Journal of Multivariate Analysis 185 (2021) 104778

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Testing the equality of a large number of means of functional data

M. Dolores Jiménez-Gamero [a], Alba M. Franco-Pereira [b,*]

[a] Dpto. Estadística e Investigación Operativa, Universidad de Sevilla, Spain
[b] Dpto. Estadística e Investigación Operativa, Universidad Complutense de Madrid, Spain

ARTICLE INFO

Article history:
Received 23 July 2020
Received in revised form 18 May 2021
Accepted 18 May 2021
Available online 25 May 2021

AMS 2010 subject classifications:
62G10

Keywords:
Asymptotic normality
Consistency
Finite dimensional data
Functional data
Testing the equality of means

ABSTRACT

Given k independent samples of functional data, this paper deals with the problem of testing for the equality of their mean functions. In contrast to the classical setting, where k is kept fixed and the sample size from each population increases without bound, here k is assumed to be large and the size of each sample is either bounded or small in comparison to k. A new test is proposed. The asymptotic distribution of the test statistic is stated under the null hypothesis of equality of the k mean functions as well as under alternatives, which allows us to study the consistency of the test. Specifically, it is shown that the test statistic is asymptotically free distributed under the null hypothesis. The finite sample performance of the test based on the asymptotic null distribution is studied via simulation. Although we start by assuming that the data are functions, the proposed test can also be applied to finite dimensional data. The practical behavior of the test for one dimensional data is numerically studied and compared with other tests.

1. Introduction

Let $X_1, \ldots, X_k$ be k random vectors taking values in $\mathbb{R}^d$, for some $d \in \mathbb{N}$, with finite means and variances. Let $X_{i,1}, \ldots, X_{i,n_i}$ be i.i.d. (independent and identically distributed) from X_i, $1 \leq i \leq k$. The k samples are assumed to be independent. A classical problem in Statistics is that of testing for the equality of the means of the k populations, that is, testing for

$$H_0 : \mu_1 = \cdots = \mu_k,$$

with $\mu_j = E(X_j)$, $1 \leq j \leq k$. When $d = 1$, if the populations are homoscedastic, following normal distributions and k is a fixed quantity, then the ANOVA F-test provides an exact solution to the above problem. If the normality assumption is dropped but the sample sizes are large, then the F-test is asymptotically valid. For heteroscedastic populations, Cochran's test gives an asymptotically free distributed test. A similar scenario could be described for d-variate data.

In the context of functional data, it is also of interest testing for the equality of mean functions. Some papers have dealt with this problem: Benko et al. [4] (for $k = 2$), Horváth and Kokoszka [19] (for $k = 2$) and Horváth and Rice [20] (for $k \geq 2$) have proposed tests based on projections onto the space determined by the leading eigenfunctions of an estimator of the covariance operator of the joint population; Ghiglietti et al. [11] (for $k = 2$) have proposed a test based on a generalized Mahalanobis distance for infinite dimensional spaces whose computation involves calculating the eigenvalues and eigenfunctions of an estimator of the covariance operator of the joint population; Yuan et al. [28] and Zhang et al. [33]

* Corresponding author.
E-mail address: albfranc@ucm.es (A.M. Franco-Pereira).

https://doi.org/10.1016/j.jmva.2021.104778

(for $k = 2$), and Cuevas et al. [9], Zhang et al. [31,32] and the approaches reviewed in Zhang [30] (for $k \geq 2$) consider tests based on calculating the norms of the difference between sample means; Cuesta-Albertos and Febrero-Bande [8] (for $k \geq 2$) have proposed a test based on random projections; among many others. A common way of approximating the null distribution of a test statistics is by means of its asymptotic null distribution, whenever it does not depend on unknown quantities. The test statistics in [4,19,20] are (under the null) asymptotically free distributed, although the associated tests are not universally consistent. Other test statistics have asymptotic null distributions depending on unknowns and, hence, their null distribution must be approximated by using alternative estimators. The consistency of the bootstrap approximation has been stated in Paparoditis and Sapatinas [24].

So far we have assumed that k is arbitrary but fixed and used the term asymptotically to mean $n_i \to \infty$, $1 \leq i \leq k$. The case in which k is allowed to increase has also received attention in the statistical literature. In particular, when $d = 1$, if the sample sizes $n_1 = \cdots = n_k$ are fixed and assuming that the data are i.i.d., Boos and Brownie [5], and Akritas and Arnold [1] have proven that, conveniently normalized, the F-test has asymptotically a standard normal distribution. This result is also true for unbalanced designs, but in such a case the asymptotic variance depends on the kurtosis of the population (the fourth moment is assumed to exist). Harrar and Gupta [17] devised a transformation that speeds up the convergence to the limiting distribution. Assuming independent data from a d-variate population ($d \geq 2$) with the same distribution across the groups except for their means, which are allowed to vary between groups, Gupta et al. [15] have proven that, conveniently normalized, commonly used statistics in MANOVA are also asymptotically normal distributed. Besides, they proved that the variance of the asymptotic distribution depends on fourth order moments of the population, which are assumed to exist. Although it is not required the sample sizes to be equal, they are assumed to be close (in certain sense). Notice that those papers assume, among other things, that the populations have equal variances. For univariate normal populations, that may have different variances, when the samples have "similar" sample sizes (in the sense that $\max_i n_i < a \min_i n_i$, for some positive constant a), Park and Park [25] proposed two tests whose associated statistics, conveniently normalized, are asymptotically normal. The assumption that the populations have all of them a normal distribution is crucial in order to derive the asymptotic distribution. In fact, we have carried out some simulations, with data coming from a negative exponential distribution (and other non-normal populations) for several values of the common mean and several values of k, and observed that the empirical type I errors were far apart from the nominal value (results are reported in Section 7). Akritas and Papadatos [2] and Harrar and Bathke [16] proposed tests for the equality of a large number of univariate and d-variate means, not requiring homoscedasticity, respectively. Assumption 1 in [16], needed to derive the asymptotic null distribution of the tests statistics in that paper, and some of these test statistics cannot be carried to the infinite dimensional case.

This paper proposes and studies a test of H_0 with k large, requiring neither normality nor homoscedasticity nor equality of distributions. We start by assuming that the data are functions. Our procedure is fully functional, not based on projections. The test statistic is asymptotically free distributed, not relying on resampling or Monte-Carlo methods to obtain critical values. Then, we show the way the new test can be applied to data with finite (but arbitrary) dimension. The sample sizes can be bounded or they can increase with k at a certain rate that will be specified later in the text. The design can be balanced or unbalanced but, for unbalanced designs, the ratio between the maximum and the minimum sample size cannot be greater than a certain power of k. One of the advantages of our proposal is that, in comparison with existing tests for finite dimensional data, it requires weaker assumptions.

The paper unfolds as follows. Based on a characterization of H_0, Section 2 proposes a test statistic. In Section 3 it is shown that, adequately normalized, it asymptotically has a standard normal distribution. The practical use of such statistic requires an estimator of its null variance. This issue is dealt with in Section 4. The power of the test rejecting the null for large values of the test statistic is studied in Section 5, in which alternatives that this test is able to detect are identified. The results in Sections 2–5 are valid for $k \to \infty$. Section 6 summarizes the results of a simulation study designed to study the goodness of the asymptotic null distribution as an approximation of the null distribution of the test statistic, and to study the power of the proposed test, for a finite k. This section also contains two real data set applications. In the previous paragraph we cited some tests for the comparison of the means of $k(\geq 2)$ independent populations, that were designed for k fixed and increasing sample sizes. One may wonder if such tests are still valid in the setting considered in this paper. Section 7 deals with this issue for the tests proposed in [8,9,20,31,32] and the approaches reviewed in [30]. This section also investigates the validity of the new test in the case of a fixed k and large sample sizes. Although in those sections we have assumed functional data, Section 8 shows how the test can be applied to finite dimensional data. This section also displays the results of some simulation experiments, where the proposed test is compared with those in [25] for one-dimensional data. A real data set application is included as well. Section 9 concludes the paper. All proofs are deferred to Section 10. All computations have been carried out using programs written in the R language [26]. The R code for the calculation of the proposed test statistic is available in the Supplementary Material.

The following notation will be used along the paper: all limits in this paper are taken when $k \to \infty$; $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution; $\xrightarrow{P}$ denotes convergence in probability; $\xrightarrow{a.s.}$ denotes the almost sure convergence; M denotes a positive constant whose exact value is unimportant and may vary across the text; and an unspecified integral denotes integration over the compact interval [0,1].

2. The test statistic

Assume that $X_1, \ldots, X_k$ are random functions defined on a common probability space $(\Omega, \mathcal{A}, P)$ with values in the separable Hilbert space $L^2 = L^2([0,1], \mathbb{R})$, the space of square-integrable real-valued functions defined on the compact interval [0,1], with the usual inner product $\langle f, g\rangle = \int f(t)g(t)dt$ and norm $\|f\| = \langle f, f\rangle^{1/2}$, $f, g \in L^2$. Assume that X_i is integrable, that is, that $E\|X_i\| < \infty$, which implies that there is a unique function $\mu_i \in L^2$, the mean function of X_i, satisfying $E(\langle X_i, x\rangle) = \langle \mu_i, x\rangle$, $\forall x \in L^2$, $1 \le i \le k$. It follows that $\mu_i(t) = E\{X_i(t)\}$ for almost all $t \in [0,1]$. Let $c_i(t,s) = E[\{X_i(t) - \mu_i(t)\}\{X_i(s) - \mu_i(s)\}]$, $s, t \in [0,1]$, stand for the covariance function of X_i, and let $\mathcal{C}_i : L^2 \to L^2$ denote the covariance operator of X_i, defined as $\mathcal{C}_i f(t) = \int c_i(t,s) f(s)ds$, $\forall f \in L^2$, which is assumed to satisfy

$$0 < \theta_i = \iint c_i^2(t,s)dtds < \infty, \quad 1 \le i \le k. \tag{1}$$

Along this paper it is assumed that the available data consist of independent random samples from each population, that is, we assume that $\mathbf{X}_i = \{X_{i1}, \ldots, X_{in_i}\}$ are n_i independent random functions from X_i, $1 \le i \le k$, and that $\mathbf{X}_1, \ldots, \mathbf{X}_k$ are independent. On the basis of the available data, we are interested in testing H_0, where the equalities under H_0 are understood in the L^2 sense, that is, $\|\mu_i - \mu_j\| = 0$, $\forall i \ne j$. Notice that H_0 is equivalent to $D_k = 0$, where

$$D_k = \frac{1}{k}\sum_{i=1}^{k} \|\mu_i - \bar{\mu}_.\|^2 = \frac{1}{k}\sum_{i=1}^{k} \langle \mu_i, \mu_i\rangle - \langle \bar{\mu}_., \bar{\mu}_.\rangle, \quad \bar{\mu}_. = \frac{1}{k}\sum_{i=1}^{k} \mu_i.$$

Since the sample mean is unbiased for the population mean, the above equivalence leads us to consider the following test statistic for testing H_0,

$$\widetilde{T}_k = \frac{1}{k}\sum_{i=1}^{k} \|\bar{X}_{i.} - \bar{X}_{..}\|^2, \quad \text{with} \quad \bar{X}_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}, \ 1 \le i \le k, \quad \bar{X}_{..} = \frac{1}{k}\sum_{i=1}^{k} \bar{X}_{i.}.$$

Routine calculations show that

$$\widetilde{T}_k = \left(\frac{1}{k} - \frac{1}{k^2}\right)\sum_{i=1}^{k} \tilde{h}_1(\mathbf{X}_i) - \frac{1}{k^2}\sum_{\substack{i,l=1\\ i\ne l}}^{k} h_2(\mathbf{X}_i, \mathbf{X}_l),$$

with $\tilde{h}_1(\mathbf{X}_i) = h_2(\mathbf{X}_i, \mathbf{X}_i)$, $h_2(\mathbf{X}_i, \mathbf{X}_l) = \langle \bar{X}_{i.}, \bar{X}_{l.}\rangle$. We have that $E\{h_2(\mathbf{X}_i, \mathbf{X}_l)\} = \langle \mu_i, \mu_l\rangle$, $\forall i \ne l$, and, assuming $\int c_i(t,t)dt < \infty$,

$$E\{\tilde{h}_1(\mathbf{X}_i)\} = \langle \mu_i, \mu_i\rangle + \frac{1}{n_i}\int c_i(t,t)dt.$$

Therefore $\widetilde{T}_k$ is a biased estimator of D_k. The bias can be removed by replacing $\tilde{h}_1(\mathbf{X}_i)$ with $h_1(\mathbf{X}_i)$, defined as

$$h_1(\mathbf{X}_i) = \frac{1}{n_i(n_i-1)}\sum_{\substack{u,v=1\\ u\ne v}}^{n_i} \langle X_{iu}, X_{iv}\rangle.$$

Let

$$T_k = \left(\frac{1}{k} - \frac{1}{k^2}\right)\sum_{i=1}^{k} h_1(\mathbf{X}_i) - \frac{1}{k^2}\sum_{\substack{i,l=1\\ i\ne l}}^{k} h_2(\mathbf{X}_i, \mathbf{X}_l).$$

Then, $E(T_k) = D_k$. Since $E(T_k) = 0$ under H_0 and $E(T_k) > 0$ under alternatives, the null hypothesis is rejected for large values of T_k. Now, in order to determine what large values are in this context, we have to calculate its distribution under the null hypothesis, or (at least) an approximation of it. The null distribution of T_k is clearly unknown. Next, we try to approximate it by means of its asymptotic null distribution.

3. The asymptotic distribution of the test statistic

By applying the Decomposition Lemma of Efron and Stein [10], after some rearrangements, one gets that

$$T_k = D_k + T_{k,Lin} + R_k, \tag{2}$$

with $T_{k,Lin} = (1/k)\sum_{i=1}^{k} L_{ik}$, $L_{ik} = h_1(\mathbf{X}_i) - \langle \mu_i, \mu_i\rangle - 2\langle \bar{X}_{i.} - \mu_i, \bar{\mu}_.\rangle$, $1 \le i \le k$, and

$$R_k = -\frac{1}{k^2}\sum_{i=1}^{k} \{h_1(\mathbf{X}_i) - \langle \mu_i, \mu_i\rangle\} + \frac{2}{k^2}\sum_{i=1}^{k} \langle \bar{X}_{i.} - \mu_i, \mu_i\rangle - \frac{1}{k^2}\sum_{\substack{i,l=1\\ i\ne l}}^{k} \langle \bar{X}_{i.} - \mu_i, \bar{X}_{l.} - \mu_l\rangle.$$

Clearly, $E(T_{k,Lin}) = E(R_k) = 0$. The variance of $T_{k,Lin}$ is given in the next lemma.

Lemma 1. *Suppose that $X_1, \ldots, X_k$ are integrable satisfying (1), that $\mathbf{X}_i = \{X_{i1}, \ldots, X_{in_i}\}$ are $n_i \geq 2$ independent random functions from X_i, $1 \leq i \leq k$, and that $\mathbf{X}_1, \ldots, \mathbf{X}_k$ are independent. Then*

$$var(T_{k,Lin}) = \frac{1}{k^2}\sum_{i=1}^{k}\left(\frac{2}{n_i(n_i-1)}\theta_i + \frac{4}{n_i}\gamma_i\right),$$

where

$$\gamma_i = \iint c_i(t,s)\{\mu_i(t) - \bar{\mu}_.(t)\}\{\mu_i(s) - \bar{\mu}_.(s)\}dtds, \quad 1 \leq i \leq k.$$

Notice that $var(T_{k,Lin}) > 0$ is equivalent to $\sum_{i=1}^{k}\theta_i > 0$, and the latter is implied by (1).
If H_0 is true, then $var(T_{k,Lin})$ simplifies to

$$var_0(T_{k,Lin}) = \frac{2}{k^2}\sum_{i=1}^{k}\frac{1}{n_i(n_i-1)}\theta_i. \tag{3}$$

The next result shows that the term R_k in (2) is negligible in comparison to $T_{k,Lin}$.

Lemma 2. *Suppose that the assumptions in Lemma 1 are fulfilled. Then $var(R_k)/var(T_{k,Lin}) \to 0$.*

Since $T_{k,Lin}$ is an average of independent random variables it readily follows that, conveniently scaled and under some assumptions (specifically, that the Lindeberg condition in (4) holds true), it is asymptotically normally distributed. This fact together with the result in Lemma 2 leads to the following result.

Theorem 1. *Suppose that the assumptions in Lemma 1 are fulfilled and that*

$$\frac{1/k^2}{var(T_{k,Lin})}\sum_{i=1}^{k}E[L_{ik}^2 I\{L_{ik}^2 > \varepsilon\, k^2\, var(T_{k,Lin})\}] \to 0, \quad \forall \varepsilon > 0. \tag{4}$$

Then

$$\frac{T_k - D_k}{\sqrt{var(T_{k,Lin})}} \xrightarrow{\mathcal{L}} Z, \tag{5}$$

where Z has a standard normal distribution.

As an immediate consequence of Theorem 1, taking into account that H_0 is equivalent to $D_k = 0$, we next derive the asymptotic null distribution of T_k.

Corollary 1. *Suppose that the assumptions in Theorem 1 are fulfilled and that H_0 is true. Then $\sqrt{k}T_k/\sigma_{0k} \xrightarrow{\mathcal{L}} Z$, where Z has a standard normal distribution and $\sigma_{0k}^2 = k\, var_0(T_{k,Lin})$, with $var_0(T_{k,Lin})$ as defined in (3).*

Recall that the test is one-sided rejecting the null hypothesis for large values of T_k. If σ_{0k}^2 were a known quantity, in view of Corollary 1, the test that rejects H_0 when

$$\sqrt{k}T_k/\sigma_{0k} \geq z_{1-\alpha},$$

for some $\alpha \in (0, 1)$, where $\Phi(z_{1-\alpha}) = 1-\alpha$ and Φ stands for the cumulative distribution function of the standard normal distribution, would have (asymptotic) level α.

In general, checking that Lindeberg condition holds is not an easy task. Because of this reason, in most cases it is easier to see that some sufficient conditions are met. Next we provide some conditions for (4) to hold. First we list such conditions, then we comment on them, and finally Proposition 1 shows that they imply (4).

Assumption 1. $\exists\, \tau > 0$ and $k_0 = k_0(\tau) \in \mathbb{N}$ such that $\frac{1}{k}\sum_{i=1}^{k}\theta_i \geq \tau$, $\forall k \geq k_0$.

Assumption 2. $E(\|X_i - \mu_i\|^4) \leq M$, $\forall i$.

Assumption 3. $\frac{n_{max}^4}{n_{min}^4}\frac{1}{k} = o(1)$, where $n_{max} = \max_{1\leq i\leq k} n_i$ and $n_{min} = \min_{1\leq i\leq k} n_i$.

Assumption 4. Either $\gamma_1 = \cdots = \gamma_k = 0$, or not all γ_is are equal to 0 and $(1/k)\Gamma_k = o(1)$, where $\Gamma_k = (1/k)\left\{\sum_{i=1}^{k}E(\langle X_i - \mu_i, \mu_i - \bar{\mu}_.\rangle^4)/n_i^2\right\}/\left(\frac{1}{k}\sum_{i=1}^{k}\gamma_i/n_i\right)^2$.

Assumption 1 is tantamount to say that there is a percentage of $100 \cdot p\%$ of covariance operators, for some $0 < p \leq 1$, say $\theta_1, \ldots, \theta_{[pk]}$, such that $\theta_i \geq c$, for some $c > 0$, $1 \leq i \leq [nk]$. This is clearly a nonrestrictive assumption. Assumption 2 is similar or even weaker than other comparable conditions used in the finite dimensional case: for univariate data, if the design is not balanced, the derivations in Boos and Brownie [5] require the fourth moment to exist; for data with dimension greater than 1 but finite, Assumption 1 in Gupta et al. [15] is stronger since it requires the finiteness of the moment of order $4 + \delta$ of the absolute value of each component of the random vector, for some $\delta > 0$. Assumption 3 tells us how different the sample sizes can be in relation to k. In particular, for either balanced designs, $n_1 = \cdots = n_k$, or bounded sample sizes, it imposes no restriction. In the finite dimensional setting, for univariate normal populations, Park and Park [25] assume that $\max_i n_i < a \min_i n_i$, for some positive constant a, which is more limiting than our assumption. Finally, we will see that Assumption 4 is not a severe restriction. Under H_0 it is automatically satisfied because $\gamma_1 = \cdots = \gamma_k = 0$. Thus, it only happens to be a constraint under alternatives. Let $Y_i = \langle X_i - \mu_i, \mu_i - \bar{\mu}_. \rangle / \sqrt{n_i}$ $1 \leq i \leq k$. Notice that $Y_1, \ldots, Y_k$ are k independent random variables. Let W_k be an equal mixture of the random variables $Y_1, \ldots, Y_k$. Then Γ_k is the kurtosis of W_k. This assumption allows the kurtosis of W_k to increase with k, but at a lower rate than k, which seems not to be a hard limitation.

Proposition 1. *Suppose that the assumptions in Lemma 1 and Assumptions 1–4 are fulfilled. Then (4) holds.*

4. Estimating the null variance

In order to get a critical region for testing H_0 using the result in Corollary 1, we need a consistent estimator of σ_{0k}. Recall that $\sigma_{0k}^2 = k \times var_0(T_{k,Lin}) = 2/k \sum_{i=1}^{k} \theta_i / n_i(n_i - 1)$. σ_{0k}^2 is unknown because $\theta_1, \ldots, \theta_k$ are unknown quantities. From expression (1), θ_i can be unbiasedly estimated by replacing $c_i^2(t, s)$ with an unbiased estimator. Assume that $n_i \geq 4$, $1 \leq i \leq k$, then $c_i^2(t, s)$ can be unbiasedly estimated by

$$\widehat{c_i^2}(t, s) = \frac{1}{n_i(n_i - 1)(n_i - 2)(n_i - 3)} \sum_{1 \leq u \neq v \neq w \neq z \leq n_i} h(X_{iu}, X_{iv}; t, s) h(X_{iw}, X_{iz}; t, s),$$

where $h(X_{iu}, X_{iv}; t, s) = \frac{1}{2}\{X_{iu}(t) - X_{iv}(t)\}\{X_{iu}(s) - X_{iv}(s)\}$.

Note that

$$\widehat{c_i^2}(t, s) = \frac{1}{n_i(n_i - 1)} \sum_{1 \leq u \neq v \leq n_i} h(X_{iu}, X_{iv}; t, s) S_{i(u,v)}^2(t, s), \tag{6}$$

where $S_{i(u,v)}^2(t, s) = \frac{1}{n_i - 3} \sum_{w \neq u,v} \{X_{iw}(t) - \bar{X}_{i(u,v)}(t)\}\{X_{iw}(s) - \bar{X}_{i(u,v)}(s)\}$, and $\bar{X}_{i(u,v)}(t) = \frac{1}{n_i - 2} \sum_{w \neq u,v} X_{iw}(t)$. Formula (6) is useful for computational purposes.

Let $\hat{\sigma}_{0k}^2 = (2/k) \sum_{i=1}^{k} \hat{\theta}_i / n_i(n_i - 1)$, where $\hat{\theta}_i = \iint \widehat{c_i^2}(t, s) dt ds$. Next proposition shows that, under some mild assumptions, the quotient $\hat{\sigma}_{0k}^2 / \sigma_{0k}^2$ converges in probability to 1.

Proposition 2. *Suppose that the assumptions in Lemma 1 and Assumptions 1–3 are fulfilled, and that $n_i \geq 4$, $1 \leq i \leq k$. Then $\hat{\sigma}_{0k}^2 / \sigma_{0k}^2 \xrightarrow{P} 1$.*

As an immediate consequence of Corollary 1 and Proposition 2, we have the following result.

Theorem 2. *Suppose that the assumptions in Corollary 1 and Proposition 2 are fulfilled. Then $\sqrt{k} T_k / \hat{\sigma}_{0k} \xrightarrow{\mathcal{L}} Z$, where Z has a standard normal distribution.*

Let $\alpha \in (0, 1)$. As a consequence of Theorem 2, the test rejecting H_0 when $\sqrt{k} T_k / \hat{\sigma}_{0k} \geq z_{1-\alpha}$ is asymptotically correct in the sense that its type I error is asymptotically equal to the nominal value α.

Remark 1. Let $R(X_{11}, \ldots, X_{1n_1}, \ldots, X_{k1}, \ldots, X_{kn_k}) = \sqrt{k} T_k / \hat{\sigma}_{0k}$. Notice that $R(aX_{11} + \mu, \ldots, aX_{1n_1} + \mu, \ldots, aX_{k1} + \mu, \ldots, aX_{kn_k} + \mu) = R(X_{11}, \ldots, X_{1n_1}, \ldots, X_{k1}, \ldots, X_{kn_k})$, $\forall a \in \mathbb{R}$, $a \neq 0$, $\forall \mu \in L^2$, that is, the test statistic is invariant under location and scale transformations.

Before ending this section, we want to underline that $\hat{\sigma}_{0k}$ is consistent for σ_{0k} under quite weak assumptions. If stronger conditions are assumed, then other estimators of σ_{0k} may be considered. For instance, if we could reasonably assume that $X_1, \ldots, X_k$ have a common variance function, $c_1(\cdot, \cdot) = \cdots = c_k(\cdot, \cdot) := c(\cdot, \cdot)$, then $\sigma_{0k}^2 = k \times var_0(T_{k,Lin}) = (\theta/k) \sum_{i=1}^{k} 1/n_i(n_i - 1)$, where $\theta = \iint c^2(t, s) dt ds$. In this case we have to estimate θ. Under H_0, if $E[\{X_i(t) - \mu_i(t)\}^2 \{X_i(s) - \mu_i(s)\}^2] \leq M$, $1 \leq i \leq k$, $0 \leq t, s \leq 1$, for some positive constant M, and if $\iint c(t, t) dt < \infty$, then θ can be consistently estimated by means of $\hat{\theta} = \iint \hat{c}(t, s)^2 dt ds$, where

$$\hat{c}(t, s) = (N - k)^{-1} \sum_{i,j} \left\{X_{ij}(t) - \bar{X}_{i.}(t)\right\} \left\{X_{ij}(s) - \bar{X}_{i.}(s)\right\}.$$

A sketch of the proof of the consistency of $\hat{\theta}$ can be found in Section 10.

5. Power

This section studies what sort of alternatives can be detected by the proposed test. Let $\sigma_k^2 = k \times var(T_{k,Lin})$. Notice that

$$0 \le \sigma_{0k}^2/\sigma_k^2 \le 1. \tag{7}$$

Under the assumptions in Theorem 1 and Proposition 2, we have that

$$P\left(\sqrt{k}\frac{T_k}{\hat{\sigma}_{0k}} > z_{1-\alpha}\right) = P\left(\sqrt{k}\frac{T_k - D_k}{\sigma_k} > \frac{\hat{\sigma}_{0k}}{\sigma_k} z_{1-\alpha} - \sqrt{k}\frac{D_k}{\sigma_k}\right) \approx \Phi\left(\sqrt{k}\frac{D_k}{\sigma_k} - \frac{\sigma_{0k}}{\sigma_k} z_{1-\alpha}\right). \tag{8}$$

In the light of (7) and (8), and taking into account that $D_k \ge 0$, we consider the following three cases: (i) $\sqrt{k}D_k/\sigma_k \to 0$, (ii) $\sqrt{k}D_k/\sigma_k \to \delta \in (0, \infty)$, and (iii) $\sqrt{k}D_k/\sigma_k \to \infty$.

In case (iii) it is clear that $P(\sqrt{k}T_k/\hat{\sigma}_{0k} > z_{1-\alpha}) \to 1$, and thus the test is consistent against that sort of alternatives. From (8), it is apparent that to derive the power in cases (i) and (ii) we must first study the quotient σ_{0k}/σ_k. Next proposition shows that, in these cases, $\sigma_{0k}/\sigma_k \to 1$.

Proposition 3. *Suppose that the assumptions in Lemma 1 and Assumptions 1 and 2 are fulfilled, that*

$$n_{max}^2 = o(k), \tag{9}$$

and $\sqrt{k}D_k/\sigma_k \to \delta \in [0, \infty)$. *Then* $\sigma_{0k}/\sigma_k \to 1$.

As an immediate consequence of Theorem 1 and Propositions 2 and 3, we have the following.

Corollary 2. *Suppose that the assumptions in Theorem 1 and Propositions 2 and 3 are fulfilled. Then* $\sqrt{k}T_k/\hat{\sigma}_{0k} \xrightarrow{\mathcal{L}} Z + \delta$, *where Z has a standard normal distribution.*

From Corollary 2, in cases (i) and (ii) we have that

$$P\left(\sqrt{k}\frac{T_k}{\hat{\sigma}_{0k}} > z_{1-\alpha}\right) \to \Phi\left(\delta - z_{1-\alpha}\right) = \begin{cases} \alpha & \text{if } \delta = 0, \\ > \alpha & \text{if } \delta \in (0, \infty). \end{cases}$$

Summarizing, the proposed test asymptotically detects those alternatives such that $\sqrt{k}D_k/\sigma_k \to \delta \in (0, \infty]$. It is consistent against those alternatives satisfying $\sqrt{k}D_k/\sigma_k \to \infty$. The alternatives fulfilling $\sqrt{k}D_k/\sigma_k \to \delta \in (0, \infty)$ play the role of contiguous alternatives in the classical setting of fixed k and large sample sizes. The test is not able to detect those cases in which $\sqrt{k}D_k/\sigma_k \to 0$. This is not surprising since it entails that the between means variance D_k is much more smaller than the standard deviation of T_k.

As an illustration, we consider the case where $1 \le m = m_k \le k$ populations have an equal mean, say $\mu \ne 0$, and the other $k - m$ populations have a mean equal to 0. Let $p_k = m/k$. For simplicity, we will also assume that the populations are homoscedastic, that is, $c_i(t, s) = c(t, s)$, $1 \le i \le k$, $\forall t, s \in [0, 1]$, and that the design is balanced, that is, $n_i = n$, $1 \le i \le k$. Let $\gamma = \iint \mu(t)\mu(s)c(t, s)dtds$, $\theta = \iint c^2(t, s)dtds$, and assume that both quantities are positive. With this notation, $\delta_k := \sqrt{k}D_k/\sigma_k = \sqrt{k}p_k(1 - p_k)\|\mu\|^2/\sqrt{2\theta/n(n-1) + 4\gamma p_k(1 - p_k)/n}$.

In this setting, the question that naturally arises is how small (or big) must m be so that the test can detect that H_0 is not true? In order to answer this question, we will assume that condition (9) in Proposition 3 holds. If m is a fixed number that does not depend on k (or equivalently $k - m$ is a fixed number that does not depend on k), then $\delta_k \to 0$. Therefore, the test it is not able to detect alternatives with a finite number of different means. This is a special case of $p_k \to 0$. Next, we see that the test can detect alternatives with $p_k \to 0$ but at a lower speed than $1/k$. The following cases can arise:

(a) $p_k \to p \in (0, 1)$: then $\delta_k \to \infty$, for any choice of n, and thus we are in case (iii);
(b) $p_k \to 0$ and n remains bounded: then the asymptotic behavior of δ_k coincides with that of $\sqrt{k}p_k$, and thus,

  (b.1) if $\sqrt{k}p_k \to 0$, which is equivalent to $p_k = o(k^{-1/2})$, then we are in case (i),
  (b.2) if $\sqrt{k}p_k \to \delta \in (0, \infty)$, then we are in case (ii),
  (b.3) if $\sqrt{k}p_k \to \infty$, then we are in case (iii);

(c) $p_k \to 0$, $n \to \infty$, $np_k \to \rho \in (0, +\infty]$: then $\delta_k \to \infty$ and thus we are in case (iii).
(d) $p_k \to 0$, $n \to \infty$, $np_k \to 0$: then the asymptotic behavior of δ_k coincides with that of $\sqrt{knp_k}$, which from (9) is $\sqrt{knp_k} = o(kp_k)$, and thus,

  (d.1) if $\sqrt{knp_k} \to 0$, then we are in case (i),
  (d.2) if $\sqrt{knp_k} \to \delta \in (0, \infty)$, then we are in case (ii),
  (d.3) if $\sqrt{knp_k} \to \infty$, then we are in case (iii).

The answer to the question *how small must m be so that the test can detect that H_0 is not true* is that we must take $m = O(f(k))$ and $n = O(\sqrt{k}/f(k))$, with f such that $f(k) \to \infty$, $f(k)/k \to 0$ and $n \to \infty$. For instance, we could take $m = k^a$ and $n = k^{0.5-a}$, for some $0 < a < 0.5$, or $m = \log(k)$ and $n = \sqrt{k}/\log(k)$, among many possible choices.

Remark 2. With the aim of improving the asymptotic power of the test in case (d.2), an anonymous reviewer suggested us to consider a max-type statistic, instead of T_k. Reasoning as in Section 2, H_0 is equivalent to $\max_{i\neq j} \|\mu_i - \mu_j\|^2 = 0$. Now, since $\|\mu_i - \mu_j\|^2$ can be unbiasedly estimated by means of $T_{ij} = h_1(\mathbf{X}_i) + h_1(\mathbf{X}_j) - 2h_2(\mathbf{X}_i, \mathbf{X}_j)$, one could take $T_{\max} = \max_{i\neq j} T_{ij}$ as the test statistic. Let $T_{all} = (T_{12}, T_{13}, \dots, T_{k-1k})^\top \in \mathbb{R}^D$, where the superindex $^\top$ denotes transpose and $D = k(k-1)/2$. If $n_i/n_1 \to \tau_i \in (0, \infty)$ $\forall i$, when $\min_i n_i \to \infty$, then for each fixed k, $\sqrt{n_1}T_{all}$ converges in law to a D-variate normal random vector, $Z = (Z_1, \dots, Z_D)^\top$, with mean $\mu_{all} = (\|\mu_1 - \mu_2\|^2, \dots, \|\mu_{k-1} - \mu_k\|^2)^\top$ and a certain covariance matrix. Therefore, $T_{\max}$ converges in law to $\max_i Z_i$. To derive the asymptotic distribution of $T_{\max}$ we have assumed that $\min_i n_i \to \infty$ (not required to get the asymptotic distribution of T_k). In addition, the law of $\max_i Z_i$ is not free of nuisance parameters, even under the null hypothesis, and thus cannot be used to approximate the null distribution of $T_{\max}$. A sound study of $T_{\max}$ deserves to be considered in a separate paper.

Remark 3. All our results have been stated under the tacit assumption that realizations of $X_{11}, \dots, X_{kn_k}$, i.e., complete trajectories of functions are observable. In practice, these functions are observed at a finite grid of points and the curves $X_{11}, \dots, X_{kn_k}$ are recovered by using nonparametric techniques, such as local linear regression. The statistic is then calculated from $\widehat{X}_{11}, \dots, \widehat{X}_{kn_k}$, which stand for the resulting curve estimators. Under suitable assumptions, all previous results remain valid when the test statistic is calculated from $\widehat{X}_{11}, \dots, \widehat{X}_{kn_k}$. See, e.g., Jiang et al. [21], in particular the comments made after the proof of their Theorem 2.

6. Simulation results for a finite number of populations

Recall that the proposed test, which rejects H_0 when $\sqrt{k}T_k/\hat{\sigma}_{0k} \geq z_{1-\alpha}$, has asymptotically type I error equal to α. To study the level of the proposed test for a finite value of k, we have carried out a simulation experiment as follows. We have generated a sample from each of k populations ($k = 30, 40, 50, 100, 200, 300, 400, 500$), all of them with the same mean that, without loss of generality since the test statistic is location invariant, we took equal to 0. The size of each sample has been randomly generated from a discrete uniform random law $UD\{a, a+1, \dots, b\}$, with $(a, b) = (5, 10), (11, 20)$. The data $\{X_{ij}(t),\ t \in [0, 1],\ 1 \leq j \leq n_i,\ 1 \leq i \leq k\}$ have been generated in discretized versions $X_{ij}(t_r)$, for $r \in \{1, \dots, 50\}$, where the values t_r were chosen equispaced in the interval [0, 1]. Four different types of random functions have been considered: the data were generated from a Wiener process with dispersion parameter $\sigma = 1$ for all populations (recall from Remark 1 that the test statistic is invariant under scale transformations, so the same results would be obtained for other choices of σ), and three heteroscedastic versions, where the data from population i were generated from a Wiener process with standard deviation σ_i, which in turn was randomly generated from a uniform distribution on $(1, s)$, $1 \leq i \leq k$, $s = 1.5, 2, 3$. Each case was run 10,000 times. Table 1 shows the fractions of p-values less than or equal to 0.05 and 0.10, which are the estimated type I error probabilities for nominal significance level α=0.05 and 0.10, respectively. Looking at this table we observe that, in general, the actual levels are not far from the nominal values even for small values of k, specially in the case of homoscedastic populations; as the value of σ is allowed to vary in a wider interval, a larger value of k is required for the empirical levels to closely match the nominal significance values. We also see that the approximation is better for larger sample sizes.

To study the power, we have considered the same type of random functions as those used for the level, but now the mean function for population i was taken $\mu_i(t) = \frac{i}{k}I(t > t_0)$, $1 \leq i \leq k$, for several values of t_0. Although the test statistic is invariant under scale and location changes, the power is greatly influenced by the choice of the scale. Recall that the power is (asymptotically) an increasing function of $\sqrt{k}D_k/\sigma_k$, and thus the power increases with D_k (which measures how different the means are), with k, with the sample sizes (recall that σ_k decreases as the sample sizes increase), and decreases when the scale increases. Because of this reason, in order to study the power we have considered several values of the scale parameter for the homoscedastic case. Table 2 shows the fractions of p-values less than or equal to 0.05 and 0.10, which are the estimated powers for nominal significance level α=0.05 and 0.10, respectively. They have been obtained by generating 2,000 samples in each case (for the power we observed, in some preliminary simulations, that the results using 2,000 samples and those obtained with 10,000 samples were very close, so in order to save computation time, we took 2,000 samples). Looking at this table we observe that the above considerations, which are valid asymptotically, also hold for finite sample sizes.

We illustrate the applicability of our methodology through two real data examples. First, we have applied the new test to compare the mean of the occupancy rate across the 52 Spanish provinces. The data were taken from the website of the Spanish National Institute of Statistics, *http://www.ine.es*. For each province, we took as sample the monthly observed occupancy rate from 2011 to 2019. Therefore, we have $k = 52$ populations, each sample curve is observed in 12 points (months) and the sample sizes are $n_1 = n_2 = \cdots = n_{52} = 9$. We can consider the data as i.i.d. since the accompanying note states "due to different updates in the Establishments Directory, data from different years are not directly comparable". Fig. 1 displays the sample curves of the occupancy rate for each province and the group sample means. Each color

Table 1
Observed proportion of rejections in 10,000 simulated data sets for each scenario, for functional data, under the null hypothesis. The size of each sample has been randomly generated from a discrete uniform random law $UD\{a, a+1, \dots, b\}$, with $(a, b) = (5, 10), (11, 20)$. The data were generated from a Wiener process with mean 0 and with dispersion parameter $\sigma = 1$ for all populations and three heteroscedastic versions. The nominal significance levels are $\alpha = 0.05$ and 0.10.

Number of populations k	Sample sizes limits n_i	Dispersion parameters of the Wiener processes							
		$\sigma_i = 1$		$\sigma_i \sim U(1, 1.5)$		$\sigma_i \sim U(1, 2)$		$\sigma_i \sim U(1, 3)$	
		5%	10%	5%	10%	5%	10%	5%	10%
30	5–10	.0717	.1176	.0714	.1143	.0753	.1191	.0787	.1228
	11–20	.0651	.1104	.0629	.1053	.0656	.1105	.0668	.1069
40	5–10	.0666	.1102	.0716	.1195	.0696	.1134	.0740	.1198
	11–20	.0617	.1063	.0631	.1097	.0699	.1129	.0662	.1078
50	5–10	.0704	.1142	.0682	.1152	.0685	.1128	.0722	.1181
	11–20	.0612	.1093	.0677	.1115	.0582	.1036	.0694	.1135
100	5–10	.0634	.1102	.0662	.1148	.0661	.1153	.0654	.1162
	11–20	.0597	.1074	.0605	.1086	.0610	.1062	.0588	.1062
200	5–10	.0575	.1115	.0590	.1082	.0592	.1084	.0595	.1110
	11–20	.0596	.1088	.0560	.1061	.0577	.1025	.0633	.1092
300	5–10	.0588	.1081	.0553	.1024	.0641	.1138	.0595	.1096
	11–20	.0549	.1067	.0527	.1011	.0587	.1071	.0590	.1058
400	5–10	.0544	.1001	.0572	.1075	.0565	.1013	.0622	.1100
	11–20	.0546	.1045	.0569	.1059	.0590	.1023	.0546	.1046
500	5–10	.0574	.1080	.0584	.1081	.0610	.1089	.0565	.1046
	11–20	.0545	.1029	.0516	.1004	.0573	.1085	.0590	.1076

Table 2
Observed proportion of rejections in 2,000 simulated data sets for each scenario, for functional data, under alternatives. The size of each sample has been randomly generated from a discrete uniform random law $UD\{a, a+1, \dots, b\}$, with $(a, b) = (5, 10), (11, 20)$. The data were generated from a Wiener process where the mean function for population i is $\mu_i(t) = \frac{i}{k} I(t > t_0)$, $1 \le i \le k$, for several values of t_0 and different dispersion values σ. The nominal significance levels are $\alpha = 0.05$ and 0.10.

Number of populations k	t_0	Sample sizes limits n_i	Dispersion parameters of the Wiener processes											
			$\sigma_i = s$						$\sigma_i \sim U(1, s)$					
			$s = 1$		$s = 1.5$		$s = 2$		$s = 1.5$		$s = 2$		$s = 3$	
			5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
100	0.5	5–10	.9860	.9955	.6365	.7435	.3335	.4480	.8255	.9025	.5725	.6990	.2555	.3515
		11–20	1.000	1.000	.9695	.9870	.7105	.8050	.9985	.9995	.9425	.9740	.5845	.7070
	0.7	5–10	.8420	.9025	.3620	.4930	.1760	.2700	.4990	.6160	.3230	.4470	.1700	.2450
		11–20	.9985	.9995	.7725	.8620	.4010	.5400	.9350	.9720	.7125	.8165	.3515	.4595
	0.9	5–10	.2975	.4295	.1295	.2260	.0950	.1620	.1745	.2720	.1355	.2030	.0955	.1575
		11–20	.6750	.7835	.2480	.3685	.1495	.2360	.3845	.5065	.2260	.3355	.1190	.1985
200	0.5	5–10	.9995	1.000	.8665	.9255	.5005	.6300	.9735	.9870	.7750	.8635	.4005	.5270
		11–20	1.000	1.000	1.000	1.000	.9020	.9475	1.000	1.000	.9995	1.000	.7950	.8730
	0.7	5–10	.9805	.9950	.5300	.6610	.2790	.3950	.7650	.8605	.4955	.6175	.2105	.3110
		11–20	1.000	1.000	.9545	.9830	.6210	.7505	.9990	.9995	.9445	.9755	.4950	.6270
	0.9	5–10	.4240	.5645	.1595	.2560	.1185	.2020	.2525	.3575	.1640	.2490	.0925	.1720
		11–20	.9060	.9540	.3695	.5010	.1890	.2940	.5985	.7305	.3365	.4795	.1665	.2525
300	0.5	5–10	1.000	1.000	.9420	.9690	.6425	.7435	.9960	1.000	.9065	.9560	.4830	.6305
		11–20	1.000	1.000	1.000	1.000	.9740	.9905	1.000	1.000	1.000	1.000	.9240	.9640
	0.7	5–10	.9990	.9995	.6740	.7925	.3550	.4740	.8920	.9495	.6200	.7410	.2625	.3885
		11–20	1.000	1.000	.9935	.9960	.7645	.8450	1.000	1.000	.9810	.9920	.6120	.7440
	0.9	5–10	.5800	.7060	.2245	.3355	.1225	.2000	.3110	.4535	.1995	.3115	.1110	.1840
		11–20	.9885	.9960	.4975	.6335	.2575	.3695	.7265	.8245	.4415	.5870	.1960	.2960

corresponds to a province. We have applied the proposed test to check the equality of the 52 mean functions and got a p-value < 0.001 which leads us to reject the null hypothesis.

The second data set consists of samples of Auslan (Australian Sign Language) signs. Auslan is the language used by the Australian Deaf and non-vocal communities. This data set is available in the R package `mfds` (Górecki and Smaga [13]). In particular, 27 examples of each of 95 Auslan signs were captured from a native signer using high-quality position trackers and instrumented gloves and were collected over a period of nine weeks. The average length of each sign was approximately 57 frames. Each hand generated a total of 11 features: 3 for orientation (roll, pitch, yaw), 3 for position (x, y, z) and 5 for finger bends. Therefore, we have $k = 95$ populations, 22 variables, and the sample sizes are $n_1 = n_2 = \cdots = n_{95} = 27$. For a more detailed description of the variables we refer to Kadous and Sammut [23]. In Fig. 2 the sample curves of the 2565 signs and the group sample means for the variable that measures the x-position of the left hand are represented. We have applied the test to check the equality of the 95 mean functions for each of the 22 variables. In all the cases the null hypothesis is rejected.

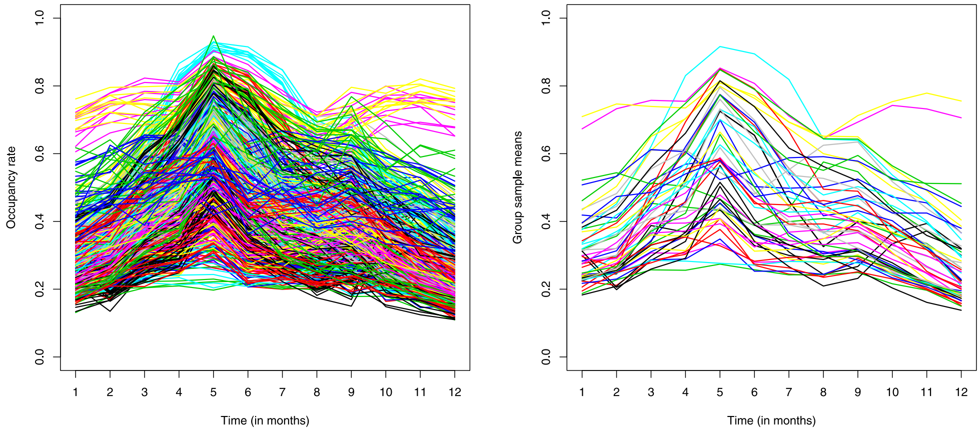

Fig. 1. All curves (left) and sample mean of each group (right) of the occupancy rate per month in Spain, from January 2011 to December 2019. There are $k = 52$ populations (the 52 Spanish province), each sample curve is observed in 12 points (months) and the sample sizes are $n_1 = n_2 = \cdots = n_{52} = 9$.

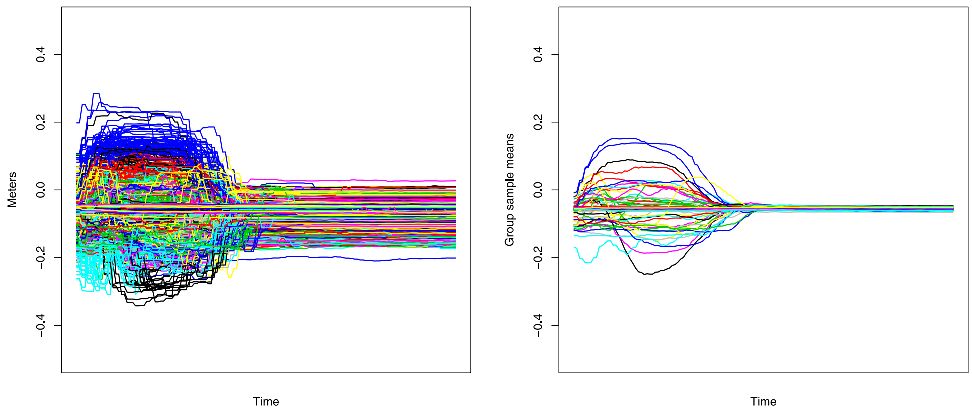

Fig. 2. All curves (left) and sample mean of each group (right) of the x-position of the left hand in the Auslan data set. There are $k = 95$ populations (the different Auslan signs), 22 variables, and the sample sizes are $n_1 = n_2 = \cdots = n_{95} = 27$.

7. Comparison with existing tests

In the Introduction we cited some tests for the comparison of the means of k independent populations, that were designed for k fixed and increasing sample sizes. One may wonder if such tests are still valid in the setting considered in this paper. This section discusses such issue for the tests in Cuesta-Albertos and Febrero-Bande [8], Horváth and Rice [20], Cuevas et al. [9], Zhang et al. [31,32] and the approaches reviewed in Zhang [30]. The test in Cuesta-Albertos and Febrero-Bande [8] consists of randomly generating an element of L^2 (according to some adequate distribution), and then projecting the data on this element. The projected data has dimension 1, and so one can apply any technique designed for the univariate case. Notice that a random projection may lead to the rejection of the null hypothesis, while another random projection could induce the opposite conclusion. To avoid this inconvenience, these authors have proposed to take several random projections, calculate the p-value for each projection, and then apply some correction, as for example the procedure in Benjamini and Yekutieli [3], which controls the false discovery rate. In order to numerically investigate the validity of this test for large k, we partially repeated the experiment in Table 1. To apply the test in [8] we used the function `fanova.RPm` from the `R` package `fda.usc`. For each case, we took 5, 10 and 30 random projections and adjusted the global p-value by using the procedure in [3]. Results based on 1000 simulated functions for each scenario are reported in Table 3. Looking at this table we see that the test becomes more liberal as k increases, specially for smaller sample sizes. In the light of the outputs of this experiment, it can be concluded that the test in [8] is not valid in the setting of a large k and small n_i.

The test in Horváth and Rice [20] is also based on projections. Specifically, it considers the projections of the data on the eigenfunctions associated with the d largest eigenvalues of an estimator of a certain covariance operator. These projections are also known as the scores. This way each function in the data set is transformed into a point of $\mathbb{R}^d$. Then, these authors build a test statistic that compares the sample mean of the scores of each sample with an estimator of the

Table 3
Observed proportion of rejections in 1000 simulated data sets for each scenario, for functional data, under the null hypothesis, obtained by applying the test in Cuesta-Albertos and Febrero-Bande [8] with different numbers of random projections (RP). The nominal significance levels are $\alpha = 0.05$ and 0.10.

Number of populations k	Sample sizes limits \ Number of projections n_i \ RP	Dispersion parameters of the Wiener processes											
		$\sigma_i = 1$						$\sigma_i \sim U(1,3)$					
		5%			10%			5%			10%		
		5	15	30	5	15	30	5	15	30	5	15	30
30	5–10	.098	.098	.075	.180	.171	.153	.096	.103	.104	.171	.173	.172
	11–20	.076	.072	.060	.133	.121	.116	.080	.076	.072	.134	.118	.119
40	5–10	.107	.093	.098	.193	.176	.161	.102	.105	.103	.185	.189	.182
	11–20	.087	.081	.084	.153	.137	.130	.071	.073	.070	.130	.117	.129
50	5–10	.140	.134	.121	.236	.226	.216	.129	.112	.114	.229	.203	.199
	11–20	.085	.072	.073	.142	.132	.111	.084	.087	.076	.145	.141	.149
100	5–10	.202	.182	.205	.330	.324	.309	.178	.165	.171	.305	.270	.273
	11–20	.108	.095	.099	.186	.182	.162	.098	.095	.085	.175	.167	.160
200	5–10	.389	.384	.388	.558	.578	.554	.306	.325	.322	.473	.479	.488
	11–20	.167	.167	.158	.285	.270	.263	.159	.156	.147	.248	.250	.246

Table 4
Observed proportion of rejections in 1000 simulated data sets for each scenario, for functional data, under the null hypothesis, obtained by applying the test in Horváth and Rice [20]. The nominal significance levels are $\alpha = 0.05$ and 0.10.

Sample sizes limits n_i	Number of populations											
	$k = 30$				$k = 40$				$k = 50$			
	Dispersion parameters of the Weiner processes											
	$\sigma_i = 1$		$\sigma_i \sim U(1,3)$		$\sigma_i = 1$		$\sigma_i \sim U(1,3)$		$\sigma_i = 1$		$\sigma_i \sim U(1,3)$	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
5–10	.844	.887	.867	.896	.937	.961	.932	.959	.972	.982	.966	.978
11–20	.429	.534	.423	.536	.514	.618	.488	.616	.563	.675	.574	.683

common mean under the null hypothesis. When the null hypothesis is true, the test statistic is asymptotically (in the sense that k is fixed and $n_i \to \infty$) free distributed. In order to numerically investigate the validity of this test for large k, we partially repeated the experiment in Table 1. Results based on 1000 simulated curves for each case are reported in Table 4. Looking at this table we see that the test is very liberal for all tried values of k and that it becomes more liberal as k increases, specially for smaller sample sizes. Therefore, it can be concluded that the test in [20] is not valid in the setting of a large k and small n_i.

Next, we deal with the test in Cuevas et al. [9] which, in a sense that will be explained next, is close to our proposal. These authors proposed to use two test statistics which, sharing the usual terminology of ANOVA, measure the between groups variability. The statistic considered in this paper, T_k, also measures such variability. If the design were balanced ($n_1 = \cdots = n_k$), then the three statistics (the two ones in [9] and T_k) are equivalent. Of course, we could have considered any other statistic measuring the between groups variability. We took T_k for three main reasons: it is unbiased for D_k and the null hypothesis can be rewritten in terms of D_k; the asymptotics for T_k (as $k \to \infty$) possess convenient expressions than can be approximated; and finally, T_k is, in some sense, similar to other statistics than have been previously considered for testing problems when the number of populations increases (see, e.g. Cousido-Rocha et al. [6] and Zhan and Hart [29] for the comparison of univariate continuous populations; Jiménez-Gamero et al. [22] for the comparison of d-variate populations; Park and Park [25] for the equality of means of univariate normal populations).

As said before, Cuevas et al. [9] proposed to use two test statistics for testing H_0. For technical reasons, they only derived the limit distribution of one of then, specifically that of $V_{n,k} = \sum_{i<j} n_i \|\bar{X}_{i.} - \bar{X}_{j.}\|^2$, which was shown to converge in law (assuming that k is fixed and $n_i \to \infty$) under the null hypothesis to $V_k = \sum_{i<j} \|Z_i - p_{ij} Z_j\|^2$, where $Z_1, \ldots, Z_k$ are independent zero mean Gaussian elements taking values in L^2, and Z_i has the same covariance function as X_i, $c_i(t,s)$, $1 \le i \le k$. Such a result was derived by assuming $n_i/n \to p_i > 0$, $1 \le i \le k$, n being the total sample size, $n = n_1 + \cdots + n_k$. With this notation, $p_{ij} = p_i/p_j$, $1 \le i, j \le k$. Rejection of H_0 is for large values of $V_{n,k}$. In order to decide what is large, one must estimate the null distribution of $V_{n,k}$. With this aim, these authors propose to use a bootstrap in the limit procedure, that consists in approximating the distribution of $V_{n,k}$ by that of V_k with p_{ij} and c_i replaced by n_i/n_j and $\hat{c}_i(t,s) = 1/(n_i - 1)\sum_{j=1}^{n_i}\{X_{ij}(t) - \bar{X}_{i.}(t)\}\{X_{ij}(s) - \bar{X}_{i.}(s)\}$, $1 \le i, j \le k$, respectively, which in turn is approximated by simulation. At this point one may wonder if the above convergence still holds when $k \to \infty$, that is, if $V_{n,k}$ and V_k converge in law to the same limit. Theorems 3 and 4 provide a decomposition for $V_{n,k}$ and V_k, respectively, which is similar to that in (2) for T_k. From those decompositions, it will be seen that, in general, $V_{n,k}$ and V_k have different limits (in law), which implies that the test in [9] is not suitable for testing H_0 in the setting of a large k and small n_i.

Theorem 3. *Suppose that $X_1,\dots,X_k$ satisfy Assumption 2, that $\mathbf{X}_i=\{X_{i1},\dots,X_{in_i}\}$ are $n_i\ge 2$ independent random functions from X_i, $1\le i\le k$, that $\mathbf{X}_1,\dots,\mathbf{X}_k$ are independent, and that H_0 is true with $\mu_1=\cdots=\mu_k=0$. Let θ_i be as defined in (1), $\tau_i=\int c_i(t,t)dt$, $\varrho_i=\iint E\left\{X_i^2(t)X_i^2(s)\right\}dtds$, $1\le i\le k$. Then*

$$\begin{aligned}
V_{n,k} &= E(V_{n,k})+V_{n,k,lin}+V_{n,k,rem},\\
E(V_{n,k}) &= \sum_{i=1}^{k-1}(k-i)\tau_i+\sum_{i=2}^{k}(n_1+\cdots+n_{i-1})\tau_i/n_i,\\
V_{n,k,lin} &= \sum_{i=1}^{k}\phi(\mathbf{X}_i),\quad \phi(\mathbf{X}_i)=\{(n_1+\cdots+n_{i-1}+n_i(k-i))\}\left(\|\bar{X}_{i.}\|^2-\frac{\tau_i}{n_i}\right),\quad 1\le i\le k,\\
E\{\phi(\mathbf{X}_i)\} &= 0,\quad 1\le i\le k,\\
var\{\phi(\mathbf{X}_i)\} &= \{(n_1+\cdots+n_{i-1}+n_i(k-i))\}^2\left\{2\frac{n_i-1}{n_i^3}\theta_i+\frac{1}{n_i^3}(\varrho_i-\tau_i)\right\},\quad 1\le i\le k.
\end{aligned}$$

If, in addition, $\sum_{i=1}^k\theta_i>0$ and $n_{max}=o(\sqrt{k})$, then $var(V_{n,k,rem})/var(V_{n,k,lin})\to 0$.

Reasoning as for T_k, from Theorem 3 it follows that, under the conditions in it and if we also assume that $\phi(\mathbf{X}_1),\dots,\phi(\mathbf{X}_k)$ meet Lindeberg condition, then

$$\frac{V_{n,k}-E(V_{n,k})}{\sqrt{var\{\phi(\mathbf{X}_1)\}+\cdots+var\{\phi(\mathbf{X}_k)\}}}\xrightarrow{\mathcal{L}} Z,$$

where Z has a standard normal distribution. This result is parallel to that obtained in Theorem 1 for T_k. Notice that Theorem 3 requires stronger assumptions than Theorem 1. In addition, the practical use of Theorem 3 would require the estimation of $var\{\phi(\mathbf{X}_1)\}+\cdots+var\{\phi(\mathbf{X}_k)\}$, which is more involved than the estimation of σ_{0k}^2.

Since we are deriving asymptotics when $k\to\infty$, instead of deriving the limit of V_k, we next derive the limit in law of $\tilde{V}_k$, defined as V_k with p_{ij} replaced by n_i/n_j.

Theorem 4. *Suppose that $Z_1,\dots,Z_k$ are independent zero mean Gaussian elements taking values in L^2 and that Z_i has the same covariance function as X_i, $1\le i\le k$. With the notation in the statement of Theorem 3 we have that*

$$\begin{aligned}
\tilde{V}_k &= E(V_{n,k})+\tilde{V}_{k,lin}+\tilde{V}_{k,rem},\\
\tilde{V}_{k,lin} &= \sum_{i=1}^{k}\varphi(Z_i),\quad \varphi(Z_i)=\{(n_1+\cdots+n_{i-1}+n_i(k-i))\}\left(\|Z_i\|^2-\tau_i\right)/n_i,\quad 1\le i\le k,\\
E\{\varphi(Z_i)\} &= 0,\quad 1\le i\le k,\\
var\{\varphi(Z_i)\} &= 2\{(n_1+\cdots+n_{i-1}+n_i(k-i))\}^2\theta_i/n_i^2,\quad 1\le i\le k.
\end{aligned}$$

If, in addition, $\sum_{i=1}^k\theta_i>0$ and $n_{max}=o(\sqrt{k})$, then $var(\tilde{V}_{k,rem})/var(\tilde{V}_{k,lin})\to 0$.

Notice that, in general, the variance of $\varphi(Z_i)$ and that of $\phi(\mathbf{X}_i)$ are different. If $X_1,\dots,X_k$ are Gaussian, then those variances coincide and, under some general conditions, both $V_{n,k}$ and $\tilde{V}_k$ have the same asymptotic distribution (when $k\to\infty$). In order to illustrate this fact numerically, we have partially repeated the experiment in Table 1 as follows: we first considered data generated from a Wiener process, as described in Section 6. This case is labeled in Table 5 as W. Since these data are Gaussian, from the above discussion, it is expected to obtain actual levels close to the nominal values. To see that this is not the case when the data are not Gaussian, we have also generated non-Gaussian data. Specifically, we have generated samples from

$$Y(t)=A_0+\sqrt{2}\sum_{j=1}^{5}C_j\cos(2\pi jt)+\sqrt{2}\sum_{j=1}^{5}S_j\sin(2\pi jt),$$

where A_0, $C_1,\dots,C_5$ and $S_1,\dots,S_5$ are independent random variables, having a Laplace distribution (two-sided exponential distribution). This case is labeled in Table 5 as L. The results in Table 5 are based on 1000 samples from each scenario, and each p-value was calculated by generating 1000 samples from (an estimation of) the asymptotic null distribution, that is, from V_k. Since the bootstrap in the limit approximation is very time consuming, we only tried $k=30,40$. Table 5 summarizes the output of this experiment. Looking at this table we see that, as expected, for the Gaussian data the observed proportion of rejections is in all cases close to the nominal value; but for non-Gaussian data the test is extremely conservative, notably for smaller sample sizes. Thus, to safely apply this test to functional data, one should first apply some Gaussianity test (see, for example, the tests in Górecki et al. [12], Cuesta-Albertos et al. [7], and Henze and Jiménez-Gamero [18]). Nevertheless, the main disadvantage of the test in [9] resides in the computational time required to calculate the p-value.

Table 5
Observed proportion of rejections in 1000 simulated data sets for each scenario, for functional data, under the null hypothesis, obtained by applying the test in Cuevas et al. [9]. W denotes the Wiener process and L the non-Gaussian data. The nominal significance levels are $\alpha = 0.05$ and 0.10.

Number of populations k	Sample sizes limits n_i	W				L			
		Dispersion parameters of the Wiener processes							
		$\sigma_i = 1$		$\sigma_i \sim U(1,3)$		$\sigma_i = 1$		$\sigma_i \sim U(1,3)$	
		5%	10%	5%	10%	5%	10%	5%	10%
30	5–10	.057	.105	.043	.095	.005	.023	.002	.015
	11–20	.047	.101	.060	.109	.005	.032	.009	.030
40	5–10	.048	.098	.049	.098	.005	.021	.002	.013
	11–20	.051	.098	.048	.089	.009	.031	.006	.028

Table 6
Observed proportion of rejections in 1000 simulated data sets for each scenario, for functional data, under the null hypothesis, obtained by applying the test in Zhang et al. [31,32] and the approaches reviewed in Zhang [30] that are described at the end of Section 7. W denotes the Wiener process and L the non-Gaussian data. The nominal significance levels are $\alpha = 0.05$ and 0.10.

	Number of populations k	Sample sizes limits n_i		Tests							
				LN	*LB*	*Lb*	*FN*	*FB*	*Fb*	*GPF*	*Fm*
W $\sigma_i = 1$	30	5–10	5%	.066	.070	.042	.040	.042	.005	.066	.048
			10%	.118	.127	.103	.101	.101	.016	.113	.094
		11–20	5%	.062	.065	.047	.054	.054	.009	.058	.048
			10%	.100	.101	.095	.089	.089	.040	.103	.093
	40	5–10	5%	.070	.072	.055	.055	.055	.003	.076	.048
			10%	.121	.126	.108	.103	.104	.008	.114	.105
		11–20	5%	.064	.065	.059	.057	.057	.021	.061	.053
			10%	.107	.108	.098	.102	.102	.044	.112	.108
L $\sigma_i = 1$	30	5–10	5%	.064	.071	.011	.052	.052	.000	.062	.056
			10%	.133	.139	.036	.113	.122	.000	.132	.106
		11–20	5%	.042	.046	.015	.037	.040	.001	.047	.042
			10%	.087	.090	.035	.079	.085	.005	.091	.086
	40	5–10	5%	.059	.061	.008	.049	.051	.000	.059	.066
			10%	.106	.118	.027	.093	.096	.000	.108	.112
		11–20	5%	.045	.047	.008	.040	.043	.001	.047	.047
			10%	.103	.103	.040	.094	.096	.002	.100	.092

We repeated the simulation experiment in Table 5 for the tests in Zhang et al. [31,32] and the approaches reviewed in Zhang [30], that are implemented in the R package `fdANOVA` (see Górecki and Smaga [14]). The tests in [31,32] are designed to compare populations with the same covariance function. Moreover, to derive their properties, it is assumed that the populations only differ in their mean curves (see Condition A (2) in [32] and Condition (A2) in [31]). The null distribution of the test statistic in [31], denoted as *Fm* in Table 6, is approximated by means of a bootstrap estimator and the null distribution of the test statistic in [32], which is denoted as *GPF* in Table 6, is approximated by using the Welch–Satterthwaite χ^2-approximation. The approaches reviewed in Zhang [30] use as test statistics $S = \sum_i n_i \|\bar{X}_i - \tilde{X}_{..}\|^2$, where $\tilde{X}_{..}(\cdot) = (1/N)\sum_{i,j} X_{ij}(\cdot)$ and $N = \sum_{i=1}^k n_i$, and $F = S/(k-1)\sum_{i,j}\|X_{ij} - \bar{X}_i\|^2$. Although they can be applied in the case of populations with different covariance functions, in `fdANOVA` they are only implemented in the homoscedastic case. To approximate the null distribution, `fdANOVA` offers three possibilities: the Welch–Satterthwaite approximation with naive and bias reduced estimators (valid when the curves are Gaussian) of the target parameters and via bootstraping. The tests based on S (F) that employ these approximations are denoted in Table 6 as *SN*, *SB* and *Sb* (*FN*, *FB* and *Fb*), respectively. Table 6 summarizes the output of this experiment for the homoscedastic case ($\sigma_i = 1$). For the heteroscedastic case ($\sigma_i \sim U(1,3)$), all tests are rather liberal. Looking at Table 6 we see that *FB*, *FN*, and *Fm* have reasonable sizes; *Lb* only works for the Gaussian data; *LN*, *LB* and *GPF* work better for large sample sizes and *Fb* is extremely conservative.

To end this Section, we investigate the adequacy of the proposed test in the classical setting: k is fixed and $\min_i n_i \to \infty$. Specifically, we derive its asymptotic null distribution in such a case. Under H_0, we always have that

$$k \times var(T_k) = \frac{(k-1)^2}{k^3}\sum_{i=1}^{k}\frac{2}{n_i(n_i-1)}\theta_i + \frac{2}{k^3}\sum_{i\neq j}\theta_{ij},$$

with $\theta_{ij} = \iint c_i(t,s)c_j(t,s)dtds$. Thus, for moderate and large k, $k \times var(T_k) \approx \sigma_{0k}^2$, which is the variance we used to normalize T_k. If $n_1/n_i \to \tau_i \in (0,\infty)$ $\forall i$, as $\min_i n_i \to \infty$, then, under H_0 and assuming that $\eta_i = \int c_i(t,t)dt < \infty$ $\forall i$, routine calculations show that

$$n_1 T_k \xrightarrow{\mathcal{L}} \frac{k-1}{k^2}\sum_{i=1}^{k}(\langle G_i, G_i\rangle - \eta_i) - \frac{1}{k^2}\sum_{i\neq j}\langle G_i, G_j\rangle,$$

Table 7
Observed proportion of rejections in 10,000 simulated data sets for each scenario, for functional data, under the null hypothesis in the classical framework. That is, k is fixed and $\min_i n_i \to \infty$. W denotes the Wiener process and L the non-Gaussian data. The nominal significance levels are $\alpha = 0.05$ and 0.10.

Number of populations k	Sample sizes limits n_i	W				L			
		Dispersion parameters of the Wiener processes							
		$\sigma_i = 1$		$\sigma_i \sim U(1,3)$		$\sigma_i = 1$		$\sigma_i \sim U(1,3)$	
		5%	10%	5%	10%	5%	10%	5%	10%
3	5–10	.0769	.1064	.0758	.1071	.0417	.0779	.0458	.0788
	11–20	.0577	.0880	.0623	.0901	.0378	.0732	.0381	.0734
	21–30	.0604	.0894	.0563	.0817	.0388	.0722	.0359	.0688
	31–40	.0556	.0802	.0515	.0738	.0383	.0722	.0343	.0664
	41–50	.0543	.0804	.0580	.0833	.0406	.0756	.0334	.0633
	51–60	.0568	.0831	.0544	.0797	.0374	.0737	.0333	.0627
5	5–10	.0535	.0969	.0800	.1142	.0535	.0960	.0573	.0955
	11–20	.0679	.1031	.0705	.0996	.0453	.0851	.0516	.0912
	21–30	.0680	.1000	.0674	.1007	.0477	.0855	.0521	.0908
	31–40	.0661	.0920	.0627	.0941	.0499	.0921	.0445	.0842
	41–50	.0673	.1023	.0607	.0912	.0450	.0853	.0452	.0823
	51–60	.0620	.0895	.0634	.0924	.0456	.0858	.0444	.0816

where $G_1, \ldots, G_k$ are zero mean, independent Gaussian processes on L^2, with covariance functions $\tau_1 c_1(t,s), \ldots, \tau_k c_k(t,s)$. Therefore, the asymptotic null distribution of T_k differs in each setting. Nevertheless, one may wonder if the test, which has been specifically designed for large k, gives tolerable results in the classical framework. In order to study this point numerically, we have repeated the simulation experiments in Table 5 for small k ($k = 3, 5$) and several sample sizes, based on 10,000 simulated data sets. In each case, the test rejects H_0 when $\sqrt{k}T_k/\hat{\sigma}_{0k} > z_{1-\alpha}$. Table 7 displays the results obtained. In view of these outcomes, we conclude that the actual levels are quite reasonable, specially for $k = 5$.

8. Data with finite dimension

So far we have assumed that the available data consist of functional data. A close inspection of the developments reveals that all stated results remain valid whenever the data have finite dimension by defining adequately all involved norms and operators as follows. Assume that the data take values in $\mathbb{R}^d$, where we consider the usual scalar product and the Euclidean norm. In this setting, each $c_i = (c_{i,j,r})_{1\le j,r\le d}$ is the covariance matrix of X_i, which is a $d\times d$-matrix, $c_i \in \mathcal{M}_{d\times d}$, $1 \le i \le k$, and the covariance operator is the usual product of the covariance matrix by a vector. If c_i has eigenvalues $\lambda_{i1}, \ldots, \lambda_{id}$, then

$$\theta_i = \sum_{j=1}^{d} \lambda_{ij}^2 = \mathrm{trace}(c_i c_i) = \sum_{j,r=1}^{d} c_{i,j,r}^2$$

and $\gamma_i = (\mu_i - \bar{\mu}_.)^\top c_i(\mu_i - \bar{\mu}_.)$. We will not reformulate all previous results, which keep on being true mutatis mutandis.

For the special case of $d = 1$, and assuming that the data are normally distributed, Park and Park [25] have proposed two tests, that will be denoted by $\mathcal{T}_1$ and $\mathcal{T}_2$, whose associated statistics, conveniently normalized, are also asymptotically normal. The statistic of test $\mathcal{T}_1$ is closely related to T_k: the numerator is the same for both statistics, only differing in their denominators, being in both cases the square root of an estimator of the variance of the numerator. In our proposal, the variance estimator does not assume any parametric model, while the variance estimator used in [25], heavily relies on the normality assumption. Now, naturally two questions arise:

(a) Since the tests in [25] are built by assuming normal populations, one may wonder if those tests still work for non-normal data.
(b) If the data were normally distributed, it should be expected that test $\mathcal{T}_1$ had a better behavior than our proposal, since the former incorporates this information (the normality of the data) in its construction. In this setting, one may wonder if the loss due to use the test based on T_k is considerable or by contrast it is negligible.

To numerically investigate these two questions, we have carried out two simulation experiments. In order to investigate question (a) we have generated data with equal means, that is under H_0, for normal data and non-normal data. Specifically, for each scenario we generated 10,000 samples of data coming from homoscedastic normal populations with equal mean, heteroscedastic normal populations with equal mean and negative exponential populations with equal mean. In all cases the sizes of the samples from each population were generated from a discrete uniform random law $UD\{8, \ldots, 20\}$, because the practical application of test $\mathcal{T}_2$ requires sample sizes greater than or equal to 8. Table 8 reports the fractions of p-values less than or equal to 0.05 and 0.10, which are the estimated type I error probabilities for nominal significance level α=0.05 and 0.10, respectively. The results for the proposal in this paper are headed by $\mathcal{T}$. Looking at this table we see that in

Table 8
Observed proportion of rejections in 10,000 simulated data sets for each scenario, for univariate data, with sample sizes $n_i \sim UD\{8, \dots, 20\}$, under the null hypothesis. The proposed test, denoted by $\mathcal{T}$, is compared in terms of the estimated type I error to the two tests proposed by Park and Park [25], denoted by $\mathcal{T}_1$ and $\mathcal{T}_2$. The nominal significance levels are $\alpha = 0.05$ and 0.10.

Number of populations k		$N(0,\sigma^2)$						$Exp(1)$		
		$\sigma_i = 1$ $\forall i$			$\sigma_i \sim U(1,3)$					
		$\mathcal{T}$	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}$	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}$	$\mathcal{T}_1$	$\mathcal{T}_2$
30	5%	.0657	.0687	.0596	.0664	.0691	.0633	.0564	.0434	.3993
	10%	.1105	.1141	.1005	.1056	.1080	.1049	.0981	.0806	.4801
40	5%	.0636	.0655	.0609	.0638	.0659	.0619	.0527	.0396	.4700
	10%	.1078	.1099	.0994	.1050	.1067	.1057	.0984	.0798	.5472
50	5%	.0620	.0633	.0605	.0663	.0676	.0607	.0518	.0372	.5378
	10%	.1065	.1086	.1035	.1124	.1149	.1050	.0997	.0768	.6188
100	5%	.0513	.0522	.0568	.0606	.0614	.0573	.0539	.0371	.7419
	10%	.1028	.1039	.1011	.1052	.1058	.1054	.1011	.0759	.8091
200	5%	.0528	.0533	.0540	.0606	.0612	.0576	.0565	.0367	.9155
	10%	.1018	.1022	.1017	.1090	.1094	.1050	.1039	.0779	.9470
300	5%	.0547	.0549	.0532	.0559	.0561	.0566	.0532	.0337	.9761
	10%	.1083	.1087	.1025	.1028	.1031	.1016	.1003	.0734	.9863
400	5%	.0551	.0554	.0537	.0565	.0566	.0548	.0506	.0319	.9940
	10%	.1015	.1021	.1005	.1055	.1060	.1030	.1008	.0704	.9965
500	5%	.0538	.0542	.0534	.0515	.0516	.0499	.0505	.0321	.9974
	10%	.1025	.1022	.0977	.1002	.1009	.0966	.1004	.0693	.9985

Table 9
Observed proportion of rejections in 10,000 simulated data sets for each scenario, for univariate data, with sample sizes $n_i \sim UD\{8, \dots, 20\}$, under alternatives. The proposed test, denoted by $\mathcal{T}$, is compared in terms of the estimated power to the two tests proposed by Park and Park [25], denoted by $\mathcal{T}_1$ and $\mathcal{T}_2$. The nominal significance levels are $\alpha = 0.05$ and 0.10.

Number of populations k		Dispersion parameters								
		$\sigma_i = 1$ $\forall i$			$\sigma_i = 3$ $\forall i$			$\sigma_i \sim U(1,3)$		
		$\mathcal{T}$	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}$	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}$	$\mathcal{T}_1$	$\mathcal{T}_2$
30	5%	.7757	.7809	.7311	.2036	.2102	.2940	.1900	.1965	.2830
	10%	.8368	.8401	.7982	.2876	.2947	.3839	.2714	.2777	.3707
40	5%	.8560	.8589	.8247	.1354	.1397	.1277	.2145	.2199	.3217
	10%	.9018	.9038	.8781	.2112	.2153	.1971	.3070	.3134	.4183
50	5%	.9065	.9084	.8749	.1366	.1402	.1298	.2437	.2489	.3642
	10%	.9382	.9399	.9148	.2181	.2211	.1983	.3402	.3439	.4662
100	5%	.9907	.9908	.9852	.1788	.1809	.1688	.3393	.3414	.5388
	10%	.9959	.9959	.9928	.2718	.2739	.2586	.4527	.4544	.6412
200	5%	1.000	1.000	.9998	.2582	.2592	.2413	.4983	.4998	.7559
	10%	1.000	1.000	1.000	.3648	.3661	.3465	.6276	.6286	.8393
300	5%	1.000	1.000	1.000	.3151	.3162	.2919	.6452	.6464	.8843
	10%	1.000	1.000	1.000	.4370	.4383	.4153	.7549	.7557	.9332
400	5%	1.000	1.000	1.000	.3751	.3757	.3486	.7424	.7430	.9457
	10%	1.000	1.000	1.000	.5080	.5086	.4797	.8346	.8350	.9709
500	5%	1.000	1.000	1.000	.4331	.4339	.4033	.8201	.8204	.9765
	10%	1.000	1.000	1.000	.5673	.5679	.5346	.8930	.8933	.9880

the case of normal populations the three tests behave quite closely, but for exponential data test $\mathcal{T}_1$ is very conservative (the level decreases as k increases) and test $\mathcal{T}_2$ is very liberal. The observed level of the test proposed in this paper is very close to the nominal values in all cases. It is concluded that $\mathcal{T}_1$ and $\mathcal{T}_2$ should not be used for non-normal data.

Now, we numerically explore question (b). From the previous experiment, we have learnt that, under the null hypothesis and for normal data, the proposal in this paper and the tests in [25] perform very closely. We have also learnt that it is not advisable to apply the tests in [25] for non-normal data (since the actual levels are far apart from the nominal values). So, to compare the powers we must restrict to the case of normal populations with different means. Therefore, to numerically compare the powers we generated samples from homoscedastic (with $\sigma_i = 1$ and $\sigma_i = 3$, $\forall i$) and heteroscedastic ($\sigma_i \sim U(1, 3)$) normal populations, with the 80% of the populations with mean equal to 0 and the mean of the other 20% randomly generated from a law $U(0, 1)$. As in the previous experiment, 10,000 samples were generated for each scenario. Table 9 displays the results obtained. Looking at this table, we see that $\mathcal{T}$ and $\mathcal{T}_1$ have very close power, so it seems that there is no advantage in using a variance estimator relying on the normality assumption. In some cases $\mathcal{T}$ and $\mathcal{T}_1$ outperform $\mathcal{T}_2$, while in other cases the opposite is observed.

Before ending this section, we summarize the results of a real data set application. Specifically, we applied the new test to compare the mean of the number of births per month, relative to the number of women, across the 52 Spanish provinces in 2019. So we have $k = 52$ populations. The data were taken from the website of the Spanish National Institute of Statistics, *http://www.ine.es*. For each province, we took as sample the observed number of births each month in 2019,

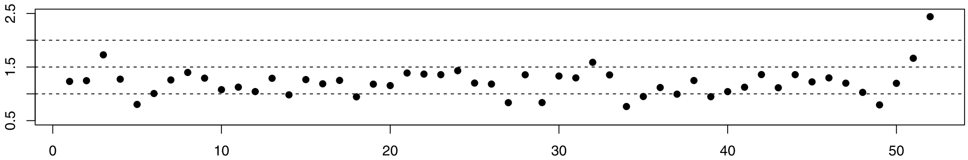

Fig. 3. 1000 times the sample mean of the number of births per month, relative to the number of woman, for each province in Spain in 2019.

over the number of women in the province at the beginning of that year, therefore $n_1 = n_2 = \cdots = n_{52} = 12$. We graphically checked that there is no seasonality (the month does not have an effect on the number of births), so we can consider the data as i.i.d. The value of the test statistic for this data set is $T_k = 227.1637$, which clearly led us to reject the null hypothesis that the mean of the number of births per month, relative to the number of woman, is constant along the 52 Spanish provinces. Fig. 3 displays 1000 times the sample mean of the number of births per month, relative to the number of woman, for each province. Looking at this figure one can see that there are some provinces with a small relative number of births, most of these quantities are between 1 and 1.5, and there are three provinces having a relative high number of births, specially province 52.

9. Concluding remarks

This paper proposes and studies a test for the equality of the means of a large number of populations. No parametric assumption is made on the populations and the procedure can be applied to functional data and to finite dimensional data. The test statistic converges under the null hypothesis to a standard normal distribution, so the critical points are available. When it was numerically compared with other tests specifically designed for normal univariate data, our proposal gave very close results.

10. Proofs

This section sketches the proofs of the results stated in the previous sections.

Proof of Lemma 1. Because $L_{1k}, \dots, L_{kk}$ are independent random variables, it follows that $var(T_{K,Lin}) = \frac{1}{k^2}\sum_{i=1}^k var(L_{ik})$. To calculate $var(L_{ik})$ we rewrite L_{ik} as follows,

$$L_{ik} = \frac{1}{n_i(n_i-1)} \sum_{1\le u<v\le n_i} \langle X_{iu}-\mu_i, X_{iv}-\mu_i\rangle + 2\frac{1}{n_i}\sum_{r=1}^{n_i}\langle X_{ir}-\mu_i, \mu_i-\bar{\mu}_.\rangle. \tag{10}$$

Taking into account that $var(L_{ik}) = E(L_{ik}^2)$, $E(\langle X_{iu}-\mu_i, X_{iv}-\mu_i\rangle^2) = \theta_i$, $\forall u \neq v$, $E(\langle X_{iu}-\mu_i, X_{iv}-\mu_i\rangle\langle X_{il}-\mu_i, X_{ij}-\mu_i\rangle) = 0$, $\forall 1 \le u < v \le n_i$, $1 \le j < l \le n_i$ so that $(u,v) \neq (j,l)$, $E(\langle X_{ir}-\mu_i, \mu_i-\bar{\mu}_.\rangle^2) = \gamma_i$ and $E(\langle X_{iu}-\mu_i, X_{iv}-\mu_i\rangle)(\langle X_{ir}-\mu_i, \mu_i-\bar{\mu}_.\rangle) = 0$, $1 \le u \neq v, r \le n_i$, the result follows. □

Proof of Lemma 2. Taking into account that $R_k = -\frac{T_{k,Lin}}{k} + R_1 - R_2$, with $R_1 = (2/k^2)\sum_{i=1}^k\langle \bar{X}_i - \mu_i, \mu_i - \bar{\mu}_.\rangle$, $R_2 = (2/k^2)\sum_{1\le i<l\le k}\langle \bar{X}_{i.}-\mu_i, \bar{X}_{l.}-\mu_l\rangle$, it suffices to show that $E(R_i^2)/var(T_{k,Lin}) \to 0$, $i = 1, 2$. Since $E(R_1^2) = (4/k^4)\sum_{i=1}^k \gamma_i/n_i$, it readily follows that $E(R_1^2)/var(T_{k,Lin}) \le 1/k^2 \to 0$. As for R_2,

$$E(R_2^2) = \frac{2}{k^4}\sum_{1\le i\neq l\le k}\frac{1}{n_in_l}\iint c_i(t,s)c_l(t,s)dtds \le \frac{2}{k^2}\left(\frac{1}{k}\sum_{i=1}^k \frac{\sqrt{\theta_i}}{n_i}\right)^2 \le \frac{2}{k^3}\sum_{i=1}^k\frac{\theta_i}{n_i^2}, \tag{11}$$

which implies that $E(R_2^2)/var(T_{K,Lin}) \le 1/k \to 0$, and the proof is complete. □

Proof of Proposition 1. From Corollary 1.9.3 in Serfling [27], it suffices to show that

$$\frac{1}{k^4}\sum_{i=1}^k E(L_{ik}^4) = o\left(var^2(T_{k,Lin})\right). \tag{12}$$

From (10) and taking into account that $|x+y|^r \le c_r(|x|^r + |y|^r)$, $\quad x, y \in \mathbb{R}, \quad r > 0$, with $c_r = 1$ if $0 < r \le 1$ and $c_r = 2^r$, otherwise, it follows that

$$E(L_{ik}^4) \le \tfrac{16}{n_i^4(n_i-1)^4}E\left\{\left(\sum_{1\le u\neq v\le n_i}\langle X_{iu}-\mu_i, X_{iv}-\mu_i\rangle\right)^4\right\} + \tfrac{16}{n_i^4}E\left\{\left(\sum_{u=1}^{n_i}\langle X_{iu}-\mu_i, \mu_i-\bar{\mu}_.\rangle\right)^4\right\}. \tag{13}$$

From Assumption 2,

$$E\left\{\left(\sum_{1\le u\neq v\le n_i}\langle X_{iu}-\mu_i, X_{iv}-\mu_i\rangle\right)^4\right\} \le n_i^4 M. \tag{14}$$

From (14), Assumptions 1 and 3 and taking into account that $k^2\, var(T_{k,Lin}) \ge \sum_{i=1}^k \frac{1}{n_i(n_i-1)}\theta_i$,

$$\frac{1}{k^4}\frac{\sum_{i=1}^k \frac{1}{n_i^4(n_i-1)^4} E\left\{\left(\sum_{1\le u\neq v\le n_i}\langle X_{iu}-\mu_i, X_{iv}-\mu_i\rangle\right)^4\right\}}{var^2(T_{k,Lin})} \le M\frac{n_{max}^4}{n_{min}^4}\frac{1}{k} = o(1). \tag{15}$$

We have that

$$E\left\{\left(\sum_{u=1}^{n_i}\langle X_{iu}-\mu_i, \mu_i-\bar{\mu}_{\cdot}\rangle\right)^4\right\} = n_i E\left(\langle X_i-\mu_i, \mu_i-\bar{\mu}_{\cdot}\rangle^4\right) + 3n_i(n_i-1)E^2\left(\langle X_i-\mu_i, \mu_i-\bar{\mu}_{\cdot}\rangle^2\right)$$
$$\le 4n_i^2 E\left(\langle X_i-\mu_i, \mu_i-\bar{\mu}_{\cdot}\rangle^4\right). \tag{16}$$

From (16) and Assumption 4,

$$\frac{1}{k^4}\frac{\sum_{i=1}^k \frac{1}{n_i^4} E\left\{\left(\sum_{u=1}^{n_i}\langle X_{iu}-\mu_i, \mu_i-\bar{\mu}_{\cdot}\rangle\right)^4\right\}}{var^2(T_{k,Lin})} \le 4\frac{1}{k}\frac{\frac{1}{k}\sum_{i=1}^k E(\langle X_i-\mu_i, \mu_i-\bar{\mu}_{\cdot}\rangle^4)/n_i^2}{\left(\frac{1}{k}\sum_{i=1}^k \gamma_i/n_i\right)^2} = o(1). \tag{17}$$

Finally, (13), (15) and (17) imply (12), and hence the result is proven. □

Proof of Proposition 2. Let $\varepsilon > 0$, by Markov inequality,

$$P\left(|\hat{\sigma}_{0k}^2-\sigma_{0k}^2| > \sigma_{0k}^2\varepsilon\right) \le E\{(\hat{\sigma}_{0k}^2-\sigma_{0k}^2)^2\}/\varepsilon^2\sigma_{0k}^2. \tag{18}$$

Since $E(\hat{\theta}_i^2)=\theta_i^2$, $1\le i\le k$, we have that

$$E\{(\hat{\sigma}_{0k}^2-\sigma_{0k}^2)^2\} = (1/k^2)\sum_{i=1}^k E\{(\hat{\theta}_i-\theta_i)^2\}/n_i^2(n_i-1)^2. \tag{19}$$

From Assumption 2, routine calculations show that

$$E\{(\hat{\theta}_i-\theta_i)^2\} \le M, \quad \forall i. \tag{20}$$

From (18)–(20) and Assumption 1, we get that

$$P\left(\frac{|\hat{\sigma}_{0k}^2-\sigma_{0k}^2|}{\sigma_{0k}^2} > \varepsilon\right) \le \frac{M}{\varepsilon^2}\frac{n_{max}^4}{n_{min}^4}\frac{1}{k}, \tag{21}$$

From Assumption 3, the right-hand side of (21) is o(1), which implies the result. □

Proof of the consistency $\hat{\theta}$ in the homoscedastic case. We have that

$$|\hat{\theta}-\theta| = \iint |\hat{c}(t,s)-c(t,s)||\hat{c}(t,s)-c(t,s)|dtds.$$

Therefore

$$E\left(|\hat{\theta}-\theta|\right) \le \iint E^{1/2}[\{\hat{c}(t,s)-c(t,s)\}^2]E^{1/2}[\{\hat{c}(t,s)+c(t,s)\}^2]dtds.$$

Under the stated assumptions, routine calculations show that $E[\{\hat{c}(t,s)-c(t,s)\}^2] \le C/k$ and $E[\{\hat{c}(t,s)+c(t,s)\}^2] \le C$, for all $0\le t,s\le 1$, for a positive constant C, which implies that $E\left(|\hat{\theta}-\theta|\right) \to 0$, and hence the consistency of $\hat{\theta}$ as an estimator of θ. □

Proof of Proposition 3. We will show it by reduction to absurdity.

Suppose that $\frac{1}{\sigma_{0k}^2}\frac{1}{k}\sum_{i=1}^k \frac{\gamma_i}{n_i} \nrightarrow 0$ which is equivalent to

$$\exists \varepsilon_1 > 0 \text{ such that } \varepsilon_1 < \frac{1}{\sigma_{0k}^2}\frac{1}{k}\sum_{i=1}^k \frac{\gamma_i}{n_i} \text{ for an infinite number of values of } k. \tag{22}$$

By applying the Cauchy–Schwarz inequality and Assumption 2, one gets

$$\frac{1}{\sigma_{0k}^2}\frac{1}{k}\sum_{i=1}^{k}\frac{\gamma_i}{n_i} \le M\frac{D_k}{\sigma_{0k}^2}. \tag{23}$$

From (22) and (23),

$$0 < \varepsilon_2 = \frac{\varepsilon_1}{M} \le \frac{D_k}{\sigma_{0k}^2}, \quad \text{for an infinite number of values of } k. \tag{24}$$

From Assumption 1 it follows that

$$\sigma_{0k}^2 \ge \tau/n_{max}^2 \tag{25}$$

Let $\varepsilon_3 > 0$ be arbitrary but fixed. We are assuming that $\sqrt{k}\frac{D_k}{\sigma_k} \to \delta \in [0, \infty)$, which implies that

$$\exists k_0 = k_0(\varepsilon_3) \text{ such that } \sqrt{k}\frac{D_k}{\sigma_k} \le \delta + \varepsilon_3, \quad \forall k \ge k_0. \tag{26}$$

From (23), (25) and (26),

$$M\left(\frac{k}{n_{max}^2}\right)^{1/2}\frac{D_k/\sigma_{0k}^2}{\sqrt{1 + D_k/\sigma_{0k}^2}} \le \delta + \varepsilon_3, \quad \forall k \ge k_0. \tag{27}$$

Taking into account that the function $f(x) = x/\sqrt{1+x}$ is increasing $\forall x \ge 0$, from (24) it follows that

$$\frac{D_k/\sigma_{0k}^2}{\sqrt{1 + D_k/\sigma_{0k}^2}} \ge \frac{\varepsilon_2}{\sqrt{1+\varepsilon_2}}, \quad \text{for an infinite number of values of } k. \tag{28}$$

From (27) and (28) it follows that $M(k/n_{max}^2)^{1/2} \le \delta + \varepsilon_3$, for an infinite number of values of k, which contradicts (9), implying that $\frac{1}{\sigma_{0k}^2}\frac{1}{k}\sum_{i=1}^{k}\frac{\gamma_i}{n_i} \to 0$. $\square$

Proof of Theorems 3 and 4. Their proofs are parallel to that of Lemmas 1 and 2. To save space we omit it. $\square$

Acknowledgments

The authors thank two anonymous referees for their valuable time and careful comments, which improved the presentation of this paper. The authors have been partially supported by grant MTM2017-89422-P (Spanish Ministry of Economy, Industry and Competitiveness, the State Agency of Investigation, the European Regional Development Fund). MD Jiménez-Gamero was also partially supported by grant P18-FR-2369 (Junta de Andalucía, Spain).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2021.104778.

References

[1] M.G. Akritas, S. Arnold, Asymptotics for analysis of variance when the number of levels is large, J. Amer. Statist. Assoc. 95 (2000) 212–226.
[2] M.G. Akritas, N. Papadatos, Heteroscedastic one-way ANOVA and lack-of-fit tests, J. Amer. Statist. Assoc. 99 (2004) 368–382.
[3] Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency, Ann. Statist. 29 (2001) 1165–1188.
[4] M. Benko, W. Härdle, A. Kneip, Common functional principal components, Ann. Statist. 37 (2009) 1–34.
[5] D.D. Boos, C. Brownie, ANOVA and rank tests when the number of treatments is large, Statist. Probab. Lett. 23 (1995) 183–191.
[6] M. Cousido-Rocha, J. de Uña Álvarez, J.D. Hart, Testing equality of a large number of densities under mixing conditions, Test 28 (2019) 1203–1228.
[7] J.A. Cuesta-Albertos, E. del Barrio, R. Fraiman, C. Matrán, The random projection method in goodness of fit for functional data, Comput. Statist. Data Anal. 51 (2007) 4814–4831.
[8] J.A. Cuesta-Albertos, M. Febrero-Bande, A simple multiway ANOVA for functional data, Test 19 (2010) 537–557.
[9] A. Cuevas, M. Febrero, R. Fraiman, An ANOVA test for functional data, Comput. Statist. Data Anal. 47 (2004) 111–122.
[10] B. Efron, C. Stein, The jackknife estimate of variance, Ann. Statist. 9 (1981) 586–596.
[11] A. Ghiglietti, F. Ieva, A.M. Paganoni, Statistical inference for stochastic processes: two-sample hypothesis tests, J. Statist. Plann. Inference 180 (2017) 49–68.
[12] T. Górecki, S. Hörmann, L. Horváth, P. Kokoszka, Testing normality of functional time series, J. Time Anal. 39 (2018) 471–487.
[13] T. Górecki, L. Smaga, mfds: Multivariate Functional Data Sets. R package version 0.1.0, 2017, URL https://github.com/Halmaris/mfds.
[14] T. Górecki, L. Smaga, fdANOVA: an R software package for analysis of variance for univariate and multivariate functional data, Comput. Statist. 34 (2019) 571–597.
[15] A.K. Gupta, S.W. Harrar, Y. Fujikoshi, MANOVA for large hypothesis degrees of freedom under non-normality, Test 17 (2008) 120–137.

[16] S.W. Harrar, A.C. Bathke, A modified two-factor multivariate analysis of variance: asymptotics and small sample approximations, Ann. Inst. Statist. Math. 64 (2012) 135–165.
[17] S.W. Harrar, A.K. Gupta, Asymptotic expansion for the null distribution of the F-statistic in one-way ANOVA under non-normality, Ann. Inst. Statist. Math. 59 (2007) 531–556.
[18] N. Henze, M.D. Jiménez-Gamero, A test for Gaussianity in Hilbert spaces via the empirical characteristic functional, Scand. J. Stat. (2020) http://dx.doi.org/10.1111/sjos.12470.
[19] L. Horváth, P. Kokoszka, Inference for Functional Data with Applications, Springer, New York, 2012.
[20] L. Horváth, G. Rice, An introduction to functional data analysis and a principal component approach for testing the equality of mean curves, Rev. Mat. Complut. 28 (2015) 505–548.
[21] Q. Jiang, M. Hušková, S.G. Meintanis, L. Zhu, Asymptotics, finite-sample comparisons and applications for two-sample tests with functional data, J. Multivariate Anal. 170 (2019) 202–220.
[22] M.D. Jiménez-Gamero, M. Cousido-Rocha, M.V. Alba-Fernández, F. Jiménez-Jiménez, Testing the equality of a large number of populations, TEST (2021) http://dx.doi.org/10.1007/s11749-021-00769-9.
[23] M.W. Kadous, C. Sammut, Classification of multivariate time series and structured data using constructive induction, Mach. Learn. 58 (2005) 179–216.
[24] E. Paparoditis, T. Sapatinas, Bootstrap-based testing of equality of mean functions or equality of covariance operators for functional data, Biometrika 103 (2016) 727–733.
[25] J. Park, D. Park, Testing the equality of a large number of normal population means, Comput. Statist. Data Anal. 56 (2012) 1131–1149.
[26] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2020, URL https://www.R-project.org/.
[27] R.J. Serfling, Approximation Theorems of Mathematical Statistics, Wiley, New York, 1980.
[28] A. Yuan, H.B. Fang, H. Li, C.O. Wu, M.T. Tan, Hypothesis testing for multiple mean and correlation curves with functional data, Statist. Sinica 30 (2020) 1095–1116.
[29] D. Zhan, J.D. Hart, Testing equality of a large number of densities, Biometrika 101 (2014) 449–464.
[30] J.T. Zhang, Analysis of Variance for Functional Data, Chapman & Hall, London, 2013.
[31] J.T. Zhang, M.Y. Cheng, H.T. Wu, B. Zhou, A new test for functional one-way ANOVA with applications to ischemic heart screening, Comput. Statist. Data Anal. 132 (2019) 3–17.
[32] J.T. Zhang, X. Liang, One-way ANOVA for functional data via globalizing the pointwise F-test, Scand. J. Stat. 41 (2014) 51–71.
[33] C. Zhang, H. Peng, J.T. Zhang, Two samples tests for functional data, Commun. Stat. - Theory Methods 39 (2010) 559–578.