ORIGINAL ARTICLE

Application of distance standard deviation in functional data analysis

Mirosław Krzyśko[1] · Łukasz Smaga[2]

Abstract

This paper concerns the measurement and testing of equality of variability of functional data. We apply the distance standard deviation constructed based on distance correlation, which was recently introduced as a measure of spread. For functional data, the distance standard deviation seems to measure different kinds of variability, not only scale differences. Moreover, the distance standard deviation is just one real number, and for this reason, it is of more practical value than the covariance function, which is a more difficult object to interpret. For testing equality of variability in two groups, we propose a permutation method based on centered observations, which controls the type I error level much better than the standard permutation method. We also consider the applicability of other correlations to measure the variability of functional data. The finite sample properties of two-sample tests are investigated in extensive simulation studies. We also illustrate their use in five real data examples based on various data sets.

Keywords Distance correlation · Functional data analysis · Projection correlation · Standard deviation · Variability measure

Mathematics Subject Classification 62R10 · 62H15

✉ Łukasz Smaga
 ls@amu.edu.pl

 Mirosław Krzyśko
 mkrzysko@amu.edu.pl

[1] Interfaculty Institute of Mathematics and Statistics, Calisia University-Kalisz, Nowy Świat 4, Kalisz 62-800, Poland

[2] Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Uniwersytetu Poznańskiego 4, Poznań 61-614, Poland

 ⌂ Springer

1 Introduction

Székely et al (2007) and Székely and Rizzo (2009) introduced distance covariance and correlation into the statistical literature. These quantities are useful for measuring the association between and testing the independence of multivariate data sets. In the last few years, there has been great interest in the theory and applications of distance correlation. Theoretical extensions were investigated in, for example, Berrett and Samworth (2019), Dueck et al (2015), Fokianos and Pitsillou (2017), Gretton et al (2005), Gretton et al (2008), Lyons (2013), Pan et al (2018), Székely and Rizzo (2013, 2014), Zhu et al (2017), and Zhu et al (2020). These papers concern various issues, such as dependence measures, statistical inference, time series analysis, affinely invariant distance correlation, and metric spaces. Some of the important applications of distance correlation are feature screening (Li et al 2012), detection of long-range concerted motions in protein (Roy and Post 2012), and assessment of associations of familial relationships, lifestyle factors, diseases, and mortality (Kong et al 2012). An efficient implementation of the statistical inference methods based on distance covariance is provided by the R packages energy and dcortools (R Core Team 2022; Rizzo and Székely 2021; Edelmann and Fiedler 2022) and the Python package dcor (Ramos-Carreño, 2022).

Distance covariance is a measure of dependence between two random vectors of arbitrary dimensions, defined as

$$dcov^2(\mathbf{X}, \mathbf{Y}) = \int_{\mathbb{R}^{p+q}} \frac{\|\varphi_{\mathbf{X}, \mathbf{Y}}(\mathbf{t}, \mathbf{s}) - \varphi_{\mathbf{X}}(\mathbf{t})\varphi_{\mathbf{Y}}(\mathbf{s})\|^2}{c_p c_q \|\mathbf{t}\|^{1+p} \|\mathbf{s}\|^{1+q}} d\mathbf{t} d\mathbf{s},$$

where $\varphi_{\mathbf{X}, \mathbf{Y}}$ is the joint characteristic function of $(\mathbf{X}^\top, \mathbf{Y}^\top)^\top$, $\varphi_{\mathbf{X}}$ and $\varphi_{\mathbf{Y}}$ are the characteristic functions of $\mathbf{X} \in \mathbb{R}^p$ and $\mathbf{Y} \in \mathbb{R}^q$ respectively, $\| \cdot \|$ is the complex Euclidean norm, and $c_b = \pi^{(1+b)/2}/\Gamma((1+b)/2)$. Using the notation of Edelmann et al (2020), for a single random vector $\mathbf{X}$, we can define the distance variance as

$$dvar(\mathbf{X}) = \int_{\mathbb{R}^{2p}} \frac{\|\varphi_{\mathbf{X}}(\mathbf{t} + \mathbf{s}) - \varphi_{\mathbf{X}}(\mathbf{t})\varphi_{\mathbf{X}}(\mathbf{s})\|^2}{c_p^2 \|\mathbf{t}\|^{1+p} \|\mathbf{s}\|^{1+p}} d\mathbf{t} d\mathbf{s}.$$

Then, the distance standard deviation is defined as the nonnegative square root of $dvar(\mathbf{X})$. We denote it as $dsd(\mathbf{X})$ or simply dsd. For measuring the amount of dependence, one can use the distance correlation coefficient of the form

$$\mathcal{R}(\mathbf{X}, \mathbf{Y}) = \begin{cases} \frac{dcov(\mathbf{X}, \mathbf{Y})}{\sqrt{dsd(\mathbf{X})dsd(\mathbf{Y})}}, & \text{if } dsd(\mathbf{X}) \neq 0, dsd(\mathbf{Y}) \neq 0, \\ 0, & \text{if } dsd(\mathbf{X}) = 0 \text{ or } dsd(\mathbf{Y}) = 0. \end{cases}$$

It is well known that $\mathcal{R}(\mathbf{X}, \mathbf{Y}) = 0$ is equivalent to the mutual independence of $\mathbf{X}$ and $\mathbf{Y}$. Moreover, $\mathcal{R}(\mathbf{X}, \mathbf{Y}) \in [0, 1]$, and for one-dimensional random variables, we have $\mathcal{R}(X, Y) = 1$ if and only if Y is a linear function of X, almost surely. Other coefficients that are useful for testing the independence of random vectors are the Hilbert-Schmidt Independence Criterion (HSIC) (Gretton et al 2005, 2008), ball covariance (Pan et al

2018), mutual information (Berrett and Samworth 2019), and projection correlation (Zhu et al 2017). In the following sections we will also discuss the applicability of these to the topic of this paper.

Recently, Edelmann et al (2020) considered the distance standard deviation as measure of scale, which can be useful especially for heavy-tailed distributions. In the univariate case, they proved that the distance standard deviation is a measure of spread in the axiomatic sense of Bickel and Lehmann (2012), i.e., it satisfies the following conditions:

(C1) $dsd(X) \geq 0$,
(C2) $dsd(a + bX) = |b|dsd(X)$ for all $a, b \in \mathbb{R}$,
(C3) $dsd(X) \leq dsd(Y)$ if for all $0 < \alpha \leq \beta < 1$,

$$F^{-1}(\beta) - F^{-1}(\alpha) \leq G^{-1}(\beta) - G^{-1}(\alpha),$$

where F and G are the cumulative distribution functions of X and Y respectively, and F^{-1} and G^{-1} are the corresponding right-continuous inverses.

The conditions (C1) and (C2) are also satisfied for multivariate data. Edelmann et al (2020) also constructed a two-sample test for testing equality of distance standard deviations. In comparison with tests based on the standard deviation and Gini's mean difference for comparing scales, the dsd-based test has superior performance when the underlying distributions are heavy-tailed. Moreover, the distance standard deviation is defined for all random variables with finite first moments, while finite second moments are required for the classical standard deviation. Finally, Edelmann et al (2020) also showed the applicability of the distance standard deviation in multivariate statistical quality control, comparing it with the generalized variance.

However, some issues were not directly considered by Edelmann et al (2020). First of all, they mainly considered the univariate case, except in the application to multivariate statistical quality control, where the two-dimensional case was investigated in the simulation studies. Secondly, their tests were based on the asymptotic distribution or standard permutation procedure. These approaches have some disadvantages. For these reasons, this paper considers the application of distance standard deviation to specific high-dimensional data, namely functional data, and modification of the standard permutation method. This modified permutation procedure seems to avoid the problem with the standard approach.

In many practical tasks, great advances in computational and analytical techniques result in high-dimensional measurements. Such measurements are often observed repeatedly at different time or space points. For simplicity, we call them time points or design time points, regardless of what they refer to. Such data are realizations of some random process. In many applications, it is convenient to treat them not as a sequence of single measurements taken one after another, but as whole functional entities, e.g., functions, curves, surfaces, or images. Such data are called functional data, and their analysis is referred to as functional data analysis (FDA). This is a relatively new branch of statistics, which offers a powerful set of methodologies for analyzing complex data structures. The following books and review papers contain a good overview of the main FDA methods and their applications: Cuevas (2014), Ferraty and Vieu (2006),

Horváth and Kokoszka (2012), Ramsay et al (2009), Ramsay and Silverman (2002), Ramsay and Silverman (2005), Jacques and Preda (2014), Wang et al (2015), Zhang (2013). The parametric and nonparametric methods concern change point detection, classification, cluster analysis, dimension reduction, hypothesis testing, regression analysis, and times series analysis. Some examples of functional data are as follows: temperature or precipitation in a given location over some time, environmental diurnal ozone and NO_x cycles, monitoring of water quality, cash flows in finance, fetal heart rate monitoring, and the angle formed at the right elbow between the upper and lower arms of a driver measured three times for each of 20 locations within a test car. Other specific examples are presented in Sect. 4.

In this paper, we consider measurement of the variability of functional data. There are many studies on testing the equality of mean functions for two or several groups (see, for example, Górecki and Smaga 2019; Zhang 2013, and the references therein). The mean function is quite easy to visualize and interpret. However, what parameter can be used to distinguish between groups of functional observations when the mean functions cannot be found to be significantly different? Naturally, there is a covariance function, which describes the dependence structure of functional data. The test procedures proposed by Guo et al (2018) and Guo et al (2019), among others, verify the equality of several covariance functions. However, the covariance function is more difficult to interpret. Moreover, it is a function, while practitioners like to have just one real number, which is much easier to interpret. For these reasons, we apply the distance standard deviation for measuring and testing the equality of variability of functional data. In contrast to the covariance function, the interpretation of distance standard deviation is easy, namely the larger the distance standard deviation the larger variability. This is confirmed in simulation studies and real data examples of Sects. 3-4. In practice, the functional observation has the form of a time series, which can be interpreted as a multidimensional random vector, perhaps with highly correlated variables. For such data, the distance standard deviation is correctly defined and the statistical methods dedicated to it can be used. In particular, we study the testing of hypotheses about the equality of variability of two groups of functional data. We consider the tests proposed by Edelmann et al (2020), but due to their disadvantages, we propose a new permutation method based on centered observations. The results of intensive simulation studies and real data examples suggest that the new tests based on the distance standard deviation exhibit good performance. In this way, we extend the results of Edelmann et al (2020) and show their applicability to comparing the variability of functional data. We also attempted to construct "new" standard deviations based on HSIC, ball covariance, mutual information, and projection correlation, but this is not a clear issue, as will also be discussed in the following sections.

The remainder of the paper is organized as follows. In Sect. 2, we present the methodology. In particular, we consider the estimation of distance standard deviation, statistical tests for its equality in the two-sample problem, and the new permutation method. Section 3 presents an investigation of the finite sample properties of tests applied to functional data in simulation studies. In Sect. 4, we consider five different real data examples for illustrative purposes. Section 5 concludes the paper.

2 Methodology

In this section, we formulate the problem, statistical hypotheses, and test procedures which can be used for testing the equality of variability of functional data.

Let X_1 and X_2 be two independent random processes defined on the interval $[a, b]$, where $a, b \in \mathbb{R}$ and $a < b$. These processes represent functional variables. We wish to test the equality of variability of the processes X_1 and X_2. The question arises of how to define the variability of functional variables. In this paper, we propose to use the distance standard deviation. Assume that we have two independent random samples drawn from these variables. Denote them by

$$X_{11}, \ldots, X_{1n_1} \text{ and } X_{21}, \ldots, X_{2n_2}.$$

In practice, they are observed in a discrete way at certain design time points. Let

$$X_{ijk} = X_{ij}(t_k)$$

for the design time points $t_k \in [a, b]$, $i = 1, 2$, $j = 1, \ldots, n_i$, $k = 1, \ldots, K$. Thus, we have two random samples

$$\mathbf{X}_{ij} = (X_{ij1}, \ldots, X_{ijK})^\top$$

from certain random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ in the space $\mathbb{R}^K$, which can be high-dimensional. For testing the variability of the processes X_1 and X_2, we propose to test the following null hypothesis:

$$H_0 : dsd(\mathbf{X}_1) = dsd(\mathbf{X}_2). \tag{1}$$

For this purpose, we use the test statistic of Edelmann et al (2020) and a new permutation method, which are described in detail below. Therefore, we consider a simple procedure which applies the test for random vectors to discrete functional observations.

Remark 1 Note that in many FDA methods, the discrete functional data are transformed to continuous functions of time using some smoothing method. For the purposes of this paper, we also considered smoothing by the basis representation (Ramsay and Silverman 2005, Chapter 3). However, because the results were equally good as those for the approach presented below, we decided not to use smoothing. Therefore, although we use discrete functional data as multivariate observations, we can consider the functional nature of functional data. For example, in case of missing values or measurement errors, we can use smoothing to solve this problem, and then apply our methods to coefficients of basis representation or recalculated values of functional data in appropriate design time points.

Before we present the testing procedure, let us notice possible applications of verifying the null hypothesis (1). One of them was already mentioned in the introduction (Sect. 1), namely the application in the multivariate statistical quality control by Edelmann et al (2020). The other one is the detection of significant differences between two

populations when the mean functions are the same. The distribution does not have to be different in mean only. On the other hand, many tests for equality of mean functions assume that the variability is the same. In particular, equality of covariance functions is a common such assumption. As we show in simulation studies (see Sect. 3), the tests verifying the null hypothesis (1) can be useful for testing equality of variability being alternatives to the tests for equality of covariance functions. Finally, an important application is an issue of data integration, i.e., we investigate if two samples are part of the same larger dataset, or if they should be treated as originating from two different sources. When the tests for location (mean function) as well as for variability (distance standard deviation) do not reject the null hypotheses, one can conclude (with a certain amount of uncertainty) that the data from two samples can be combined into one. On the other hand, in case of rejecting one of the hypotheses, we suspect that the data should be separated.

Let $\mathbf{X}_i^s = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{in_i})$ be a sample from the random vector $\mathbf{X}_i$, $i = 1, 2$. For the estimation of distance standard deviation, we recall that Székely et al (2007) showed that if $\mathbf{X}_i'$ and $\mathbf{X}_i''$ are independent copies of the random vector $\mathbf{X}_i$, $i = 1, 2$, such that $E \|\mathbf{X}_i\|^2 < \infty$, then

$$dsd(\mathbf{X}_i) = \left(E \left(\|\mathbf{X}_i - \mathbf{X}_i'\|^2 \right) + \left(E \|\mathbf{X}_i - \mathbf{X}_i'\| \right)^2 - 2E \left(\|\mathbf{X}_i - \mathbf{X}_i'\| \|\mathbf{X}_i - \mathbf{X}_i''\| \right) \right)^{1/2}.$$

Hence, the estimator of $dsd(\mathbf{X}_i)$ is given by

$$\widehat{dsd}(\mathbf{X}_i^s) = \left(\frac{1}{n_i(n_i - 3)} \sum_{p=1}^{n_i} \sum_{q=1}^{n_i} \|\mathbf{X}_{ip} - \mathbf{X}_{iq}\|^2 \right.$$

$$+ \frac{1}{n_i(n_i - 1)(n_i - 2)(n_i - 3)} \left(\sum_{p=1}^{n_i} \sum_{q=1}^{n_i} \|\mathbf{X}_{ip} - \mathbf{X}_{iq}\| \right)^2$$

$$\left. - \frac{2}{n_i(n_i - 2)(n_i - 3)} \sum_{p=1}^{n_i} \sum_{q=1}^{n_i} \sum_{r=1}^{n_i} \|\mathbf{X}_{ip} - \mathbf{X}_{iq}\| \|\mathbf{X}_{ip} - \mathbf{X}_{ir}\| \right)^{1/2}.$$

Note that $(\widehat{dsd}(\mathbf{X}_i^s))^2$ is the unbiased estimator of $dvar(\mathbf{X}_i)$.

For testing the null hypothesis (1), we use the following test statistic of Edelmann et al (2020):

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\widehat{dsd}(\mathbf{X}_1^s) - \widehat{dsd}(\mathbf{X}_2^s)}{\sqrt{\frac{n_1 \hat{\sigma}^2(\mathbf{X}_1^s) + n_2 \hat{\sigma}^2(\mathbf{X}_2^s)}{n_1 + n_2}}},$$

where

$$\hat{\sigma}^2(\mathbf{X}_i^s) = (n_i - 1) \sum_{p=1}^{n_i} \left(\widehat{dsd}(\mathbf{X}_{i,-p}^s) - \frac{1}{n_i} \sum_{q=1}^{n_i} \widehat{dsd}(\mathbf{X}_{i,-q}^s) \right)^2$$

is the jackknife estimator for the variance of the asymptotic distribution of $n_i^{1/2}(\widehat{dsd}(\mathbf{X}_i^s) - dsd(\mathbf{X}_i))$, $i = 1, 2$. Here $\mathbf{X}_{i,-p}^s$ denotes the sample $\mathbf{X}_i^s$ without the pth observation. The estimator $\hat{\sigma}^2(\mathbf{X}_i^s)$ is weakly consistent (Arvesen 1969, Theorem 9). The test statistic T can be used for two-sided as well as one-sided alternative hypotheses.

To construct a test based on the statistic T, one has to approximate its null distribution. The first idea is to use the asymptotic distribution of T. Edelmann et al (2020) proved that this distribution is a normal one. Thus, they constructed an asymptotic test based on the normal approximation. In the one-dimensional case, this test performed quite well, but for small sample sizes, it may have a problem with control of the type I error level. Examining this approach, we observed that the asymptotic test often has a conservative character, resulting in a loss of power for the multivariate case. A comparison of the asymptotic dsd test with corresponding permutation procedures is presented in the supplementary online material. Its results indicate that for small or moderate sample sizes, the asymptotic testing procedure may lose some power against the permutation approaches. Nevertheless, for larger samples, the asymptotic test may be a good alternative to permutation tests, since the normal approximation should be good enough to appropriately control the type I error level and it does not require extensive computation. Since we take into account small and moderate samples, we do not consider the asymptotic test further in this paper. To avoid the problem with the asymptotic test, Edelmann et al (2020) proposed to use a permutation approach. Some recent developments relating to this approach can be found in Arboretti et al (2021), Du and Wang (2020) and Corain et al (2014). In the univariate case, the permutation test avoids the problems of the asymptotic test, but it needs a more restrictive assumption. Namely, the permutation method requires that both distributions share a common location parameter. Then the permutation distribution of the test statistic T is the same as the distribution of T, and the standard permutation test has good properties. Of course, sharing a common location parameter is not always the case, and then these distributions are not them same. In such case, we will see in Sect. 3 that the standard permutation test is too liberal or conservative in many cases. For this reason, we propose a simple modification to the classical permutation approach. We apply the permutation method, but to the centered data. Then the expected values are equal to zero, and hence we can expect better finite sample behavior. This is shown in the simulation studies in Sect. 3. The new permutation procedure is as follows:

1. Center the original data, i.e., $\mathbf{X}_{ij}^c = \mathbf{X}_{ij} - \bar{\mathbf{X}}_i$, $i = 1, 2$, $j = 1, \ldots, n_i$, where $\bar{\mathbf{X}}_i = n_i^{-1} \sum_{j=1}^{n_i} \mathbf{X}_{ij}$. We will call these observations centered data.
2. Compute T for centered data $\mathbf{X}_{ij}^c$, $i = 1, 2$, $j = 1, \ldots, n_i$. Denote the value obtained by T_{obs}.
3. Create a permutation sample from the centered data in the following way: From all observations $\mathbf{X}_{ij}^c$, $i = 1, 2$, $j = 1, \ldots, n_i$, select randomly without replacement n_1 observations for the first new sample; then the remainder of the observations is the second new sample.
4. Repeat step 3 a large number of times, e.g. $B = 1,000$, and obtain B independent permutation samples $\mathbf{X}_{ij}^{c,l}$, $i = 1, 2$, $j = 1, \ldots, n_i$, $l = 1, \ldots, B$.

5. For each permutation sample, compute the value of the test statistic T. Denote these values by $T_l, l = 1, \ldots, B$.

6. The final p-value of the permutation test for a two-sided alternative hypothesis is defined by

$$2 \min \left\{ B^{-1} \sum_{l=1}^{B} I(T_l > T_{obs}), 1 - B^{-1} \sum_{l=1}^{B} I(T_l > T_{obs}) \right\},$$

where $I(A)$ stands for the usual indicator function on a set A.

For convenience, the above test procedure will be called the centered permutation test. When we apply steps 2–6 to the original data, we have the standard permutation test.

Finally, let us comment on possible "new" standard deviations. Since the distance standard deviation is constructed based on the distance correlation, which is a dependence measure, one can consider defining a "standard deviation" based on other such measures. As we mentioned above, we consider the HSIC, ball covariance, mutual information, and projection correlation. Assume that we consider the HSIC with the Gaussian or Laplace kernels and the median method of selecting the bandwidth parameter (Gretton et al 2009). Unfortunately, none of these cases is generally as satisfactory as the distance standard deviation was. First of all, the ball covariance gives the ball standard deviation, which is constant and thus of no use at all in the sense of measuring variability. All of the other "standard deviations" satisfy the condition (C1) of Bickel and Lehmann (2012), but not the condition (C2). In fact, we have $\tau(a + bX) = \tau(X)$, where τ may be any of the three "standard deviations". Thus, we have invariance under location transformation, but these values cannot be applied to test scale differences. However, for functional data, variability seems to be a more general concept than just scale, and so it is interesting to investigate this. For this reason, we will consider tests based on projection standard deviation in the next sections. We use it for illustration, since the results for HSIC were poor, while the mutual information standard deviation leads to numerical problems that prevent its application in many cases (data not shown).

Let us briefly describe the projection correlation and standard deviation. They are based on the fact that the independence of random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ is equivalent to the independence of $U = \boldsymbol{\alpha}^\top \mathbf{X}_1$ and $V = \boldsymbol{\beta}^\top \mathbf{X}_2$ for all unit vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Assume that $F_{U,V}(u, v)$ is the joint distribution of (U, V), while $F_U(u)$ and $F_V(v)$ are the marginal distributions of U and V respectively. Then, given $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, U and V are independent if and only if

$$F_{U,V}(u, v) - F_U(u) F_V(v) = cov(I(\boldsymbol{\alpha}^\top \mathbf{X}_1 \leq u), I(\boldsymbol{\beta}^\top \mathbf{X}_2 \leq v)) = 0.$$

Thus, the independence of $\mathbf{X}_1$ and $\mathbf{X}_2$ can be tested based on the following equality:

$$\int \int \int cov^2(I(\boldsymbol{\alpha}^\top \mathbf{X}_1 \leq u), I(\boldsymbol{\beta}^\top \mathbf{X}_2 \leq v)) dF_{U,V}(u, v) d\boldsymbol{\alpha} d\boldsymbol{\beta} = 0.$$

After some analytical transformations, the squared projection covariance between $\mathbf{X}_1$ and $\mathbf{X}_2$ is defined as follows:

$$
\begin{aligned}
pcov^2(\mathbf{X}_1, \mathbf{X}_2) = \\
E\left(\arccos\left\{ \frac{(\mathbf{X}_{11} - \mathbf{X}_{13})^\top (\mathbf{X}_{14} - \mathbf{X}_{13})}{\|\mathbf{X}_{11} - \mathbf{X}_{13}\| \|\mathbf{X}_{14} - \mathbf{X}_{13}\|} \right\} \arccos\left\{ \frac{(\mathbf{X}_{21} - \mathbf{X}_{23})^\top (\mathbf{X}_{24} - \mathbf{X}_{23})}{\|\mathbf{X}_{21} - \mathbf{X}_{23}\| \|\mathbf{X}_{24} - \mathbf{X}_{23}\|} \right\} \right) \\
+ E\left(\arccos\left\{ \frac{(\mathbf{X}_{11} - \mathbf{X}_{13})^\top (\mathbf{X}_{14} - \mathbf{X}_{13})}{\|\mathbf{X}_{11} - \mathbf{X}_{13}\| \|\mathbf{X}_{14} - \mathbf{X}_{13}\|} \right\} \arccos\left\{ \frac{(\mathbf{X}_{22} - \mathbf{X}_{23})^\top (\mathbf{X}_{25} - \mathbf{X}_{23})}{\|\mathbf{X}_{22} - \mathbf{X}_{23}\| \|\mathbf{X}_{25} - \mathbf{X}_{23}\|} \right\} \right) \\
- 2E\left(\arccos\left\{ \frac{(\mathbf{X}_{11} - \mathbf{X}_{13})^\top (\mathbf{X}_{14} - \mathbf{X}_{13})}{\|\mathbf{X}_{11} - \mathbf{X}_{13}\| \|\mathbf{X}_{14} - \mathbf{X}_{13}\|} \right\} \arccos\left\{ \frac{(\mathbf{X}_{22} - \mathbf{X}_{23})^\top (\mathbf{X}_{24} - \mathbf{X}_{23})}{\|\mathbf{X}_{22} - \mathbf{X}_{23}\| \|\mathbf{X}_{24} - \mathbf{X}_{23}\|} \right\} \right).
\end{aligned}
$$

Then, the projection correlation between $\mathbf{X}_1$ and $\mathbf{X}_2$ is the square root of

$$
pcor^2(\mathbf{X}_1, \mathbf{X}_2) = \frac{pcov^2(\mathbf{X}_1, \mathbf{X}_2)}{pcov(\mathbf{X}_1, \mathbf{X}_1)\, pcov(\mathbf{X}_2, \mathbf{X}_2)},
$$

when $pcov(\mathbf{X}_i, \mathbf{X}_i) > 0$ for $i = 1, 2$, and 0 otherwise. The projection correlation belongs to $[0, 1]$, and is equal to zero if and only if the random vectors are independent. We define the projection standard deviation by

$$
psd(\mathbf{X}_i) = \sqrt{pcov(\mathbf{X}_i, \mathbf{X}_i)}, \quad i = 1, 2.
$$

Its estimator is as follows:

$$
\begin{aligned}
\widehat{psd}^2(\mathbf{X}_i) = \\
\frac{1}{n_i^3} \sum_{j,k,l=1}^{n_i} \arccos^2\left\{ \frac{(\mathbf{X}_{ij} - \mathbf{X}_{ik})^\top (\mathbf{X}_{il} - \mathbf{X}_{ik})}{\|\mathbf{X}_{ij} - \mathbf{X}_{ik}\| \|\mathbf{X}_{il} - \mathbf{X}_{ik}\|} \right\} \\
+ \frac{1}{n_i^5} \sum_{j,k,l,m,n=1}^{n_i} \arccos\left\{ \frac{(\mathbf{X}_{ij} - \mathbf{X}_{ik})^\top (\mathbf{X}_{il} - \mathbf{X}_{ik})}{\|\mathbf{X}_{ij} - \mathbf{X}_{ik}\| \|\mathbf{X}_{il} - \mathbf{X}_{ik}\|} \right\} \arccos\left\{ \frac{(\mathbf{X}_{im} - \mathbf{X}_{ik})^\top (\mathbf{X}_{in} - \mathbf{X}_{ik})}{\|\mathbf{X}_{im} - \mathbf{X}_{ik}\| \|\mathbf{X}_{in} - \mathbf{X}_{ik}\|} \right\} \\
- 2\frac{1}{n_i^4} \sum_{j,k,l,m=1}^{n_i} \arccos\left\{ \frac{(\mathbf{X}_{ij} - \mathbf{X}_{ik})^\top (\mathbf{X}_{il} - \mathbf{X}_{ik})}{\|\mathbf{X}_{ij} - \mathbf{X}_{ik}\| \|\mathbf{X}_{il} - \mathbf{X}_{ik}\|} \right\} \arccos\left\{ \frac{(\mathbf{X}_{im} - \mathbf{X}_{ik})^\top (\mathbf{X}_{il} - \mathbf{X}_{ik})}{\|\mathbf{X}_{im} - \mathbf{X}_{ik}\| \|\mathbf{X}_{il} - \mathbf{X}_{ik}\|} \right\}.
\end{aligned}
$$

For this estimator there is a much more computationally effective form, similarly as for the distance standard deviation (Zhu et al 2017, page833). To test the null hypothesis $H_0 : psd(\mathbf{X}_1) = psd(\mathbf{X}_2)$, we use tests analogous to the permutation procedures based on the T statistic.

3 Simulation study

In this section, we conduct simulation studies to investigate the type I error level and power of the considered tests. For this purpose, we study different scenarios involving the variability of functional data.

In the supplementary online material, we present additional simulation studies, in which we compare the new *dsd* and *psd* tests with the testing procedures by Guo et al (2018, 2019) for testing the equality of covariance functions in several samples, which was suggested by the referee. Although the tests verify different hypotheses, it is possible to compare them as they concern variability of the functional data, but probably in a different way. These simulations conclude that the *dsd* and *psd* tests can be powerful alternatives to the tests by Guo et al (2018, 2019) in terms of testing the global variability of functional data.

3.1 Simulation setup

For generating simulated data, we consider a model constructed based on those considered in Guo et al (2019), Kraus and Panareto (2012), and Zhang and Liang (2013). To generate discrete functional samples, we use the following model:

$$x_{ij}(t) = h_i(t)(\eta_i(t) + v_{ij}(t)), \tag{2}$$

where $t \in [0, 1]$, $i = 1, 2$, $j = 1, \ldots, n_i$. We assume that the functions $x_{ij}(t)$ are observed at fifty design time points $t_r = (r - 1)/49$ for $r = 1, \ldots, 50$. We set $n_i = 25, 40$ for $i = 1, 2$.

For the group mean functions, we consider two scenarios. In the first one, we have the same mean functions, i.e., $\eta_i(t) = 1 + 2.3t + 3.4t^2 + 1.5t^3$ for $i = 1, 2$. The second scenario has different mean functions: $\eta_1(t) = 1 + 2.3t + 3.4t^2 + 1.5t^3$ and $\eta_2(t) = \cos(2.3\pi t) + 3.4t \sin(1.5\pi t)$.

We consider

$$v_{ij}(t) = \sum_{l=1}^{m} \lambda_{il}^{1/2} y_{ijl} \psi_{il}(t),$$

where $\lambda_{il} > 0$, y_{ijl} are independent random variables with mean 0 and variance 1, and $\psi_{1l}(t) = \phi_l(t)$ for $l = 1, \ldots, m$, while $\psi_{2l}(t) = \phi_l(t)$ for $l = 1, 3, 4, \ldots, m$, and $\psi_{22}(t) = \phi_2(t) + \omega$, where $\phi_1(t) = 1$, $\phi_{2r}(t) = \sqrt{2} \sin(2\pi rt)$ and $\phi_{2r+1}(t) = \sqrt{2} \cos(2\pi rt)$, $r = 1, \ldots, (m - 1)/2$, are the orthonormal basis functions, and ω is some constant. We set $m = 11$.

We consider two distributions of the i.i.d. random variables y_{ijl}. Namely, the standard normal distribution and the standardized t-distribution with five degrees of freedom are used to generate Gaussian and non-Gaussian functional observations. Note that this t-distribution has nearly the heaviest tails among the t-distributions with finite first four moments.

The other parameters of the general model (2) are specified in the following particular models:

Model 0 We consider $\lambda_{il} = 1.5\rho^l$, $\omega = 0$, and $h_i(t) = 1$ for $i = 1, 2, l = 1, \ldots, m$, and $t \in [0, 1]$.

Table 1 Empirical sizes (in %) of all tests for Model 0

dmf	distr	n	ρ	dsd_s	psd_s	dsd_c	psd_c
n	N	25	0.1	4.9	5.5	5.3	5.6
			0.5	5.2	4.5	5.3	4.8
			0.9	4.5	5.0	4.3	5.1
		40	0.1	4.9	5.0	5.3	5.2
			0.5	5.3	5.8	5.7	6.2
			0.9	5.1	4.9	5.3	5.7
	t	25	0.1	5.6	3.9	5.7	4.2
			0.5	5.9	4.1	6.2	4.2
			0.9	5.9	5.5	5.5	5.5
		40	0.1	5.0	5.5	5.1	5.4
			0.5	5.4	4.9	5.3	5.4
			0.9	5.4	4.8	5.9	4.7
y	N	25	0.1	18.1	10.5	3.8	4.5
			0.5	7.8	8.6	4.8	5.4
			0.9	3.9	1.8	5.8	4.2
		40	0.1	14.6	9.3	4.9	5.7
			0.5	7.0	9.1	5.0	5.6
			0.9	4.7	2.1	5.5	4.4
	t	25	0.1	16.8	8.3	4.7	4.3
			0.5	7.0	9.5	4.4	5.7
			0.9	3.8	1.1	4.5	3.5
		40	0.1	13.9	9.6	3.7	5.2
			0.5	6.7	7.4	5.9	4.9
			0.9	4.9	1.6	5.2	5.3

We denote: dmf–different mean functions (n–no, y–yes); distr– distribution (N–normal, t–t-distribution); $n = n_1 = n_2$; dsd_s, psd_s (respectively dsd_c, psd_c)–standard (respectively centered) permutation tests

In this model, the variability is the same in both samples, and hence the null hypothesis is true. In the following models, we consider different alternative hypotheses. They are based on Model 0, but in the second sample:

Model 1 the observations are multiplied by 1.5.
Model 2 $h_2(t) = g_1(t) = \sin(1.5t) + 1$ or $h_2(t) = g_2(t) = \cos(1.5t) + 1$.
Model 3 $\lambda_{2l} = \lambda_{1l}^{1/2}$ for $l = 1, 4, 5, \ldots, m$, and $\lambda_{2l} = 1.5\lambda_{1l}^{1/2}$ for $l = 2, 3$.
Model 4 $\lambda_{2l} = \lambda_{1l}^{1/2}$ for $l = 1, 4, 5, \ldots, m$, while $\lambda_{22} = \lambda_{13}^{1/2}$ and $\lambda_{23} = \lambda_{12}^{1/2}$.
Model 5 $\omega = 2, 3$.

In Model 1, the second sample is obtained simply by scaling the first sample. Model 2 concerns two cases, where the second sample has greater variability for greater or smaller t. In Models 3–5, the variability of the second sample is changed internally. Example trajectories of the generated data are shown in the figures in the

Table 2 Empirical powers (in %) of all tests for Model 1

dmf	distr	n	ρ	dsd_s	psd_s	dsd_c	psd_c
n	N	25	0.1	35.8	6.4	30.1	4.4
			0.5	50.7	4.1	52.7	5.6
			0.9	95.8	2.7	96.8	4.6
		40	0.1	56.8	5.1	50.1	3.7
			0.5	71.2	3.5	73.3	4.5
			0.9	99.8	4.0	99.8	6.3
	t	25	0.1	31.7	6.4	24.5	4.1
			0.5	43.1	3.4	45.5	4.2
			0.9	87.8	3.4	88.0	4.8
		40	0.1	48.7	6.9	43.3	5.6
			0.5	60.7	3.1	61.8	3.7
			0.9	98.5	3.6	98.6	5.5
y	N	25	0.1	57.6	10.1	34.2	4.9
			0.5	59.0	7.8	54.0	5.1
			0.9	95.3	2.4	97.1	5.2
		40	0.1	67.1	9.6	49.1	4.9
			0.5	78.2	6.1	74.3	4.7
			0.9	99.9	2.0	99.9	4.8
	t	25	0.1	52.0	9.8	28.4	3.9
			0.5	45.6	7.7	40.5	4.5
			0.9	87.9	3.1	89.3	5.5
		40	0.1	63.1	9.4	44.0	5.0
			0.5	66.7	6.8	61.9	4.6
			0.9	97.9	2.1	98.1	4.3

We denote: dmf–different mean functions (n–no, y–yes); distr–distribution (N–normal, t–t-distribution); $n = n_1 = n_2$; dsd_s, psd_s (respectively dsd_c, psd_c)–standard (respectively centered) permutation tests

supplementary online material. They suggest that the variability of the second sample is greater than that of the first sample.

We set $\rho = 0.1, 0.5, 0.9$ for high, moderate and low correlation of the functional data (see Guo et al 2019 for explanation).

The p-values of the permutation tests were obtained using $B = 1000$ runs of random permutations. We reject the null hypothesis when the p-value of a test is smaller than the nominal significance level α, which is set to 5%. This process is repeated 1000 times. The empirical sizes or powers of the test procedures are then the percentages of rejections. The simulation studies, as well as the real data examples of Sect. 4, were conducted using the R programming language (R Core Team 2022). The code is given in the supplementary materials.

Table 3 Empirical powers (in %) of all tests for Model 2

h_2	dmf	distr	n	ρ	dsd_s	psd_s	dsd_c	psd_c
g_1	n	N	25	0.1	60.8	8.1	46.0	5.0
				0.5	69.8	5.1	70.7	5.0
				0.9	99.6	4.5	99.8	8.2
			40	0.1	78.1	8.5	69.8	6.6
				0.5	88.3	6.2	88.9	6.1
				0.9	100.0	5.0	100.0	10.4
		t	25	0.1	55.1	8.2	40.1	4.3
				0.5	63.3	4.8	63.7	5.0
				0.9	97.1	4.1	97.4	7.8
			40	0.1	72.3	6.4	60.1	4.3
				0.5	83.0	5.3	82.9	5.4
				0.9	100.0	5.8	100.0	11.5
	y	N	25	0.1	70.4	11.2	49.2	6.3
				0.5	72.3	7.8	68.2	6.5
				0.9	99.9	5.6	99.9	10.0
			40	0.1	84.6	8.2	70.7	4.5
				0.5	92.7	7.1	91.0	5.3
				0.9	100.0	7.1	100.0	12.0
		t	25	0.1	61.8	9.5	38.2	5.0
				0.5	65.8	8.9	60.8	6.1
				0.9	97.1	4.2	97.8	9.6
			40	0.1	80.8	8.1	64.8	4.5
				0.5	83.3	6.5	80.8	4.5
				0.9	99.7	6.3	99.7	12.3
g_2	n	N	25	0.1	54.8	7.0	51.4	5.3
				0.5	70.9	4.7	72.9	5.9
				0.9	99.9	6.4	99.9	6.3
			40	0.1	77.1	6.1	74.9	4.6
				0.5	91.4	4.4	92.9	5.4
				0.9	100.0	8.2	100.0	9.9
		t	25	0.1	48.4	7.5	44.1	5.0
				0.5	62.6	4.2	64.8	4.6
				0.9	98.3	6.1	98.3	7.2
			40	0.1	68.1	7.0	65.2	6.2
				0.5	83.9	6.3	84.7	6.7
				0.9	100.0	7.8	100.0	10.2
	y	N	25	0.1	74.7	10.3	52.4	5.6

Table 3 continued

h_2	dmf	distr	n	ρ	dsd_s	psd_s	dsd_c	psd_c
				0.5	77.5	7.6	72.5	5.0
				0.9	100.0	4.0	100.0	6.5
			40	0.1	87.8	10.5	76.8	6.3
				0.5	94.1	7.6	92.4	6.2
				0.9	100.0	6.3	100.0	9.0
		t	25	0.1	63.6	11.6	42.0	6.9
				0.5	71.3	6.7	66.6	4.7
				0.9	98.4	4.9	98.4	6.7
			40	0.1	79.8	8.6	65.7	4.8
				0.5	86.8	7.2	84.7	6.2
				0.9	100.0	6.2	100.0	10.6

We denote: $g_1(t) = \sin(1.5t) + 1$ and $g_2(t) = \cos(1.5t) + 1$; dmf–different mean functions (n–no, y–yes); distr–distribution (N–normal, t–t-distribution); $n = n_1 = n_2$; dsd_s, psd_s (respectively dsd_c, psd_c) standard (respectively centered) permutation tests

3.2 Simulation results

Let us discuss the simulation results for the permutation tests obtained in Models 0–5. We consider four tests: the standard and centered permutation tests based on distance and projection standard deviations. We denote them by dsd_s, psd_s, dsd_c, and psd_c, if needed. The results of the simulation studies are presented in Tables 1, 2, 3, 4, 5 and 6.

The type I error level of the tests was investigated in Model 0, where the null hypothesis is true. The obtained empirical sizes are given in Table 1. We can observe that the centered permutation tests based on distance and projection standard deviations control the type I error level very well in all cases. Their empirical sizes always lie within the 95% binomial confidence interval [3.6%, 6.4%] (Duchesne and Francq 2015). The same holds for the standard permutation tests, but only for the case with the same mean function. When the mean functions are different, the standard permutation procedures are far from maintaining the type I error level. For highly and moderately correlated functional data, they are too liberal, while in the case of small correlation, they may have a conservative character.

In Model 1 (Table 2), we have a simple difference in the scale of the functional data. This explains the complete loss of power of the tests based on the projection standard deviation. Their empirical powers are close to the significance level $\alpha = 5\%$ or even smaller. For the case of different mean functions, the slightly greater empirical powers of the psd_s test are due to the too liberal character of this test observed in Model 0. On the other hand, the test procedures based on the distance standard deviation have much more acceptable power. This power decreases with an increase in correlation. When the mean functions are the same, both permutation tests have very similar power for low and moderate correlation, with a slight advantage for the centered permutation test. For high correlation, the dsd_s test is a few percent more powerful than the dsd_c test.

Table 4 Empirical powers (in %) of all tests for Model 3

dmf	distr	n	ρ	dsd_s	psd_s	dsd_c	psd_c
n	N	25	0.1	63.3	98.9	63.0	98.4
			0.5	38.5	92.5	37.8	92.7
			0.9	39.9	5.9	41.0	5.5
		40	0.1	85.6	100.0	85.3	100.0
			0.5	54.6	99.3	54.9	99.2
			0.9	69.5	5.6	69.5	5.3
	t	25	0.1	57.6	97.2	57.9	97.0
			0.5	33.9	81.3	34.6	80.9
			0.9	32.0	4.8	32.6	5.1
		40	0.1	82.7	99.8	82.4	99.9
			0.5	49.1	97.8	50.5	97.7
			0.9	48.7	6.0	48.6	5.9
y	N	25	0.1	78.0	99.3	64.0	98.3
			0.5	43.5	93.1	39.2	91.3
			0.9	38.8	1.1	45.6	4.4
		40	0.1	93.2	100.0	87.4	100.0
			0.5	59.5	99.2	56.6	98.8
			0.9	61.0	2.8	64.4	5.9
	t	25	0.1	75.0	99.2	58.7	97.1
			0.5	41.1	84.1	35.9	79.7
			0.9	30.5	2.0	33.2	4.7
		40	0.1	89.2	100.0	81.5	100.0
			0.5	55.2	97.5	51.6	96.4
			0.9	46.9	2.9	51.0	6.7

We denote: dmf–different mean functions (n–no, y–yes); distr– distribution (N–normal, t–t-distribution); $n = n_1 = n_2$; dsd_s, psd_s (respectively dsd_c, psd_c) standard (respectively centered) permutation tests

Nevertheless, for different mean functions, we can see that the dsd_s test is too liberal, and as a result has undesirably large power. Fortunately, the centered permutation test based on the distance standard deviation has very similar power in the cases when the mean functions are the same and when they are different.

In Model 2 (Table 3), we have two cases. In the first (respectively second) case, the much greater variability of the second sample is concentrated for greater (respectively smaller) values of t, closer to one (respectively zero). Nevertheless, the conclusions are very similar to those for Model 1. Thus, the psd-based tests have almost no power, while the dsd_s test has too much power in the case of different mean functions. The best is the dsd_c test. For this test, at least slightly greater empirical powers occur for $h_2 = g_2$ under high and moderate correlation. For a small correlation ($\rho = 0.9$) the power is almost the same.

Table 5 Empirical powers (in %) of all tests for Model 4

dmf	distr	n	ρ	dsd_s	psd_s	dsd_c	psd_c
n	N	25	0.1	54.1	95.9	53.5	95.8
			0.5	15.6	95.4	15.6	95.7
			0.9	21.8	9.0	22.0	8.7
		40	0.1	76.4	99.8	76.5	99.6
			0.5	23.5	99.9	23.6	99.8
			0.9	30.8	13.2	30.3	13.4
	t	25	0.1	48.7	88.4	48.3	88.9
			0.5	17.2	88.1	17.2	88.1
			0.9	16.3	8.1	16.1	7.5
		40	0.1	69.6	99.0	70.7	99.0
			0.5	23.9	98.0	23.9	98.1
			0.9	23.3	8.8	23.8	8.7
y	N	25	0.1	72.5	98.5	53.7	95.5
			0.5	19.8	96.9	16.8	94.6
			0.9	16.7	3.8	22.1	9.7
		40	0.1	88.1	99.9	78.4	99.8
			0.5	25.5	99.7	23.1	99.5
			0.9	27.7	5.4	33.6	13.9
	t	25	0.1	68.6	95.3	50.8	89.7
			0.5	22.2	88.7	18.3	86.8
			0.9	13.7	3.2	16.4	6.8
		40	0.1	80.9	99.3	68.7	98.8
			0.5	27.5	99.1	25.0	98.6
			0.9	23.7	4.9	25.2	8.8

We denote: dmf–different mean functions (n–no, y–yes); distr– distribution (N–normal, t–t-distribution); $n = n_1 = n_2$; dsd_s, psd_s (respectively dsd_c, psd_c) standard (respectively centered) permutation tests

Models 3 and 4 (Tables 4 and 5) alter the λ_{2l} values, which results in greater variability in the second sample. However, its character seems to be different than in Models 1 and 2, as can be seen in the simulation results. The conclusions for the dsd-based tests are very similar to those for the earlier models, but for the case with the same mean functions, the power of both permutation methods is very similar. On the other hand, we observe much better properties of the test based on the projection standard deviation for high and moderate correlation. Namely, both psd-based permutation tests have similar and high power, which is independent of the equality or inequality of the mean functions. They are also more powerful than the dsd_s and dsd_c test procedures. For low correlation, we observe the opposite situation, and the behavior of psd-based tests is still poor. Nevertheless, these two models show that there is some potential for the use of the projection standard deviation.

Table 6 Empirical powers (in %) of all tests for Model 5

ω	dmf	distr	n	ρ	dsd_s	psd_s	dsd_c	psd_c
2	n	N	25	0.1	11.9	17.6	11.8	16.7
				0.5	66.9	77.7	66.9	77.7
				0.9	86.3	89.2	86.0	88.4
			40	0.1	18.2	25.2	18.0	25.2
				0.5	88.9	94.2	89.0	95.0
				0.9	97.8	98.9	97.8	98.9
		t	25	0.1	12.5	19.2	13.5	19.5
				0.5	63.9	75.2	63.3	74.9
				0.9	72.7	78.7	72.4	77.8
			40	0.1	16.7	28.0	17.2	27.6
				0.5	84.9	91.5	85.3	92.0
				0.9	93.2	96.7	93.2	96.8
	y	N	25	0.1	26.4	26.9	10.2	16.3
				0.5	74.0	84.0	66.5	77.4
				0.9	78.8	81.8	83.5	87.1
			40	0.1	32.9	34.9	17.5	26.0
				0.5	92.2	96.6	89.4	94.8
				0.9	97.4	99.0	98.1	99.4
		t	25	0.1	31.5	28.1	13.5	18.1
				0.5	71.2	79.2	60.9	71.4
				0.9	69.4	72.5	75.8	80.2
			40	0.1	34.4	34.8	17.6	26.4
				0.5	88.7	95.3	84.8	92.8
				0.9	93.0	95.2	94.5	96.9
3	n	N	25	0.1	28.8	44.7	29.0	44.3
				0.5	94.7	98.5	94.5	98.2
				0.9	98.9	99.7	99.2	99.5
			40	0.1	47.1	64.7	46.5	64.8
				0.5	99.6	100.0	99.7	100.0
				0.9	100.0	100.0	100.0	100.0
		t	25	0.1	30.9	47.7	31.5	46.6
				0.5	90.6	95.1	90.7	94.5
				0.9	96.6	96.5	96.4	96.5
			40	0.1	48.2	71.5	47.8	71.6
				0.5	99.3	99.8	99.5	99.8
				0.9	100.0	100.0	99.9	100.0
	y	N	25	0.1	51.6	57.2	29.7	46.6

Table 6 continued

ω	dmf	distr	n	ρ	dsd_s	psd_s	dsd_c	psd_c
				0.5	96.1	98.9	94.1	98.0
				0.9	98.1	99.6	99.2	99.8
			40	0.1	62.9	74.5	45.3	66.0
				0.5	99.9	99.9	99.7	99.9
				0.9	99.9	100.0	100.0	100.0
		t	25	0.1	53.4	60.9	30.9	46.3
				0.5	93.3	96.8	89.7	94.8
				0.9	93.2	94.9	95.8	97.5
			40	0.1	66.3	77.6	47.4	68.8
				0.5	99.2	99.9	99.0	99.8
				0.9	99.6	100.0	99.6	100.0

We denote: $\omega = 2, 3$ in the second sample; dmf–different mean functions (n–no, y–yes); distr–distribution (*N*–normal, *t* - *t*-distribution); $n = n_1 = n_2$; dsd_s, psd_s (respectively dsd_c, psd_c) standard (respectively centered) permutation tests

This can also be seen in Model 5 (Table 6). Here the tests based on projection standard deviation outperform those using the distance standard deviation in most cases. The former tests are also powerful in the case of low correlation. The other conclusions are the same as for Models 3 and 4. Naturally, the power increases with an increase in ω (see the figures in the supplementary online material).

Finally, we can observe that the empirical powers for the normal distribution are usually a few percent greater than those for the *t*-distribution case. Nevertheless, in the latter case, the results are still satisfactory.

Let us sum up the above results. The centered permutation tests seem to perform very well independent of whether or not the mean functions are the same. This is not true for the standard permutation method, which may be too liberal for cases with different means. Thus, we recommend using the centered permutation procedure. The tests based on distance standard deviation are powerful for all of the variability scenarios considered in these simulation studies. However, when the variability is internally caused, the tests based on projection standard deviation may be more powerful. On the other hand, these tests are of no use at all for various scale changes of functional data. Thus, the choice between different standard deviations may require some inspection of the data.

4 Real data application

In this section, we illustrate the use of the standard and centered permutation tests based on distance and projection standard deviations. We consider five data sets representing different subject matter, types of variability, numbers of observations, and design time points. All data sets are presented in Fig. 1. For each data set, we removed the outlying observations detected by an outliergram as proposed by Arribas-Gil and Romo (2014),
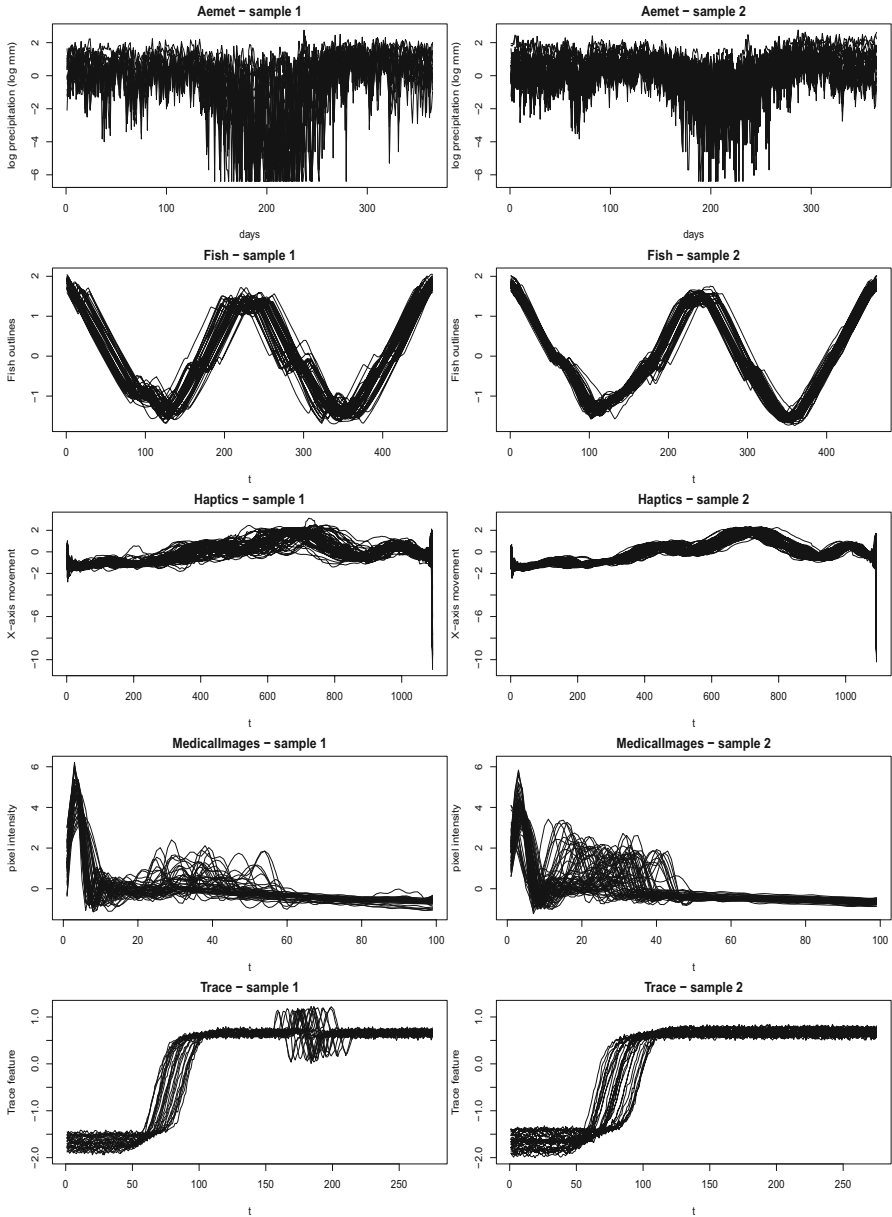
Fig. 1 Trajectories of real data sets considered in Sect. 4

and the equality of mean functions was checked using the tests contained in the R package fdANOVA (Górecki and Smaga 2018). The results of the FANOVA tests are presented in the supplementary online material.

We first consider the aemet data set, available in the R package fda.usc (Febrero-Bande and Oviedo de la Fuente 2012). This data set contains a series of daily summaries of 73 Spanish weather stations selected for the period 1980–2009. More precisely, geographic information is given for each station, along with the averages for the period 1980–2009 of daily temperature, daily precipitation, and daily wind speed. The data are obtained from the Meteorological State Agency of Spain (AEMET), Government of Spain (http://www.aemet.es/). Here, we consider the data for the mean curve of the log precipitation (in log mm) in continental Spain. We construct two samples by dividing the stations into those having an altitude of less than 100 ms (the first sample) and the remainder (the second sample). Then, the samples contain $n_1 = 20$ and $n_2 = 30$ observations. From Fig. 1, it appears that the variability is greater in the first sample. This is confirmed by the values of the estimators of standard deviations. Namely, for the first (respectively second) sample, $\widehat{dsd}$ and $\widehat{psd}$ are equal to 9.34 and 0.33 (respectively 5.42 and 0.3). However, the amount of difference in variability does not seem to be evenly distributed: it is greater in the middle of the year than at its beginning and end. This is similar to the case of Model 2 from the simulation studies, which may explain why the standard and centered permutation psd-based tests have p-values equal to 0.784 and 0.832 respectively. Thus, they do not reject the null hypothesis, perhaps due to loss of power. The dsd-based tests have p-values of 0.066 and 0.05 and are on the boundary of rejection and non-rejection, suggesting that they are much more powerful than their competitors based on the projection standard deviation. The slightly greater p-value of the standard permutation test can be explained by the fact that the mean functions seem to be significantly different (see supplementary online material), while the correlation is small, and hence this test is subject to a loss of power.

The next four data sets are available in the UEA & UCR Time Series Classification Repository (Bagnall et al 2022), which contains many real data sets for the classification of time series. The first of them is the Fish data set, which contains fish outlines originally used with contour matching in Dah-Jye et al (2008). The outlines were derived at UCR. Each class is a different species. Here, we consider the first two classes with $n_1 = 47$ and $n_2 = 46$. We have $\widehat{dsd}(\mathbf{X}_1) = 2.47$, $\widehat{dsd}(\mathbf{X}_2) = 1.55$, $\widehat{psd}(\mathbf{X}_1) = 0.4$, and $\widehat{psd}(\mathbf{X}_2) = 0.38$, which together with the second row of Fig. 1 suggests that again the variability in the first sample is greater than in the second. Moreover, the charts suggest a simple scale difference, as in Model 1 of the simulations. Thus, the tests based on the projection standard deviation are likely to lose power, and this is confirmed by their p-values, which are close to 0.45. On the other hand, the p-values of the standard and centered permutation dsd-based tests are 0.006 and 0.004 respectively. Thus, these tests reject the null hypothesis, indicating greater variability for the first species. Very similar results are obtained for the Haptics data set. Those data are taken from five people entering their pass graph on a touchscreen, and include the X-axis movement only. Here, the sample sizes, values of estimators and p-values of the tests are as follows: $n_1 = 63$, $n_2 = 82$, $\widehat{dsd}(\mathbf{X}_1) = 5.49$, $\widehat{dsd}(\mathbf{X}_2) = 3.35$, $\widehat{psd}(\mathbf{X}_1) = 0.3$, $\widehat{psd}(\mathbf{X}_2) = 0.29$, both dsd-based tests have p-values equal to zero, while the p-values of the standard and centered permutation tests based on psd are 0.538 and 0.646 respectively.

Next, we consider the MedicalImages data set, where the variability scenario seems to be different than in the above examples. In this data set, the observations are histograms of the pixel intensity of medical images, and the classes are different regions of the human body. For illustrative purposes, we use the second and third classes, with $n_1 = 49$ and $n_2 = 64$. Row four in Fig. 1 suggests that the difference in variability is not just in scale. The variability is similar at the beginning and end of the observable period, while in the middle, the variability seems to be greater in the second sample. However, it is difficult to identify the kind of variability in this case. In general, the greater variability in the second sample seems to be confirmed, since $\widehat{dsd}(\mathbf{X}_1) = 1.13$, $\widehat{dsd}(\mathbf{X}_2) = 1.73$, $\widehat{psd}(\mathbf{X}_1) = 0.27$, and $\widehat{psd}(\mathbf{X}_2) = 0.28$. Moreover, all tests reject the null hypothesis. The dsd-based tests have zero p-values. The p-values of the standard and centered permutation tests based on projection standard deviation are 0.014 and 0.004 respectively. The greater p-value of the standard method is due to the significant differences in the mean functions (see supplementary online material) and the relatively small correlation–the standard permutation test may have a conservative character, resulting in loss of power.

Finally, we study the Trace data set, which is a subset of the Transient Classification Benchmark (trace project), an initiative at the turn of the century to collate data from the application domain of the process industry (e.g. nuclear, chemical, etc.). It is a synthetic data set designed to simulate instrumentation failures in a nuclear power plant, created by Davide Roverso. In this data set, there are four classes, and we use the last two, with $n_1 = 28$ and $n_2 = 31$ observations. The values of the estimators are as follows: $\widehat{dsd}(\mathbf{X}_1) = 1.69$, $\widehat{dsd}(\mathbf{X}_2) = 2.35$, $\widehat{psd}(\mathbf{X}_1) = 0.44$, and $\widehat{psd}(\mathbf{X}_2) = 0.62$. They suggest greater variability in the second sample. Figure 1 confirms this, but we can observe a kind of perturbation for $t \in [150, 215]$. This may be a reason for the non-rejection of the null hypothesis by the dsd-based tests, with p-values close to 10%. On the other hand, the psd-based tests seem to be robust to this perturbation and reject the null hypothesis with zero p-values. This example shows that the projection standard deviation can sometimes be useful. The similar p-values of both standard and centered permutation tests can be explained by the fact that the mean functions seem to be equal (see supplementary online material).

5 Conclusion

In this paper, we have studied the applicability of distance and projection standard deviations to measure and test the equality of variability of functional data. Of course, the variability is not strictly defined in the context of functional data. One of its measures is the covariance function, which, being a function, is more difficult to interpret. The standard deviations considered are more practical, as they are real numbers. The distance standard deviation appeared to be useful for different (all considered) kinds of variability of functional data. On the other hand, the projection standard deviation failed to detect scale differences, but was more useful when the variability was caused by "internal" issues. These observations were made in extensive simulation studies, where the type I error level and power of the two-sample tests for equality of stan-

dard deviations were investigated. The test procedures were based on approximating the distribution of the test statistic as described by Edelmann et al (2020), using the new permutation method. In this method, we first centered the observations, and then applied the permutation approach. This simple modification of the standard permutation method resulted in maintenance of the type I error level, even when the mean functions were different. The use of the proposed methods was illustrated in five real data examples, which were based on data sets with various characteristics. For most of the data sets, the tests based on distance standard deviation were more powerful, but for some of them, the projection method gave better results. To sum up, the proposed methods seem to be promising for measuring and testing the variability of functional data in an interpretive way, but they need to be further evaluated with additional real and artificial data.

Supplementary information In the supporting material, we present additional simulation studies, results of FANOVA tests for real data examples of Sect. 4, the example of generated functional data in the simulation studies of Sect. 3, and the R code for simulation studies and real data examples.

Declarations

Conflict of interest The authors declare no conflict of interest

References

Arboretti R, Pesarin F, Salmaso L (2021) A unified approach to permutation testing for equivalence. Stat Methods Appl 30:1033–1052

Arribas-Gil A, Romo J (2014) Shape outlier detection and visualization for functional data: the outliergram. Biostatistics 15:603–619

Arvesen JN (1969) Jackknifing U-statistics. Ann Math Stat 40:2076–2100

Bagnall A, Lines J, Vickers W, et al (2022) The UEA & UCR time series classification repository. www.timeseriesclassification.com

Berrett TB, Samworth RJ (2019) Nonparametric independence testing via mutual information. Biometrika 106:547–566

Bickel PJ, Lehmann EL (2012) Descriptive statistics for nonparametric models, iv. spread. In: Rojo J (ed.) Selected works of E. L. Lehmann. selected works in probability and statistics, Springer, New York, pp 519–526

Corain L, Melas VB, Pepelyshev A et al (2014) New insights on permutation approach for hypothesis testing on functional data. Adv Data Anal Classif 8:339–356

Cuevas A (2014) A partial overview of the theory of statistics with functional data. J Stat Plann Inference 147:1–23

Dah-Jye L, Archibald J, Schoenberger R, et al (2008) Contour matching for fish species recognition and migration monitoring, vol 122, pp 183–207. https://doi.org/10.1007/978-3-540-78534-7_8

Du H, Wang L (2020) Testing variance components in linear mixed modeling using permutation. Multivar Behav Res 55:120–136

Duchesne P, Francq C (2015) Multivariate hypothesis testing using generalized and {2}-inverses - with applications. Statistics 49:475–496

Dueck J, Edelmann D, Richards D (2015) A generalization of an integral arising in the theory of distance correlation. Stat Probab Lett 97:116–119

Edelmann D, Fiedler J (2022) dcortools: providing fast and flexible functions for distance correlation analysis. https://CRAN.R-project.org/package=dcortools, R package version 0.1.6

Edelmann D, Richards D, Vogel D (2020) The distance standard deviation. Ann Stat 48:3395–3416

Febrero-Bande M, Oviedo de la Fuente M (2012) Statistical computing in functional data analysis: The R package fda.usc. J Stat Softw 51:1–28

Ferraty F, Vieu P (2006) Nonparametric functional data analysis: theory and practice. Springer

Fokianos K, Pitsillou M (2017) Consistent testing for pairwise dependence in time series. Technometrics 59:262–270

Górecki T, Smaga Ł (2018) fdANOVA: analysis of variance for univariate and multivariate functional data. https://CRAN.R-project.org/package=fdANOVA, R package version 0.1.2

Górecki T, Smaga Ł (2019) fdANOVA: An R software package for analysis of variance for univariate and multivariate functional data. Comput Stat 34:571–597

Gretton A, Bousquet O, Smola A, et al (2005) Measuring statistical dependence with Hilbert-Schmidt norms. In: ALT, Springer, pp 63–77

Gretton A, Fukumizu K, Teo CH et al (2008) A kernel statistical test of independence. In: Cambridge MA (ed) Advances in neural information processing systems 20. MIT Press, Cambridge, pp 585–592

Gretton A, Fukumizu K, Harchaoui Z, et al (2009) A fast, consistent kernel two-sample test. In: Advances in neural information processing systems, 22, Curran Associates, Inc., pp. 673–681

Guo J, Zhou B, Zhang JT (2018) Testing the equality of several covariance functions for functional data: a supremum-norm based test. Comput Stat Data Anal 124:15–26

Guo J, Zhou B, Zhang JT (2019) New tests for equality of several covariance functions for functional data. J Am Stat Assoc 114:1251–1263

Horváth L, Kokoszka P (2012) Inference for functional data with applications. Springer, Berlin

Jacques J, Preda C (2014) Functional data clustering: a survey. Adv Data Anal Classif 8:231–255

Kong J, Klein B, Klein R et al (2012) Using distance correlation and ss-anova to assess associations of familial relationships, lifestyle factors, diseases, and mortality. Proc Natl Acad Sci USA 109:20352–20357

Kraus D, Panareto VM (2012) Dispersion operators and resistant second-order functional data analysis. Biometrika 99:813–832

Li R, Zhong W, Zhu L (2012) Feature screening via distance correlation learning. J Am Stat Assoc 107:1129–1139

Lyons R (2013) Distance covariance in metric spaces. Ann Probab 41:3284–3305

Pan W, Tian Y, Wang X et al (2018) Ball divergence: nonparametric two sample test. Ann Stat 46:1109–1137

R Core Team (2022) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ramos-Carreño C (2022) dcor: distance correlation and energy statistics in python. https://doi.org/10.5281/zenodo.3468124, [Computer software]

Ramsay JO, Silverman BW (2002) Applied functional data analysis: methods and case studies. Springer, Berlin

Ramsay JO, Silverman BW (2005) Functional data analysis, Second Edition. Springer

Ramsay JO, Hooker G, Graves S (2009) Functional data analysis with R and MATLAB. Springer, New York

Rizzo ML, Székely GJ (2021) Energy: E-statistics: multivariate inference via the energy of data. https://CRAN.R-project.org/package=energy, R package version 1.7-8

Roy A, Post C (2012) Detection of long-range concerted motions in protein by a distance covariance. J Chem Theory Comput 8:3009–3014

Székely GJ, Rizzo ML (2009) Brownian distance covariance. Ann Appl Stat 3:1236–1265

Székely GJ, Rizzo ML (2013) The distance correlation t-test of independence in high dimension. J Multivar Anal 117:193–213

Székely GJ, Rizzo ML (2014) Partial distance correlation with methods for dissimilarities. Ann Stat 42:2382–2412

Székely GJ, Rizzo ML, Bakirov NK (2007) Measuring and testing dependence by correlation of distances. Ann Stat 35:2769–2794

Wang JL, Chiou JM, Müller HG (2015) Review of functional data analysis. Ann Rev Stat Appl 3:257–295

Zhang JT (2013) Analysis of variance for functional data. Chapman & Hall, London

Zhang JT, Liang X (2013) One-way ANOVA for functional data via globalizing the pointwise F-test. Scand J Stat 41:51–71

Zhu C, Zhang X, Yao S et al (2020) Distance-based and RKHS-based dependence metrics in high dimension. Ann Stat 48:3366–3394

Zhu L, Xu K, Li R et al (2017) Projection correlation between two random vectors. Biometrika 104:829–843

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.