Two-Sample Tests for Multivariate Functional Data with Applications

Zhiping Qiu
National Huaqiao University

Jin-Ting Zhang
National University of Singapore

Two-Sample Tests for Multivariate Functional Data with Applications

Zhiping Qiu[1,2], Jianwei Chen[1,2], and Jin-Ting Zhang[3*]

[1] Research Center of Appl. Statist. and Big Data, Huaqiao Univ., China

[2] School of Statistics, Huaqiao Univ., China

[3] Dept of Statist. and Appl. Prob., National Univ. of Singapore, Singapore

Abstract

Multivariate functional data are frequently obtained in many scientific or industrial areas where several functions for a statistical unit are observed over time. It is often interesting to check if the mean vector functions of two multivariate functional samples are equal. In this article, two global tests for the above two-sample problem for multivariate functional data are proposed and studied. Their asymptotic random expressions under the null and certain local alternative hypotheses are derived and their root-n consistencies are established. A simulation study and a real data application show that the proposed two tests generally have higher or not worse powers than some existing competitors.

Keywords and Phrases: Multivariate functional data; nonparametric bootstrap test; pointwise Hotelling T^2-test; two-sample test; Welch-Satterthwaite χ^2-approximation.

1 Introduction

With development of data collection technology, functional data are often encountered in many scientific fields including chemometrics, genomics, medicine, seismology to name a few. This type of data is characterized by observations of one or several functions for a statistical unit. Efficient statistical inferences of this type of data are increasingly in great

*Corresponding Author. stazjt@nus.edu.sg.

demand (Ramsay and Silverman, 2005; Ferraty, 2011; Zhang, 2013). Many standard statistical methods for multivariate data have been extended for univariate functional data that have only one function observed for a statistical unit. Examples include Boente et al. (2014) and Fremdt et al. (2014) for principal component analysis; Lian (2014) and King (2009) for canonical analysis; and Antoch et al. (2010), Febrero-Bande and González-Manteiga (2013), and Montoya (2013) for linear regression models, among others.

In the recent decades, significance tests are also considered for univariate functional data. Examples include Faraway (1997), Fan and Lin (1998), Shen and Faraway (2004), Zhang and Chen (2007), Zhang et al. (2010), and Zhang (2011) among others. In particular, for analysis of variance (ANOVA) for univariate functional data, a number of works have been done in the literature. Ramsay and Silverman (2005) proposed a pointwise F-test which is conducted for all design time points. However, the pointwise F-test is usually time-consuming and it is not guaranteed that the ANOVA problem is overall significant for a given significance level even when the pointwise F-test is significant for all design time points at the same significance level. To obtain a global conclusion, many tests have been developed using the pointwise F-test statistic. Cuevas et al. (2004) proposed an L_2-norm based test whose test statistic is obtained via integrating the numerator of the pointwise F-test statistic over time. Zhang (2013) further showed that the null distribution of the L_2-norm based test is asymptotically a χ^2-type mixture. On the other hand, Zhang and Liang (2014) developed a global test whose test statistic is obtained via integrating the pointwise F-test statistic over time while Ramsay and Silverman (2005) mentioned a global test whose test statistic is the supremum of the pointwise F-test statistic over time. This latter test is further studied by Zhang et al. (2019) and is called $F_{\max}$ test. Other interesting tests constructed using the pointwise F-test include Cox and Lee (2008). Numerical comparisons of some of these tests are presented in Górecki and Smaga (2015). A survey about hypothesis testing methodologies can be found in Zhang (2013).

All of the above tests are for univariate functional data. In the recent years, more and more interests are attracted for multivariate functional data that have more than one functions observed for a statistical unit. Berrendero et al. (2011) and Chiou et al. (2014) considered principal components analysis; Tokushige et al. (2007) and Jacques and Preda (2014) investigated cluster analysis; Keser and Kocakoç (2015) studied canonical analysis; and Górecki and Smaga (2017) extended some multivariate analysis of variance approaches for multivariate functional data.

In this article, we consider a two-sample problem for multivariate functional data. Our motivation example is the Canadian weather data set which is described in some details in Ramsay and Silverman (2005). It is available in R package *fda.usc*. In the Canadian weather data, the daily temperature and precipitation of 35 weather stations in Canada averaged over the years from 1960 to 1994 were collected so that there are one temperature curve and one precipitation curve observed daily at each weather station. Among the 35 weather stations, there are 15 located at Eastern Canada, 15 at Western Canada, and the remaining 5 at Northern Canada. Of interest is to check if the mean vectors of the temperature and precipitation curves at any given two locations are equal.

Mathematically, a two-sample problem for multivariate functional data can be described as follows. Let $\mathbf{SP}_p(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ denote a p-dimensional stochastic process over a given finite interval $\mathcal{T} = [a, b]$ with mean vector function $\boldsymbol{\mu}(t) \in \mathbb{R}^p, t \in \mathcal{T}$ and covariance matrix function $\boldsymbol{\Gamma}(s,t) \in \mathbb{R}^{p\times p}, s, t \in \mathcal{T}$. Let $\boldsymbol{y}_{ij}(t) = (y_{ij1}(t), y_{ij2}(t), \ldots, y_{ijp}(t))^T$, $j = 1, 2, \ldots, n_i; i = 1, 2$, denote the two multivariate functional samples. Assume that

$$\boldsymbol{y}_{i1}(t), \boldsymbol{y}_{i2}(t), \ldots, \boldsymbol{y}_{in_i}(t) \overset{i.i.d.}{\sim} \mathbf{SP}_p(\boldsymbol{\mu}_i, \boldsymbol{\Gamma}), i = 1, 2, \tag{1}$$

where $\boldsymbol{\mu}_i(t) = (\mu_{i1}(t), \mu_{i2}(t), \ldots, \mu_{ip}(t))^T \in \mathbb{R}^p, t \in \mathcal{T}, i = 1, 2,$ and $\boldsymbol{\Gamma}(s,t) = (\gamma(s,t)) \in \mathbb{R}^{p\times p}, s, t \in \mathcal{T}$ denote the mean vector functions and the common covariance matrix function, respectively. Moreover, the observed vector functions in the two multivariate functional samples are assumed to be independent. Of interest is to test the equality of the two mean vector functions

$$H_0 : \boldsymbol{\mu}_1(t) = \boldsymbol{\mu}_2(t), \quad \text{for any} \quad t \in \mathcal{T} \leftrightarrow H_1 : \boldsymbol{\mu}_1(t) \neq \boldsymbol{\mu}_2(t), \quad \text{for some} \quad t \in \mathcal{T}. \tag{2}$$

The above two-sample problem can be tested using the tests described in Górecki and Smaga (2017). However, we shall propose two global tests which generally have higher or not worse powers than those of Górecki and Smaga (2017) have. In addition, some theoretical analysis of our global tests is also presented.

The remainder of this article is organized as follows. In Section 2, the two global tests for the two-sample problem (2) for multivariate functional data are proposed. Their asymptotic random expressions under the null and some local alternative hypotheses are derived. Two methods for approximating the null distributions of the test statistics are described. Section 3 presents a simulation study, demonstrating that in terms of power, the proposed two tests generally outperform or perform similarly with some existing competitors. Applications to the Canadian weather data are presented in Section 4. Section 5 presents some concluding remarks. The technical proofs of the main results are given in the Appendix.

2 Main Results

Let us first give some notations. For any $t \in \mathcal{T}$, the sample group and grand mean vector functions are respectively given by

$$\bar{\boldsymbol{y}}_{i\cdot}(t) = n_i^{-1}\sum_{j=1}^{n_i}\boldsymbol{y}_{ij}(t), i = 1, 2, \quad \text{and} \quad \bar{\boldsymbol{y}}_{\cdot\cdot}(t) = n^{-1}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\boldsymbol{y}_{ij}(t), \tag{3}$$

where and throughout $n = n_1 + n_2$ denotes the total sample size. For any $s, t \in \mathcal{T}$, the common covariance matrix function $\boldsymbol{\Gamma}(s, t)$ can be estimated by the pooled sample covariance matrix function:

$$\hat{\mathbf{\Gamma}}(s, t) = (n-2)^{-1}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\big(\boldsymbol{y}_{ij}(s) - \bar{\boldsymbol{y}}_{i\cdot}(s)\big)\big(\boldsymbol{y}_{ij}(t) - \bar{\boldsymbol{y}}_{i\cdot}(t)\big)^T. \tag{4}$$

Note that for each given $t \in \mathcal{T}$, the two-sample problem defined by (1) and (2) reduces to a standard two-sample problem for multivariate data which can be tested by the well-known Hotelling T^2-test with test statistic

$$T_n(t) = \frac{n_1 n_2}{n}\big(\bar{\boldsymbol{y}}_{1\cdot}(t) - \bar{\boldsymbol{y}}_{2\cdot}(t)\big)^T(\hat{\mathbf{\Gamma}}(t, t))^{-1}\big(\bar{\boldsymbol{y}}_{1\cdot}(t) - \bar{\boldsymbol{y}}_{2\cdot}(t)\big). \tag{5}$$

When the two multivariate functional samples (1) are Gaussian, it is well known that for any given $t \in \mathcal{T}$, we have $\frac{n-p-1}{(n-2)p}T_n(t) \sim F_{p,n-p-1}$, where $F_{p,n-p-1}$ denotes the F-distribution with

p and $n-p-1$ degrees of freedom. Thus, a pointwise Hotelling T^2-test may be conducted for each given $t \in \mathcal{T}$. That is, for a predetermined significance level α and for a given $t \in \mathcal{T}$, the null hypothesis (2) is rejected when $T_n(t) > \frac{(n-2)p}{n-p-1} F_{p,n-p-1}(\alpha)$, where $F_{p,n-p-1}(\alpha)$ denotes the usual upper 100α percentile of the F-distribution $F_{p,n-p-1}$.

It is clear that under the Gaussian assumption, the pointwise Hotelling T^2-test can be conducted easily. However, it is usually time consuming to conduct the pointwise Hotelling T^2-test for all $t \in \mathcal{T}$. Moreover, even the pointwise Hotelling T^2-test is significant for all $t \in \mathcal{T}$ at a given significance level, it still can not guarantee that the alternative hypothesis is overall significant. To overcome these difficulties, we here propose two global tests for the above two-sample problem for multivariate functional data via globalizing the above pointwise Hotelling T^2-test.

It is well known that global test statistics for functional hypothesis testing can be obtained via integrating a pointwise test statistic or taking its maximum value over a given interval of interest (Zhang and Chen, 2007; Zhang et al., 2010; Zhang and Liang, 2014; Górecki and Smaga, 2015; Zhang et al., 2019). Following this idea, we obtain the following two global test statistics:

$$T_n = \int_{\mathcal{T}} T_n(t)dt, \quad \text{and} \quad T_{n,\max} = \sup_{t \in \mathcal{T}} T_n(t), \tag{6}$$

via integrating the pointwise Hotelling T^2-test statistic (5) and taking its supremum value over $\mathcal{T}$, respectively.

2.1 Asymptotical null distributions

Let $\mathcal{L}^2(\mathcal{T})$ be the set of all square-integrable functions over $\mathcal{T}$ and $C^\beta(\mathcal{T})$ be the set of functions which are Hölder continuous with exponent β $(0 < \beta \leq 1)$ and Hölder modulus $\| f \|_{C^\beta}$. Let $\|\boldsymbol{v}\| = [\sum_{i=1}^p v_i^2]^{1/2}$ and $\|\boldsymbol{A}\| = \left[\sum_{i=1}^p \sum_{j=1}^p a_{ij}^2\right]^{1/2}$ denote the usual L^2-norms of $\boldsymbol{v} = (v_i)_{i=1}^p$ and $\boldsymbol{A} = (a_{ij})_{i,j=1}^p$.

To study the asymptotic properties of T_n and $T_{n,\max}$, we impose the following regularity conditions:

(C1) Each element of the mean vector functions $\boldsymbol{\mu}_i(t), (i = 1, 2)$ satisfies $\mu_{ik}(t) \in \mathcal{L}^2(\mathcal{T}), k = 1, 2, \ldots, p$ and $\| \int_{\mathcal{T}} \boldsymbol{\Gamma}(t,t)dt \| < \infty$.

(C2) The group sample sizes tend to infinity proportionally, i.e., $n_1/n \to \tau > 0$ as $n \to \infty$.

(C3) For any $s, t \in \mathcal{T}$, each element of $\Gamma(s,t)$, denoted as $\gamma_{kl}(s,t)$, satisfies $\gamma_{kl}(s,t) \in C^{2\beta}(\mathcal{T} \times \mathcal{T})$ and $\sup_{t \in \mathcal{T}} \| \boldsymbol{\Gamma}(t,t) \| < \infty$, where $0 < 2\beta \leq 1$ and $k, l = 1, 2, \ldots, p$. For any $t \in \mathcal{T}$, $\boldsymbol{\Gamma}(t,t)$ is nonsingular, and $\sup_{t \in \mathcal{T}} \| \boldsymbol{\Gamma}(t,t) \| < \infty$.

(C4) $E\big\{ \int_{\mathcal{T}} [y_{11k}(t) - \mu_{1k}(t)]^2 dt \big\}^2 < \infty, k = 1, 2, \ldots, p$.

(C5) For any $s, t \in \mathcal{T}$, there exists a constant C, independent of s and t, such that $E\big\{ [y_{11k}(s) - \mu_{1k}(s)]^2 [y_{11l}(t) - \mu_{1l}(t)]^2 \big\} < C < \infty, k, l = 1, 2, \ldots, p$.

Let $\mathbf{GP}_p(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ denote a p-dimensional Gaussian process over $\mathcal{T}$ with mean vector function $\boldsymbol{\mu}(t) \in \mathbb{R}^p, t \in \mathcal{T}$ and covariance matrix function $\boldsymbol{\Gamma}(s,t) \in \mathbb{R}^{p \times p}, s, t \in \mathcal{T}$. Here and throughout, let "$\xrightarrow{d}$" denote convergence in distribution and $X \stackrel{d}{=} Y$ denote that X and Y have the same distribution. Under the above conditions and the null hypothesis, we have the following asymptotic random expressions for T_n and $T_{n,\max}$, whose proofs are given in the Appendix.

Theorem 1. *Under Conditions (C1)–(C5) and the null hypothesis, as $n \to \infty$, we have*

$$T_n \xrightarrow{d} T^* = \int_{\mathcal{T}} \|\boldsymbol{w}(t)\|^2 dt \stackrel{d}{=} \sum_{r=1}^{\infty} \lambda_r A_r, \quad \text{and} \quad T_{n,\max} \xrightarrow{d} T^*_{\max} = \sup_{t \in \mathcal{T}} \|\boldsymbol{w}(t)\|^2, \tag{7}$$

where $A_r, r = 1, 2, \cdots \stackrel{i.i.d.}{\sim} \chi^2_1$, and $\boldsymbol{w}(t) \sim \mathbf{GP}_p\big(\mathbf{0}, \boldsymbol{\Gamma}^*\big)$ with

$$\boldsymbol{\Gamma}^*(s,t) = \mathrm{Cov}(\boldsymbol{w}(s), \boldsymbol{w}(t)) = \boldsymbol{\Gamma}^{-\frac{1}{2}}(s,s)\boldsymbol{\Gamma}(s,t)\boldsymbol{\Gamma}^{-\frac{1}{2}}(t,t), \tag{8}$$

and $\lambda_r, r = 1, 2, \ldots, \infty$ being the decreasing-ordered eigenvalues of $\boldsymbol{\Gamma}^*(s,t)$.

2.2 Approximating the null distributions

We now study how to approximate the null distributions of T_n and $T_{n,\max}$ respectively. First of all, we consider to adopt the nonparametric bootstrap (NPB) method described in Zhang et al. (2019) as below.

Let

$$\hat{\boldsymbol{v}}_{ij}(t) = \boldsymbol{y}_{ij}(t) - \bar{\boldsymbol{y}}_{i\cdot}(t), \quad j = 1, 2, \ldots, n_i, i = 1, 2, \tag{9}$$

denote the subject-effect functions. We then resample each of the subject-effect functions in (9) equally likely and obtain a pair of bootstrap samples. We denote them as

$$\boldsymbol{v}^*_{ij}(t), \quad j = 1, 2, \ldots, n_i, i = 1, 2. \tag{10}$$

Recall that the test statistics T_n and $T_{n,\max}$ are computed in (6) using the two original samples (1). Similarly, we compute the NPB test statistics T^*_n and $T^*_{n,\max}$ based on the above bootstrap samples (10). We repeat the above process for a large number of times, say N times and record the associated T^*_n and $T^*_{n,\max}$ as T^{*l}_n and $T^{*l}_{n,\max}$ for $l = 1, 2, \cdots, N$. Then the sample percentiles of $T^{*l}_n, l = 1, 2, \cdots, N$ and $T^{*l}_{n,\max}, l = 1, 2, \cdots, N$ can be used to approximate the percentiles of T_n and $T_{n,\max}$ respectively.

Note that the above NPB tests are usually time-consuming especially when the group sample sizes are large. According to Theorem 1, the limit distribution of T_n under the null hypothesis is a χ^2-type mixture. It is well known that when the null limit distribution of a test statistic is a χ^2-type mixture, its null distribution can be accurately approximated by the Welch-Satterthwaite χ^2-approximation (Zhang 2013; Zhang and Liang 2014; Guo et al. 2019; Smaga and Zhang 2019, among others). Therefore, to save time, when the sample sizes are large, we can also approximate the null distribution of T_n using the Welch-Satterthwaite χ^2-approximation. Its key idea is to approximate the null distribution of T_n using the distribution of a random variable of form $R \sim \beta\chi^2_d$ where the unknown parameters β and d are determined via matching the means and variances of T_n and R under the null hypothesis. For this end, we first derive the asymptotic mean and variance of T_n under the null hypothesis. Let $tr(\boldsymbol{B})$ denote the usual trace of a square matrix $\boldsymbol{B}$.

Theorem 2. *Under Conditions (C1)–(C5) and the null hypothesis, as $n \to \infty$, we have* $\mathrm{E}(T_n) = p(b-a) + o(1)$ *and* $\mathrm{Var}(T_n) = 2\int_{\mathcal{T}}\int_{\mathcal{T}} \mathrm{tr}\left[\boldsymbol{\Gamma}^*(s,t)\boldsymbol{\Gamma}^{*T}(s,t)\right] dsdt + o(1)$.

Note that $\mathrm{E}(R) = d\beta$ and $\mathrm{Var}(R) = 2d\beta^2$. Matching the means and the variances of T_n and R results in the following expressions:

$$\begin{aligned}
\beta &= \tfrac{1}{p(b-a)} \int_{\mathcal{T}} \int_{\mathcal{T}} tr\left[\mathbf{\Gamma}^*(s,t)\mathbf{\Gamma}^{*T}(s,t)\right] dsdt + o(1), \\
d &= \tfrac{p^2(b-a)^2}{\int_{\mathcal{T}} \int_{\mathcal{T}} tr[\mathbf{\Gamma}^*(s,t)\mathbf{\Gamma}^{*T}(s,t)]dsdt} + o(1).
\end{aligned} \tag{11}$$

Then, to conduct the proposed test T_n, we just need estimate β and d based on the given two samples (1). Using (4), $\boldsymbol{\Gamma}^*(s,t)$ can be estimated by $\hat{\mathbf{\Gamma}}^*(s,t) = \hat{\mathbf{\Gamma}}^{-\frac{1}{2}}(s,s)\hat{\mathbf{\Gamma}}(s,t)\hat{\mathbf{\Gamma}}^{-\frac{1}{2}}(t,t)$. Then by (11), the parameters β and d can be respectively estimated by

$$\begin{aligned}
\hat{\beta} &= \frac{1}{p(b-a)} \int_{\mathcal{T}} \int_{\mathcal{T}} tr\left[\hat{\mathbf{\Gamma}}^*(s,t)\hat{\mathbf{\Gamma}}^{*T}(s,t)\right] dsdt, \\
\hat{d} &= \frac{p^2(b-a)^2}{\int_{\mathcal{T}} \int_{\mathcal{T}} tr\left[\hat{\mathbf{\Gamma}}^*(s,t)\hat{\mathbf{\Gamma}}^{*T}(s,t)\right] dsdt}.
\end{aligned}$$

The proposed test T_n can then be conducted via approximating the null distribution with the distribution of $R \sim \hat{\beta}\chi^2_{\hat{d}}$. That is, we have $T_n \sim \hat{\beta}\chi^2_{\hat{d}}$ approximately.

2.3 Asymptotic Power

In this subsection, we study the asymptotic powers of the proposed tests under the following local alternative hypothesis:

$$H_{1n} : \boldsymbol{\mu}_i(t) = \boldsymbol{\mu}_0(t) + n_i^{-1/2}\boldsymbol{d}_i(t), \quad i = 1, 2, \tag{12}$$

where $\boldsymbol{\mu}_0(t)$ is some fixed function, and $\boldsymbol{d}_i(t),\ i = 1, 2$ are any real functions that are independent of n. Note that under Condition (C2), as $n \to \infty$, the local alternative hypothesis (12) will tend to the null hypothesis with a root-n rate. Therefore, the proposed tests will be root-n consistent if they have asymptotic power one when the information provided by $\boldsymbol{d}_i(t), i = 1, 2$ diverges. For properly describing the asymptotic random expressions of the proposed tests under the local alternative hypothesis (12), let $\lambda_r, r = 1, 2, \ldots, m$ be all the positive decreasing-ordered eigenvalues of $\mathbf{\Gamma}^*(s,t)$ and $\boldsymbol{\phi}_r(t)$ be the associated eigenfunctions where $\boldsymbol{\Gamma}^*(s,t)$ is as defined in (8). Note that when all of the eigenvalues of $\mathbf{\Gamma}^*(s,t)$ are positive, we set $m=\infty$. Let $\boldsymbol{w}(t)$ be defined as in Theorem 1, $\boldsymbol{d}(t)=\boldsymbol{\Gamma}(t,t)^{-1/2}[(1-\tau)\boldsymbol{d}_1(t)-\tau\boldsymbol{d}_2(t)]$, $\delta_r=\int_{\mathcal{T}}\boldsymbol{d}^T(t)\boldsymbol{\phi}_r(t)dt$, $\delta_\lambda=(\sum_{r=1}^m\lambda_r\delta_r^2)^{1/2}$, and $\delta^2=\int_{\mathcal{T}}\|\boldsymbol{d}(t)\|^2dt$. For any given significance level α, let $\hat{T}_n(\alpha)=\hat{\beta}\chi^2_{\hat{d}}(\alpha)$ and $\hat{T}_{n,\max}(\alpha)$ denote the estimated critical values of T_n and $T_{n,\max}$ obtained using the proposed Welch-Satterthwaite χ^2-approximation based approach and the proposed NPB approach, respectively where as usual $\chi^2_{\hat{d}}(\alpha)$ denotes the upper 100α percentile of $\chi^2_{\hat{d}}$. The next two theorems present the root n-consistency of the proposed tests, respectively.

Theorem 3. *Under Conditions (C1)–(C5) and the local alternative hypothesis (12), as* $n\to\infty$*, we have* $T_n\xrightarrow{d}T_1^*$ *with*

$$\begin{gathered}T_1^*\stackrel{d}{=}\int_{\mathcal{T}}\|\boldsymbol{w}(t)+\boldsymbol{d}(t)\|^2dt\stackrel{d}{=}\sum_{r=1}^m\lambda_rB_r+\left(\delta^2-\sum_{r=1}^m\delta_r^2\right),\\ B_r\sim\chi_1^2(\delta_r^2/\lambda_r),r=1,2,\cdots,m\text{ are independent.}\end{gathered}\tag{13}$$

In addition, assume that as $n\to\infty$*, we have* $\hat{T}_n(\alpha)\to C^*(\alpha)<\infty$*. Then the asymptotic power of* T_n *can be expressed as*

$$P\big[T_n>\hat{T}_n(\alpha)\big]=P\big[T^*+2\delta_\lambda Z+\delta^2>C^*(\alpha)\big]+o(1),\tag{14}$$

where T^* *is defined in Theorem 1 and* $Z\sim N(0,1)$*. The right-hand side of (14) will tend to 1 as* $\delta^2\to\infty$*.*

Theorem 4. *Under Conditions (C1)–(C5) and the local alternative (12), as* $n\to\infty$*, we have* $T_{n,\max}\xrightarrow{d}T_{1,\max}^*$ *with*

$$T_{1,\max}^*\stackrel{d}{=}\sup_{t\in\mathcal{T}}\|\boldsymbol{w}(t)+\boldsymbol{d}(t)\|^2.\tag{15}$$

In addition, assume that as $n\to\infty$*, we have* $\hat{T}_{n,\max}(\alpha)\to C^*_{\max}(\alpha)<\infty$*. Then the asymptotic power of* $T_{n,\max}$ *can be expressed as*

$$P\big[T_{n,\max}>\hat{T}_{n,\max}(\alpha)\big]\geq P\big[T^*+2\delta_\lambda Z+\delta^2>(b-a)C^*_{\max}(\alpha)\big]+o(1),\tag{16}$$

where again T^* *is defined in Theorem 1 and* $Z\sim N(0,1)$*. Again, the right-hand side of (16)* *will tend to 1 as* $\delta^2 \to \infty$.

Under some conditions, one may show that $C^*(\alpha)$ and $C^*_{\max}(\alpha)$ are the upper 100α percentiles of T^* and $T^*_{\max}$ respectively where T^* and $T^*_{\max}$ are defined in Theorem 1. Further research in this direction is interesting and warranted.

3 Simulation Studies

In this section, we present a simulation study to compare the finite-sample performance of the proposed T_n and $T_{n,\max}$ tests against the existing competitors for two-sample problems for multivariate functional data. The T_n and $T_{n,\max}$ tests are conducted respectively using the Welch-Satterthwaite χ^2-approximation method and the NPB method as described in Section 2. The existing competitors are those tests considered by Górecki and Smaga (2017), including the Lawley-Hotelling trace, Pillai trace, Roy's maximum root, and Wilks lambda tests, denoted by LH, P, R and W respectively.

We generate two multivariate functional samples from the following model:

$$\begin{aligned}
&\boldsymbol{y}_{ij}(t_{ijl}) = \boldsymbol{\mu}_i(t_{ijl}) + \boldsymbol{\epsilon}_{ij}(t_{ijl}), \quad l = 1, 2, \ldots, M; \quad j = 1, 2, \ldots, n_i; \quad i = 1, 2, \\
&\boldsymbol{\mu}_1(t) = [\mu_{11}(t), \mu_{12}(t)]^T, \ \mu_{11}(t) = [\sin(2\pi t^2)]^5, \ \mu_{12}(t) = \boldsymbol{c}_1^T[1, t, t^2, t^3]^T, \\
&\boldsymbol{\mu}_2(t) = [\mu_{21}(t), \mu_{22}(t)]^T, \mu_{21}(t) = [\sin(2\pi t^2)]^5, \ \mu_{22}(t) = \boldsymbol{c}_2^T[1, t, t^2, t^3]^T, \\
&\boldsymbol{\epsilon}_{ij}(t) = \boldsymbol{A}[z_{ij1}(t), z_{ij2}(t)]^T, z_{ijk}(t) = \boldsymbol{b}_{ijk}^T\Psi(t), \boldsymbol{b}_{ijk} = [b_{ijk1}, b_{ijk2}, \ldots, b_{ijkq}]^T, \\
&b_{ijkr} \stackrel{d}{=} \sqrt{\lambda_r} v_{ijkr}, r = 1, 2, \ldots, q, k = 1, 2, \ j = 1, 2, \ldots, n_i, \ i = 1, 2.
\end{aligned}$$

We consider three cases of the sample sizes: $[n_1, n_2] = [20, 30]$, $[40, 60]$ or $[60, 90]$. The design time points $t_{ijl}, l = 1, 2, \cdots, M$ are equispaced in the interval $[0, 1]$ with $M = 30$ or $M = 50$. The tuning parameters in the mean functions $\boldsymbol{\mu}_1(t)$ and $\boldsymbol{\mu}_2(t)$ are specified as follows. We set $\boldsymbol{c}_1 = [1, 2.3, 3.4, 1.5]^T$ and $\boldsymbol{c}_2 = \boldsymbol{c}_1 + \delta\boldsymbol{u}$ with $\boldsymbol{u} = [1, 2, 3, 4]^T/\sqrt{30}$, where δ is used to control the difference between $\mu_{12}(t)$ and $\mu_{22}(t)$. Note that when $\delta = 0$, the null hypothesis is valid, and when δ takes bigger values, the difference between $\mu_{21}(t)$ and $\mu_{22}(t)$ becomes more significant. To generate the error term $\boldsymbol{\epsilon}_{ij}(t)$, we first set the two rows of the matrix $\boldsymbol{A}$ to be $[1, -1]$ and $[0, 1]$ respectively and then set $\Psi(t) = [\psi_1(t), \ldots, \psi_q(t)]^T$ with $\psi_1(t) = 1, \psi_{2r}(t) =$

Table 1: *Empirical sizes and powers (in %) with their associated standard deviations (in parentheses) of the tests under consideration when $\rho = 0.1$ and when v_{ijkr}'s are i.i.d. from $t_4/\sqrt{2}$.*

		$\delta = 0$		$\delta = 0.1$		$\delta = 0.2$		$\delta = 0.3$		$\delta = 0.4$	
$[n_1, n_2]$	Method	M=30	M=50	M=30	M=50	M=30	M=50	M=30	M=50	M=30	M=50
$[20, 30]$	T_n	5.7(0.23)	6.4(0.24)	17.3(0.38)	17.9(0.39)	56.1(0.50)	58.0(0.49)	89.6(0.31)	91.0(0.29)	99.5(0.07)	99.7(0.05)
	$T_{n,\max}$	3.2(0.18)	4.8(0.21)	32.4(0.47)	37.5(0.48)	93.0(0.26)	94.1(0.24)	99.8(0.04)	99.6(0.06)	100.0(0.00)	100.0(0.00)
	LH	5.1(0.22)	5.8(0.23)	15.0(0.36)	14.8(0.36)	52.1(0.50)	52.8(0.50)	87.4(0.33)	89.0(0.31)	99.3(0.08)	99.7(0.05)
	P	4.9(0.22)	5.7(0.23)	15.2(0.36)	14.8(0.36)	52.7(0.50)	53.3(0.50)	87.6(0.33)	88.9(0.31)	99.2(0.09)	99.7(0.05)
	R	5.0(0.22)	5.7(0.23)	14.9(0.36)	14.3(0.35)	49.1(0.50)	51.3(0.50)	85.8(0.35)	87.7(0.33)	99.3(0.08)	99.7(0.05)
	W	5.0(0.22)	5.7(0.23)	15.0(0.36)	14.8(0.36)	52.5(0.50)	53.0(0.50)	87.7(0.33)	89.0(0.31)	99.2(0.09)	99.7(0.05)
$[40, 60]$	T_n	5.5(0.23)	4.7(0.21)	29.1(0.45)	30.2(0.46)	85.6(0.35)	90.4(0.29)	99.7(0.05)	99.9(0.03)	100.0(0.00)	100.0(0.00)
	$T_{n,\max}$	5.4(0.23)	4.3(0.20)	69.6(0.46)	70.1(0.46)	99.9(0.03)	99.8(0.04)	100.0(0.00)	100.0(0.00)	100.0(0.00)	100.0(0.00)
	LH	5.1(0.22)	4.1(0.20)	26.8(0.44)	28.7(0.45)	83.9(0.37)	89.3(0.31)	99.6(0.06)	99.8(0.04)	100.0(0.00)	100.0(0.00)
	P	5.1(0.22)	4.1(0.20)	26.8(0.44)	28.8(0.45)	83.9(0.37)	89.3(0.31)	99.6(0.06)	99.8(0.04)	100.0(0.00)	100.0(0.00)
	R	4.9(0.22)	4.1(0.20)	25.3(0.43)	26.8(0.44)	82.4(0.38)	87.0(0.34)	99.6(0.06)	99.7(0.05)	100.0(0.00)	100.0(0.00)
	W	5.1(0.22)	4.1(0.20)	26.8(0.44)	28.8(0.45)	83.9(0.37)	89.3(0.31)	99.6(0.06)	99.8(0.04)	100.0(0.00)	100.0(0.00)
$[60, 90]$	T_n	5.1(0.22)	4.8(0.21)	38.4(0.49)	39.7(0.49)	97.3(0.16)	98.7(0.11)	100.0(0.00)	100.0(0.00)	100.0(0.00)	100.0(0.00)
	$T_{n,\max}$	4.4(0.21)	4.2(0.20)	86.3(0.34)	89.2(0.31)	100.0(0.00)	100.0(0.00)	100.0(0.00)	100.0(0.00)	100.0(0.00)	100.0(0.00)
	LH	5.0(0.22)	4.9(0.22)	36.2(0.48)	38.4(0.49)	96.8(0.18)	98.3(0.13)	100.0(0.00)	100.0(0.00)	100.0(0.00)	100.0(0.00)
	P	5.0(0.22)	4.9(0.22)	36.5(0.48)	38.6(0.49)	96.8(0.18)	98.3(0.13)	100.0(0.00)	100.0(0.00)	100.0(0.00)	100.0(0.00)
	R	5.0(0.22)	4.7(0.21)	34.7(0.48)	37.2(0.48)	96.1(0.19)	97.9(0.14)	100.0(0.00)	100.0(0.00)	100.0(0.00)	100.0(0.00)
	W	5.0(0.22)	4.9(0.22)	36.2(0.48)	38.5(0.49)	96.8(0.19)	98.3(0.03)	100.0(0.00)	100.0(0.00)	100.0(0.00)	100.0(0.00)

Table 2: *Empirical sizes and powers (in %) with their associated standard deviations (in parentheses) of the tests under consideration when $\rho = 0.5$ and when v_{ijkr}'s are i.i.d. from $t_4/\sqrt{2}$.*

		$\delta = 0$		$\delta = 0.1$		$\delta = 0.3$		$\delta = 0.5$		$\delta = 0.8$	
$[n_1, n_2]$	Method	M=30	M=50	M=30	M=50	M=30	M=50	M=30	M=50	M=30	M=50
$[20, 30]$	T_n	6.3(0.24)	6.4(0.24)	7.7(0.27)	8.1(0.27)	24.9(0.43)	25.6(0.44)	58.3(0.49)	60.7(0.49)	96.2(0.19)	97.2(0.17)
	$T_{n,\max}$	4.4(0.21)	3.1(0.17)	5.5(0.23)	5.6(0.23)	26.7(0.44)	26.5(0.44)	70.2(0.46)	75.7(0.43)	99.5(0.07)	99.7(0.05)
	LH	4.8(0.21)	5.3(0.22)	6.7(0.25)	6.2(0.24)	20.6(0.40)	22.4(0.42)	54.8(0.50)	56.0(0.50)	95.6(0.21)	95.9(0.20)
	P	4.6(0.21)	5.4(0.23)	6.6(0.25)	6.4(0.24)	20.9(0.41)	22.3(0.42)	54.8(0.50)	55.9(0.50)	95.4(0.21)	95.9(0.20)
	R	4.6(0.22)	5.5(0.23)	7.2(0.26)	6.3(0.24)	21.4(0.41)	22.4(0.42)	54.7(0.50)	57.1(0.50)	95.5(0.21)	95.8(0.20)
	W	4.9(0.22)	5.3(0.22)	6.6(0.25)	6.3(0.24)	20.5(0.40)	22.4(0.42)	54.8(0.50)	55.9(0.50)	95.5(0.21)	95.9(0.20)
$[40, 60]$	T_n	6.0(0.24)	5.4(0.23)	9.9(0.30)	9.8(0.30)	43.2(0.50)	47.6(0.50)	89.1(0.31)	92.0(0.27)	100.0(0.00)	100.0(0.00)
	$T_{n,\max}$	4.8(0.21)	4.1(0.20)	8.8(0.28)	9.1(0.29)	60.3(0.49)	65.0(0.48)	98.0(0.14)	99.4(0.08)	100.0(0.00)	100.0(0.00)
	LH	5.5(0.23)	4.9(0.22)	8.9(0.28)	8.8(0.28)	40.7(0.49)	44.6(0.50)	87.7(0.33)	90.4(0.29)	100.0(0.00)	100.0(0.00)
	P	5.6(0.23)	5.0(0.22)	8.9(0.28)	8.9(0.28)	40.5(0.49)	44.3(0.50)	87.7(0.33)	90.4(0.29)	100.0(0.00)	100.0(0.00)
	R	5.6(0.23)	4.5(0.21)	9.2(0.29)	9.4(0.29)	41.4(0.49)	44.5(0.50)	87.6(0.33)	90.6(0.29)	100.0(0.00)	100.0(0.00)
	W	5.4(0.23)	5.0(0.22)	8.8(0.28)	8.8(0.28)	40.6(0.49)	44.2(0.50)	87.7(0.33)	90.3(0.30)	100.0(0.00)	100.0(0.00)
$[60, 90]$	T_n	5.7(0.23)	4.7(0.21)	10.1(0.30)	11.2(0.32)	60.9(0.49)	65.0(0.48)	97.5(0.16)	99.2(0.09)	100.0(0.00)	100.0(0.00)
	$T_{n,\max}$	5.0(0.22)	5.6(0.23)	10.1(0.30)	12.2(0.33)	83.6(0.37)	86.4(0.34)	99.9(0.03)	99.9(0.03)	100.0(0.00)	100.0(0.00)
	LH	5.0(0.22)	5.0(0.22)	9.0(0.29)	10.4(0.31)	57.8(0.49)	62.2(0.49)	97.2(0.17)	99.0(0.10)	100.0(0.00)	100.0(0.00)
	P	4.9(0.22)	5.0(0.22)	9.0(0.29)	10.3(0.30)	58.1(0.49)	62.2(0.49)	97.1(0.17)	98.9(0.10)	100.0(0.00)	100.0(0.00)
	R	5.8(0.23)	5.3(0.22)	8.8(0.28)	10.2(0.30)	58.8(0.49)	63.6(0.48)	97.5(0.16)	99.2(0.09)	100.0(0.00)	100.0(0.00)
	W	4.9(0.22)	5.0(0.22)	9.0(0.29)	10.4(0.31)	58.1(0.49)	62.2(0.49)	97.2(0.17)	98.9(0.10)	100.0(0.00)	100.0(0.00)

Table 3: *Empirical sizes and powers (in %) with their associated standard deviations (in parentheses) of the tests under consideration when* $\rho = 0.9$ *and when* v_{ijkr}*'s are i.i.d. from* $t_4/\sqrt{2}$.

		$\delta = 0$		$\delta = 0.2$		$\delta = 0.4$		$\delta = 0.6$		$\delta = 0.9$	
$[n_1, n_2]$	Method	M=30	M=50	M=30	M=50	M=30	M=50	M=30	M=50	M=30	M=50
$[20, 30]$	T_n	5.5(0.23)	5.7(0.23)	7.7(0.27)	7.7(0.27)	12.4(0.33)	14.1(0.35)	28.9(0.45)	31.1(0.46)	59.1(0.49)	66.3(0.47)
	$T_{n,\max}$	4.5(0.21)	3.4(0.18)	5.4(0.23)	4.4(0.21)	8.0(0.27)	9.1(0.29)	19.0(0.39)	21.1(0.41)	45.2(0.50)	51.6(0.50)
	LH	5.0(0.22)	5.3(0.22)	6.8(0.25)	7.6(0.27)	11.7(0.32)	13.4(0.34)	28.2(0.45)	29.9(0.46)	58.6(0.49)	65.3(0.48)
	P	5.0(0.22)	4.9(0.22)	6.9(0.25)	7.7(0.27)	11.7(0.32)	13.2(0.34)	27.9(0.45)	30.0(0.46)	57.8(0.49)	64.7(0.48)
	R	4.7(0.21)	4.7(0.21)	6.8(0.25)	7.1(0.26)	13.4(0.34)	14.1(0.35)	28.9(0.45)	30.4(0.46)	61.6(0.49)	66.3(0.47)
	W	5.1(0.22)	5.0(0.22)	6.9(0.25)	7.6(0.27)	11.6(0.32)	13.2(0.34)	28.0(0.45)	29.9(0.46)	58.4(0.49)	65.0(0.48)
$[40, 60]$	T_n	4.8(0.21)	4.6(0.21)	9.6(0.29)	10.4(0.31)	23.1(0.42)	25.1(0.43)	58.0(0.49)	61.4(0.49)	94.4(0.23)	94.4(0.23)
	$T_{n,\max}$	4.5(0.21)	4.2(0.20)	7.4(0.26)	8.0(0.27)	18.4(0.39)	20.7(0.41)	51.3(0.50)	56.1(0.50)	92.5(0.26)	91.9(0.27)
	LH	4.0(0.20)	4.6(0.21)	9.0(0.29)	10.4(0.31)	23.2(0.42)	24.2(0.43)	57.0(0.50)	61.0(0.49)	93.7(0.24)	94.5(0.23)
	P	3.8(0.19)	4.7(0.21)	8.8(0.28)	10.6(0.31)	23.2(0.42)	23.9(0.43)	56.9(0.50)	60.5(0.49)	93.7(0.24)	94.6(0.23)
	R	4.1(0.20)	3.8(0.19)	9.6(0.29)	10.5(0.31)	24.3(0.43)	25.7(0.44)	59.4(0.49)	62.1(0.49)	95.2(0.21)	95.6(0.21)
	W	4.0(0.20)	4.6(0.21)	8.8(0.28)	10.5(0.31)	23.3(0.42)	24.1(0.43)	57.0(0.50)	60.9(0.49)	93.7(0.24)	94.6(0.23)
$[60, 90]$	T_n	5.4(0.23)	5.1(0.22)	11.0(0.31)	12.2(0.33)	36.6(0.48)	40.3(0.49)	77.0(0.42)	79.8(0.40)	99.6(0.06)	99.7(0.05)
	$T_{n,\max}$	5.0(0.22)	4.8(0.21)	10.0(0.30)	9.9(0.30)	33.2(0.47)	35.0(0.48)	76.1(0.43)	74.4(0.44)	98.7(0.11)	99.7(0.05)
	LH	6.1(0.24)	5.2(0.22)	11.0(0.31)	12.8(0.33)	36.0(0.48)	40.0(0.49)	77.0(0.42)	79.0(0.41)	99.2(0.09)	99.7(0.05)
	P	6.1(0.24)	5.2(0.22)	11.0(0.31)	12.8(0.33)	36.1(0.48)	39.9(0.49)	76.9(0.42)	79.1(0.41)	99.2(0.09)	99.7(0.05)
	R	5.8(0.23)	5.0(0.22)	12.0(0.33)	12.5(0.33)	37.4(0.49)	41.2(0.49)	79.8(0.40)	82.7(0.38)	99.6(0.06)	99.9(0.03)
	W	6.2(0.24)	5.2(0.22)	11.0(0.31)	12.8(0.33)	36.2(0.48)	39.9(0.49)	77.0(0.42)	79.0(0.41)	99.2(0.09)	99.7(0.05)

$\sqrt{2}\sin(2\pi rt)$, and $\psi_{2r+1}(t) = \sqrt{2}\cos(2\pi rt)$, $t \in [0,1]$ for $r = 1, 2, \ldots, (q-1)/2$, and $q = 7$. We let $\lambda_r = a\rho^r, r = 1, 2, \ldots, q$ with $a = 1.5$, $q = 7$ and we consider three cases of $\rho = 0.1, 0.5$ or 0.9 so that the components of the simulated functional data have high, moderate, and low correlations. Moreover, we generate v_{ijkr} from three distributions: $N(0,1)$, $t_4/\sqrt{2}$ and $(\chi_4^2 - 4)/(2\sqrt{2})$.

Under each setting, we first generate two multivariate functional samples and then apply $T_n, T_{n,\max}, LH, P, R$ and W tests to them. The p-values for all the tests are then recorded. The null hypothesis is rejected whenever the p-value of a test is smaller than the nominal size 5%. This process is repeated 1000 times so that the empirical sizes or powers of the tests under consideration can be obtained.

The empirical sizes and powers of the tests under various settings for the cases when the error-related terms v_{ijkr} are generated from $t_4/\sqrt{2}$ are presented in Tables 1-3 for $\rho = 0.1, 0.5$ and 0.9 respectively. We have the following conclusions. First of all, it is seen that with the group sample sizes increasing, the empirical sizes of T_n are generally getting closer to the nominal size 5%. This conclusion is consistent with the large sample property of T_n presented in Theorem 1. Secondly, it is seen that under each setting, the empirical powers of T_n are generally larger (see Tables 1 and 2) or not worse (see Table 3) than those of the LH, P, R, and W tests and the empirical powers of $T_{n,\max}$ are much larger (see Tables 1 and 2) or slightly worse (see Table 3) than those of the LH, P, R and W tests. This is expected since our test statistics T_n and $T_{n,\max}$ as defined in (6) are functions of the pointwise Hotelling T^2-test statistic $T_n(t)$ [see (5)] which is most powerful for each given design time point t. Finally, it is seen that in terms of empirical power, $T_{n,\max}$ generally outperforms T_n when $\rho = 0.1$ and $\rho = 0.5$, and it performs less well than T_n when $\rho = 0.9$. This shows that when the components of the multivariate functional data are highly or moderately correlated, $T_{n,\max}$ is preferred and otherwise, T_n is preferred. Alternatively speaking, when the functional data are very noisy, T_n is preferred. This phenomenon is also noticed by Zhang et al. (2019). Overall, we may say that in terms of power, both T_n and $T_{n,\max}$ outperform or perform not worse than the LH, P, R, and W tests. Similar conclusions to the above can also be drawn from the empirical sizes and powers of the tests under various settings for the cases when the error-related terms v_{ijkr} are generated from $N(0,1)$ and $(\chi_4^2 - 4)/(2\sqrt{2})$ respectively. We thus shall not present these empirical sizes and powers for space saving.

4 Applications to the Canadian weather data

A brief introduction of the Canadian weather data is given in Section 1. Figure 1 displays the raw temperature and precipitation curves for the 35 Canadian weather stations among which there are 15 located at Eastern Canada, 15 at Western Canada, and the remaining 5 at Northern Canada. In the literature, many authors aim to check whether there is a

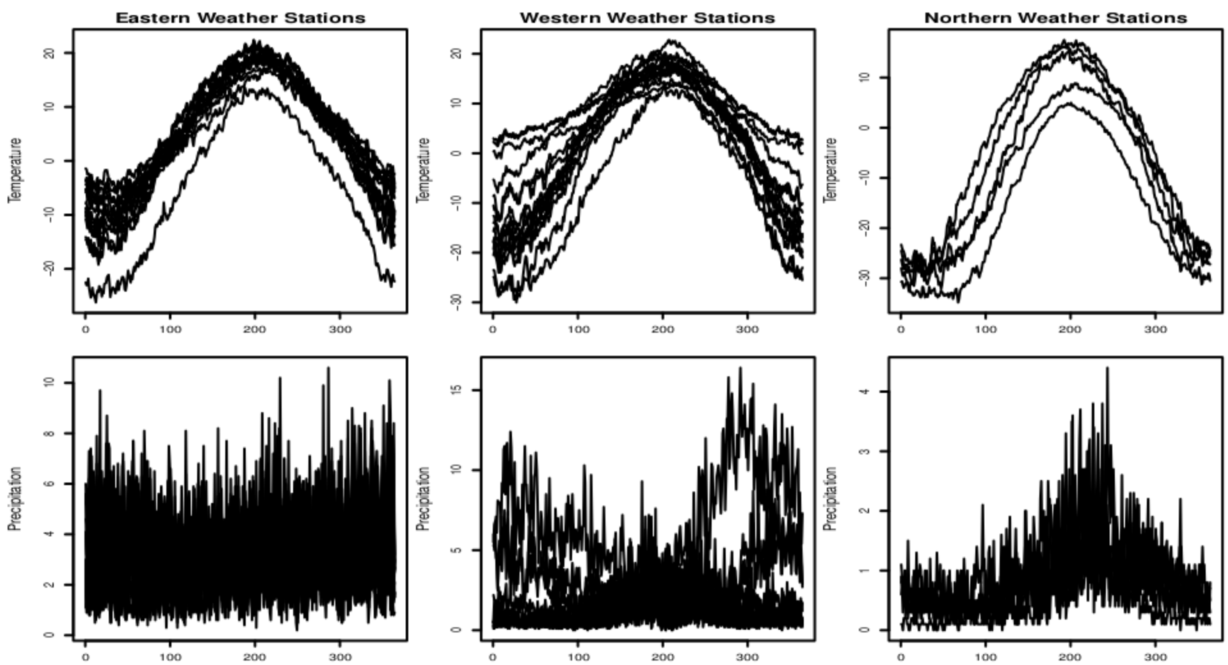

Figure 1: *Canadian weather data for* 15 *Eastern weather stations (left panels),* 15 *Western weather stations (middle panels), and* 5 *Northern weather stations (right panels).*

location effect among the mean temperature curves of the Eastern, Western and Northern weather stations. This is a one-way ANOVA problem for univariate functional data. In this section, we aim to illustrate the proposed T_n and $T_{n,\max}$ tests for two-sample problems for multivariate functional data. Therefore, the precipitation curves of the 35 weather stations are also involved. For illustrations only, we apply the T_n and $T_{n,\max}$ tests, together with the LH, P, R and W tests of Górecki and Smaga (2017), to check the equality of the mean temperature and precipitation curves of the weather stations at two different locations: (1) Eastern versus Western, (2) Eastern versus Northern, and (3) Western versus Northern.

Table 4: *P-values of* $T_n, T_{n,\max}, LH, P, R$ *and* W *tests applied to the Canadian Weather Data.*

Two-sample Problem	T_n	$T_{n,\max}$	LH	P	R	W
Eastern vs Western	0.0092	0.054	0.028	0.028	0.040	0.028
Eastern vs Northern	0.0000	0.016	0.000	0.000	0.000	0.000
Western vs Northern	0.0000	0.104	0.000	0.000	0.000	0.000

The P-values of the tests are presented in Table 4. It is seen that for all three two-sample problems, the P-values of T_n are smaller than those of LH, P, R and W, showing that T_n may be more powerful than LH, P, R and W in detecting the location effect on the temperature and precipitation curves of the Canadian weather stations simultaneously. However, it is also seen that the test results of $T_{n,\max}$ are less significant than other tests and in particular, for the first and last two-sample problems, $T_{n,\max}$ fails to reject the associated null hypotheses. This seems not a surprise since the temperature and precipitation curves of the Canadian weather stations are very noisy as seen from Figure 1 and the simulation results in 3 show that when the functional data are very noisy, $T_{n,\max}$ is less powerful.

5 Concluding Remarks

In this article, two global tests for two-sample problems for multivariate functional data are proposed and studied. The asymptotic random expressions of the test statistics under the null and some local alternative hypotheses are derived. Two methods, a Welch-Satterthwaite χ^2-approximation method and a nonparametric bootstrap method, for approximating the null distributions of the test statistics are described. A simulation study and some real data applications indicate that in terms of power, the proposed tests generally outperform the existing approaches proposed in Górecki and Smaga (2017).

The simulation results presented in Tables 1–3 indicate that when the functional data are very noisy, $T_{n,\max}$ is less powerful. To overcome this problem, one may smooth the functions using some smoothing techniques so that the resulting functional data are less noisy. A study of this problem is interesting and warranted.

Note that in the Welch-Satterthwaite χ^2-approximation method, the usual sample covariance function matrix is used to estimate the underlying covariance function matrix directly. The resulting estimators of the approximation parameters are biased. When the components of the multivariate functional data are less correlated, the bias effect may not be ignorable. Further study on how to obtain less-biased estimators of the approximation parameters is interesting and warranted.

Note also that our test statistics are constructed via globalizing the pointwise Hotelling T^2-test which is the most powerful for any given design time point. This partially explains in some degree why our tests generally outperform the tests proposed by Górecki and Smaga (2017). It is then of interest to ask if the same idea can be used to construct some global tests for k-sample problems for multivariate functional data via globalizing the pointwise Lawley-Hotelling trace test. Further study in this direction is also interesting and warranted.

Acknowledgement

Qiu's work was supported by the Education and Scientific Research Projects of Young and Middle-aged Teachers in Fujian Province, China (Grant No. JAT160027), the Natural Science Foundation of Fujian Province, China (Grant No. 2017J01002), and MOE (Ministry of Education in China) Project of Humanities and Social Sciences (Grant No. 19YJC910004). Zhang's work was supported by the National University of Singapore Academic Research grant R-155-000-212-114.

Appendix

Proof of Theorem 1. Set $\boldsymbol{w}_n(t) = (\frac{n_1 n_2}{n})^{1/2}\hat{\boldsymbol{\Gamma}}(t,t)^{-\frac{1}{2}}[\bar{\boldsymbol{y}}_{1\cdot}(t) - \bar{\boldsymbol{y}}_{2\cdot}(t)]$. Then it can be expressed as

$$\boldsymbol{w}_n(t) = \boldsymbol{w}_{n,0}(t) + (\frac{n_1 n_2}{n})^{1/2}\hat{\boldsymbol{\Gamma}}(t,t)^{-\frac{1}{2}}[\boldsymbol{\mu}_1(t) - \boldsymbol{\mu}_2(t)], \tag{17}$$

where $\boldsymbol{w}_{n,0}(t) = \sqrt{\frac{n_2}{n}}\hat{\boldsymbol{\Gamma}}(t,t)^{-1/2}\sqrt{n_1}\big(\bar{\boldsymbol{y}}_{1\cdot}(t) - \boldsymbol{\mu}_1(t)\big) - \sqrt{\frac{n_1}{n}}\hat{\boldsymbol{\Gamma}}(t,t)^{-1/2}\sqrt{n_2}\big(\bar{\boldsymbol{y}}_{2\cdot}(t) - \boldsymbol{\mu}_2(t)\big)$. Then $T_n = \int_{\mathcal{T}} \|\boldsymbol{w}_n(t)\|^2 dt$ and $T_{n,\max} = \sup_{t\in\mathcal{T}} \|\boldsymbol{w}_n(t)\|^2$. Under the null hypothesis, we have

$\boldsymbol{w}_n(t) = \boldsymbol{w}_{n,0}(t)$ and under the given conditions and the null hypothesis, by the central limit theorem of i.i.d. stochastic processes (Vaart and Wellner, 1996), we have $\sqrt{n_i}[\bar{\boldsymbol{y}}_{i\cdot}(t) - \boldsymbol{\mu}_1(t)] \xrightarrow{d} \mathbf{GP}_p(\mathbf{0}, \boldsymbol{\Gamma}), i = 1, 2$ and we can show that $\hat{\boldsymbol{\Gamma}}(s,t) \xrightarrow{p} \boldsymbol{\Gamma}(s,t)$. It follows that

$$\boldsymbol{w}_{n,0}(t) \xrightarrow{d} \boldsymbol{w}(t) \sim \mathbf{GP}_p(\mathbf{0}, \boldsymbol{\Gamma}^*) \tag{18}$$

where $\boldsymbol{\Gamma}^*(s,t)$ is defined as in (8). Then by the continuous mapping theorem, we have $T_n \xrightarrow{d} T^* = \int_{\mathcal{T}} \|\boldsymbol{w}(t)\|^2 dt$ and $T_{n,\max} \xrightarrow{d} T^*_{\max} = \sup_{t \in \mathcal{T}} \|\boldsymbol{w}(t)\|^2$.

To show the second expression of T^*, note that under Condition (C1), $\boldsymbol{\Gamma}^*(s,t)$ has the following singular value decomposition,

$$\boldsymbol{\Gamma}^*(s,t) = \sum_{r=1}^{\infty} \lambda_r \boldsymbol{\phi}_r(s) \boldsymbol{\phi}_r^T(t), \tag{19}$$

where λ_r's are eigenvalues of $\boldsymbol{\Gamma}^*(s,t)$ and $\boldsymbol{\phi}_r(t)$'s are the associated orthonormal eigenfunctions. Then it follows that $\boldsymbol{w}(t) = \sum_{r=1}^{\infty} \xi_r \boldsymbol{\phi}_r(t)$, where $\xi_r = \int_{\mathcal{T}} \boldsymbol{w}^T(t) \boldsymbol{\phi}_r(t) dt \sim N(0, \lambda_r)$. Therefore,

$$T^* = \int_{\mathcal{T}} \|\boldsymbol{w}(t)\|^2 dt = \sum_{r=1}^{\infty} \xi_r^2 \stackrel{d}{=} \sum_{r=1}^{\infty} \lambda_r A_r, \tag{20}$$

where $A_r \stackrel{i.i.d.}{\sim} \chi_1^2$. This concludes the proof of Theorem 1.

Proof of Theorem 2 . Under Conditions (C1)–(C5) and the null hypothesis, by Theorem 1, we have $\mathrm{E}(T_n) = \mathrm{E}(T^*) + o(1)$ and $\mathrm{Var}(T_n) = \mathrm{Var}(T^*) + o(1)$. By (20), we have $\mathrm{E}(T^*) = \sum_{r=1}^{\infty} \lambda_r$ and $\mathrm{Var}(T^*) = 2\sum_{r=1}^{\infty} \lambda_r^2$. On the one hand, by (8), we have $\boldsymbol{\Gamma}^*(t,t) = \boldsymbol{I}_p$ and $\mathcal{T} = [a,b]$. Thus, $\int_{\mathcal{T}} \mathrm{tr}\left[\boldsymbol{\Gamma}^*(t,t)\right] dt = p(b-a)$. On the other hand, by (19), we have

$$\begin{aligned}
\int_{\mathcal{T}} \mathrm{tr}\left[\boldsymbol{\Gamma}^*(t,t)\right] dt &= \sum_{r=1}^{\infty} \lambda_r \int_{\mathcal{T}} \boldsymbol{\phi}_r^T(t) \boldsymbol{\phi}_r(t) dt = \sum_{r=1}^{\infty} \lambda_r, \\
\int_{\mathcal{T}} \mathrm{tr}\left[\boldsymbol{\Gamma}^*(s,t) \boldsymbol{\Gamma}^{*T}(s,t)\right] dt &= \int_{\mathcal{T}} \int_{\mathcal{T}} \mathrm{tr}\Big[\sum_{r=1}^{\infty} \lambda_r \boldsymbol{\phi}_r(s) \boldsymbol{\phi}_r^T(t) \sum_{l=1}^{\infty} \lambda_l \boldsymbol{\phi}_l(t) \boldsymbol{\phi}_l^T(s)\Big] ds dt \\
&= \sum_{r=1}^{\infty} \sum_{l=1}^{\infty} \lambda_r \lambda_l \int_{\mathcal{T}} \boldsymbol{\phi}_r^T(t) \boldsymbol{\phi}_l(t) dt \int_{\mathcal{T}} \boldsymbol{\phi}_l^T(s) \boldsymbol{\phi}_r(s) ds
\end{aligned}$$

$$= \sum_{r=1}^{\infty} \lambda_r^2.$$

It follows that $\mathrm{E}(T^*) = \int_{\mathcal{T}} \mathrm{tr}\left[\boldsymbol{\Gamma}^*(t,t)\right] dt = p(b-a)$ and $\mathrm{Var}(T^*) = \int_{\mathcal{T}} \mathrm{tr}\left[\boldsymbol{\Gamma}^*(s,t)\boldsymbol{\Gamma}^{*T}(s,t)\right] dsdt$ as desired. The theorem is then proved.

Proof of Theorem 3. By (17) and (12), we have $\boldsymbol{w}_n(t) = \boldsymbol{w}_{n,0}(t) + \boldsymbol{d}_n(t)$ where $\boldsymbol{d}_n(t) = \hat{\boldsymbol{\Gamma}}(t,t)^{-1/2}\left[\sqrt{n_2/n}\boldsymbol{d}_1(t) - \sqrt{n_1/n}\boldsymbol{d}_2(t)\right]$. As $n \to \infty$, under Conditions (C1)–(C5), we have $\boldsymbol{d}_n(t) \xrightarrow{d} \boldsymbol{d}(t)$ where $\boldsymbol{d}(t) = \boldsymbol{\Gamma}(t,t)^{-1/2}\left[\sqrt{1-\tau}\boldsymbol{d}_1(t) - \sqrt{\tau}\boldsymbol{d}_2(t)\right]$. By (18), we have $\boldsymbol{w}_{n,0}(t) \xrightarrow{d} \boldsymbol{w}(t) \sim \mathbf{GP}_p(\boldsymbol{0}, \boldsymbol{\Gamma}^*)$. It follows that

$$\boldsymbol{w}_n(t) \xrightarrow{d} \boldsymbol{w}(t) + \boldsymbol{d}(t) \sim \mathbf{GP}_p(\boldsymbol{d}, \boldsymbol{\Gamma}^*). \tag{21}$$

By the continuous mapping theorem, we have $T_n = \int_{\mathcal{T}} \|\boldsymbol{w}_n(t)\|^2 dt \xrightarrow{d} T_1^*$ where $T_1^* = \int_{\mathcal{T}} \|\boldsymbol{w}(t) + \boldsymbol{d}(t)\|^2 dt$.

We now show the second expression of (13). Since $\boldsymbol{\Gamma}^*(s,t)$ has the singular value decomposition (19), we can write $\boldsymbol{w}(t) + \boldsymbol{d}(t) = \sum_{r=1}^{\infty} \zeta_r \boldsymbol{\phi}_r(t)$, where $\zeta_r = \int_{\mathcal{T}} \left[\boldsymbol{w}(t) + \boldsymbol{d}(t)\right]^T \boldsymbol{\phi}_r(t) dt \sim N(\delta_r, \lambda_r)$, where $\delta_r = \int_{\mathcal{T}} \mathbf{d}^T(t)\boldsymbol{\phi}_r(t) dt$. Therefore,

$$\begin{aligned} T_1^* &= \int_{\mathcal{T}} \|\boldsymbol{w}(t) + \boldsymbol{d}(t)\|^2 dt = \int_{\mathcal{T}} \|\sum_{r=1}^{\infty} \zeta_r \boldsymbol{\phi}_r^T(t)\|^2 dt \\ &= \sum_{r=1}^{m} \zeta_r^2 + \sum_{r=m+1}^{\infty} \zeta_r^2 \stackrel{d}{=} \sum_{r=1}^{m} \lambda_r B_r + \sum_{r=m+1}^{\infty} \delta_r^2 \\ &= \sum_{r=1}^{m} \lambda_r B_r + \left(\delta^2 - \sum_{r=1}^{m} \delta_r^2\right), \end{aligned}$$

where $B_r = \zeta_r^2/\lambda_r \sim \chi_1^2(\delta_r^2/\lambda_r), r = 1, 2, \cdots, m$ and $\delta^2 = \int_{\mathcal{T}} \|\boldsymbol{d}(t)\|^2 dt = \sum_{r=1}^{\infty} \delta_r^2$. This concludes the proof of the second expression of (13).

Note that we can write $B_r \stackrel{d}{=} \left(z_r + \delta_r/\sqrt{\lambda_r}\right)^2 \stackrel{d}{=} A_r + 2z_r\delta_r/\sqrt{\lambda_r} + \delta_r^2/\lambda_r$, where $A_r \sim \chi_1^2$ and $z_r \sim N(0,1)$. It follows that

$$T_1^* \stackrel{d}{=} \sum_{r=1}^{m} \lambda_r A_r + 2\sum_{r=1}^{m} \lambda^{1/2}\delta_r z_r + \delta^2 \stackrel{d}{=} T^* + 2\delta_\lambda Z + \delta^2,$$

where $T^* = \sum_{r=1}^m \lambda_r A_r = \sum_{r=1}^\infty \lambda_r A_r$ as in Theorem 1, $\delta_\lambda = [\sum_{r=1}^m \lambda_r \delta_r^2]^{1/2}$ and $Z = \sum_{r=1}^m \lambda_r^{1/2} \delta_r z_r / \delta_\lambda \sim N(0,1)$. Thus, under the given conditions, we have

$$\begin{aligned} P\big[T_n > \hat{T}_n(\alpha)\big] &= P\big[T^* + 2\delta_\lambda Z + \delta^2 > C^*(\alpha)\big] + o(1) \\ &= P(T^* + 2\delta_\lambda Z > C^*(\alpha) - \delta^2) + o(1). \end{aligned}$$

Then by the above expression, when $\delta_\lambda < \infty$, it is obvious that $P\big[T_n > \hat{T}_n(\alpha)\big] \to 1$ as $\delta^2 \to \infty$ and when $\delta_\lambda \to \infty$, we have

$$P\big[T_n > \hat{T}_n(\alpha)\big] = P\big[Z > -\frac{\delta^2}{2\delta_\lambda}\big] + o(1) \geq P\big[Z > -\frac{\delta}{2\sqrt{\lambda_1}}\big] + o(1) \to 1,$$

as both n and δ tend to ∞ where the last inequality holds since $\delta_\lambda^2 \leq \lambda_1 \sum_{r=1}^m \delta_r^2 \leq \lambda_1 \delta^2$. The theorem is then proved.

Proof of Theorem 4 . Under the given conditions, the expression (15) follows from the expression (21) and the continuous mapping theorem. To show the asymptotic power expression (16) of $T_{n,\max}$, note that by comparing the expressions of T_n and $T_{n,\max}$, we have $T_n \leq (b-a)T_{n,\max}$ since $\mathcal{T} = [a,b]$. Then under the given conditions and by the proof of Theorem 3, as $n \to \infty$ and $\delta^2 \to \infty$, we have

$$\begin{aligned} P\big[T_{n,\max} > \hat{T}_{\max}(\alpha)\big] &\geq P\big[T_n > (b-a)\hat{T}_{\max}(\alpha)\big] \\ &= P\big[T^* + 2\delta_\lambda Z + \delta^2 > (b-a)C^*_{\max}(\alpha)\big] + o(1) \to 1. \end{aligned}$$

This completes the proof of Theorem 4.

References

Antoch, J., Prchal, L., Rosa, M. R. D., and Sarda, P. (2010). Electricity consumption prediction with functional linear regression using spline estimators. *Journal of Applied Statistics*, 37(12):2027–2041.

Berrendero, J. R., Justel, A., and Svarc, M. (2011). Principal components for multivariate functional data. *Computational Statistics and Data Analysis*, 55(9):2619–2634.

Boente, G., Barrera, M. S., and Tyler, D. E. (2014). A characterization of elliptical distributions and some optimality properties of principal components for functional data. *Journal of Multivariate Analysis*, 131(131):254–264.

Chiou, J.-M., Chen, Y.-T., and Yang, Y.-F. (2014). Multivariate functional principal component analysis: a normalization approach. *Statistica Sinica*, 24:1571–1596.

Cox, D. D. and Lee, J. S. (2008). Pointwise testing with functional data using the westfall–young randomization method. *Biometrika*, 95(3):621–634.

Cuevas, A., Febrero, M., and Fraiman, R. (2004). An ANOVA test for functional data. *Computational Statistics and Data Analysis*, 47(1):111–122.

Fan, J. and Lin, S.-K. (1998). Test of significance when data are curves. *Journal of the American Statistical Association*, 93(443):1007–1021.

Faraway, J. J. (1997). Regression analysis for a functional response. *Technometrics*, 39(3):254–261.

Febrero-Bande, M. and González-Manteiga, W. (2013). Generalized additive models for functional data. *Test*, 22(2):278–292.

Ferraty, F. (2011). *Recent Advances in Functional Data Analysis and Related Topics*. Springer, New York.

Fremdt, S., Horvth, L., Kokoszka, P., and Steinebach, J. G. (2014). Functional data analysis with increasing number of projections. *Journal of Multivariate Analysis*, 124(2):313–332.

Górecki, T. and Smaga, Ł. (2015). A comparison of tests for the one-way ANOVA problem for functional data. *Computational Statistics*, 30(4):987–1010.

Górecki, T. and Smaga, Ł. (2017). Multivariate analysis of variance for functional data. *Journal of Applied Statistics*, 44:1–18.

Guo, J., Zhou, B., and Zhang, J.-T. (2019). New tests for equality of several covariance functions for functional data. *Journal of the American Statistical Association*, 114(527):1251–1263.

Jacques, J. and Preda, C. (2014). Model-based clustering for multivariate functional data. *Computational Statistics and Data Analysis*, 71(3):92–106.

Keser, I. K. and Kocakoç, I. D. (2015). Smoothed functional canonical correlation analysis of humidity and temperature data. *Journal of Applied Statistics*, 42(10):2126–2140.

King, D. (2009). Canonical correlation analysis of functional data. *Dissertations and Theses - Gradworks*, 11.

Lian, H. (2014). Some asymptotic properties for functional canonical correlation analysis. *Journal of Statistical Planning and Inference*, 153(5):1–10.

Montoya, E. L. (2013). On the relative efficiency of a monotone parameter curve estimator in a functional nonlinear model. *Statistics and Computing*, 23(3):425–436.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer, New York.

Shen, Q. and Faraway, J. (2004). An F test for linear models with functional responses. *Statistica Sinica*, pages 1239–1257.

Smaga, Ł. and Zhang, J.-T. (2019). Linear hypothesis testing with functional data. *Technometrics*, 61(1):1–12.

Tokushige, S., Yadohisa, H., and Inada, K. (2007). Crisp and fuzzy k -means clustering algorithms for multivariate functional data. *Computational Statistics*, 22(1):1–16.

Vaart, A. W. V. D. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

Zhang, C., Peng, H., and Zhang, J.-T. (2010). Two samples tests for functional data. *Communications in Statistics - Theory and Methods*, 39(4):559–578.

Zhang, J.-T. (2011). Statistical inferences for linear models with functional response. *Statistica Sinica*, 21:1431–1451.

Zhang, J.-T. (2013). *Analysis of variance for functional data*. Chapman and Hall/CRC.

Zhang, J.-T. and Chen, J. (2007). Statistical inferences for functional data. *Annals of Statistics*, 35(3):1052–1079.

Zhang, J.-T., Cheng, M. Y., Wu, H. T., and Bu, Z. (2019). A new test for functional one-way ANOVA with applications to ischemic heart screening. *Computational Statistics and Data Analysis*, 132:3–17.

Zhang, J.-T. and Liang, X. (2014). One-way ANOVA for functional data via globalizing the pointwise F-test. *Scandinavian Journal of Statistics*, 41(1):51–71.