Statistica Sinica **30** (2020), 1095-1116
doi:https://doi.org/10.5705/ss.202017.0262

HYPOTHESIS TESTING FOR MULTIPLE MEAN AND CORRELATION CURVES WITH FUNCTIONAL DATA

Ao Yuan[1], Hong-Bin Fang[1], Haiou Li[1], Colin O. Wu[2] and Ming T. Tan[1]

[1] *Georgetown University Medical Center and* [2] *National Institutes of Health*

Abstract: Existing statistical methods for functional data analyses tend to use local smoothing estimators or some known basis approximations. In many applications with functional observations, the main objectives of statistical inferences are to test (a) the equivalence of a set of unknown mean curves, and (b) the correlation curves of the unknown stochastic processes. However, the unknown curves of the functional data might not be approximated by a set of known basis functions. We propose a class of simple test statistics for comparing the mean and correlation curves of functional data, without relying on an estimation using local smoothing or basis approximations, and study their basic asymptotic properties. Then, we apply the proposed method to functional gene expression data, showing that it yields practical and meaningful results with minimal assumptions. Numerical justifications of our testing method are provided by a simulation study.

Key words and phrases: Comparing mean curves, correlation curve, functional data, gene expression profile, testing functional equivalence, uniform normed convergence.

1. Introduction

Functional data analyses have been used extensively in biomedical studies to evaluate multiple mean and correlation curves over time. Nonparametric analyses of functional data include the estimation and hypothesis testing of the unknown curves, without relying on potentially unrealistic parametric assumptions. A popular approach for nonparametric inferences with functional data is to assume that the unknown curves belong to a "structured regression model" (Hastie, Tibshirani and Friedman (2009)), which can be approximated by some expansions of a class of known basis functions. As a result, the estimation and testing procedures can be constructed using the unknown coefficients of the basis expansions. Basis approximation methods have been proposed by Shi, Weiss and Taylor (1996); Chiang, Rice and Wu (2001); Huang, Wu and Zhou (2002); Müller and Yao (2008) and Li, Wang and Carroll (2010), among others. When the objective is only a nonparametric estimation of the unknown curves, local smoothing methods, such as local polynomials or smoothing splines, are often used in conjunction with structured regression models; for example, see Hoover et al. (1998); Wu and Chiang (2000); Chiang, Rice and Wu (2001) and Fan and Zhang (2000). Cai and Yuan (2011) studied the optimal convergence rate of general estimators and a penalized smoothing estimator of the mean function. Kim and Zhao (2013) introduced two self-normalization methods to overcome the kernel smooth-convergence issue for the sparse and dense longitudinal model. Local polynomials and smoothing splines are often used to estimate functional curves. However, to the best of our knowledge, few studies have examined the statistical properties of hypothesis testing with functional data.

In many biological studies, biological characteristics are measured over time. Thus, we need to be able to compare the curves of various biological characteristics and their correlations, but without *a priori* knowledge of suitable basis functions with which to approximate the curves. For example, the objectives of a temporal gene expression (TGE) study might include testing whether different genes have the same mean expression profiles, or whether some gene expression profiles are correlated.

Two sample hypothesis tests for functional data have been proposed. To test the differences in the mean functions, Zhang and Chen (2007) applied a local polynomial kernel smoothing technique and constructed an L^2-norm-based global test statistic. Cao et al. (2016) constructed a polynomial spline confidence band for mean curves, and Degras (2011, 2017) constructed simultaneous confidence bands. The hypothesis test problem is addressed by testing whether the bands for a difference contain the zero function. Wang et al. (2017) considered unified empirical likelihood ratio tests based on a functional concurrent linear model. An extension to k independent samples of curves was provided by Cuevas, Febrero and Fraiman (2004) who introduced an ANOVA-like test.

In order to test the equality of the distributions of two sets of curves, Hall and Keilegom (2007) proposed a Cramer–von Mises (CVM)-type test based on a second-order smoother. Estévez-Pérez and Vilar (2008) extended the method to k sample cases and applied it to air quality data. Benko, Härdle and Kneip (2009) studied this problem by first introducing a functional principal components decomposition, along with a bootstrap. Pomann, Staicu and Gosh (2016) also decomposed the curves using a functional principle component analysis (FPCA), and applied the Anderson–Darling statistic. Some studies focus on detecting differences in the covariance structures of curves. Gaines, Kaphle and Ruymgaart (2011) used a likelihood ratio-type approach. Fremdt et al. (2012) developed a chi-square asymptotic test statistic based on an FPCA. A regularized M-test based on the Hilbert–Schmidt norm was introduced by Kraus and Panaretos (2012). Horváth and Kokoszka (2012) describe an estimation and testing approach for mean and covariance curves, in which they assume that subjects are observed over a very dense set of time points. As a result, they are able to construct statistical inferences if the entire curves are observed. This assumption is unrealistic in most studies. The above methods can surely be applied in practice with moderately dense functional data, including the method proposed here. However, both will produce inaccurate inferences if the observation grid is not sufficiently dense.

We study the problem of testing the equivalence of mean curves or correlation curves when the curves are observed at a set of time design points chosen based on the scientific objectives. In addition, the estimated curves can be constructed through interpolations at the observed time points. As pointed out by Cai and Yuan (2011, p.2332) when estimating the mean function, smoothing (i.e. a more complicated method) does not result in an improved convergence rate. Therefore, we propose a simple empirical linear interpolation mean as an estimator of the mean curve to test the difference in the means, and for correlation functions between two curves. The asymptotic results of our testing procedures are more general, in the stronger uniform sense, than those given in Horváth and Kokoszka (2012). The proposed method offers two main advantages. First, existing methods first transform the observations into smoothed curves, and then use these curves as functional data. Thus the results may deviate more or less from the truth. In contrast, we use the observed raw data, which is particularly important when the number of observation time points is not big, as in our problem. Second, the proposed method uses empirical mean functions to test the null hypothesis, making it simple to use. In comparison, most existing methods use splines, basis functions, reproducing kernels, kernel smoothers, and so on, which are not as simple to use and require additional assumptions. The main limitation of the proposed method is that it might not be applicable to problems outside the mean, such as those involving a functional regression.

We describe the data and our testing procedures in Section 2, and investigate the asymptotic properties of our proposed test statistics in Section 3. The finite-sample properties of our test statistics are investigated using data from a TGE study in Section 4, and by means of a simulation study in Section 5. We conclude with a discussion in Section 6. The online Supplementary Material contains the proofs of the main results.

2. The Data and Hypothesis Testing

2.1. The data

Functional data can be viewed as observed stochastic processes on the real line, or as random functions in some functional spaces (e.g., Ramsay and Silverman (2005)). For bivariate functional data, we consider stochastic processes $\{(X(t), Y(t))^T : t \in \mathcal{T}\}$, where, given $t \in \mathcal{T}$, a bounded subset in $[0, \infty)$, $X(t)$ and $Y(t)$ are real-valued random variables, which may be correlated. For each fixed $t \in \mathcal{T}$, let $\mu(t) = E[X(t)]$ and $\eta(t) = E[Y(t)]$ be the mean curves of $X(t)$ and $Y(t)$, respectively, and

$$R(t) = \frac{E\{[X(t) - \mu(t)][Y(t) - \eta(t)]\}}{\sigma_{X(t)}\sigma_{Y(t)}} \tag{2.1}$$

be the correlation curve of $X(t)$ and $Y(t)$, where $\sigma_{X(t)}$ and $\sigma_{Y(t)}$ are the corresponding standard deviations of $X(t)$ and $Y(t)$. Because t changes within $\mathcal{T}$, $\mu(t)$ and $\eta(t)$, $\sigma_{X(t)}$ and $\sigma_{Y(t)}$, and $R(t)$ are the curves for the means, standard deviations, and correlation coefficient, respectively, over $t \in \mathcal{T}$.

In real applications, subjects are assumed to be independent, but some subjects may have only $X(\cdot)$ or $Y(\cdot)$ observed. Therefore, we denote by $\mathcal{S}_X$ the set of subjects with observations of $X(\cdot)$ only, $\mathcal{S}_Y$ the set of subjects with observations of $Y(\cdot)$ only, and $\mathcal{S}_{XY}$ the set of subjects with observations of $(X(\cdot), Y(\cdot))^T$. Let n_x, n_y, and n_{xy} be the numbers of subjects in $\mathcal{S}_X$, $\mathcal{S}_Y$, and $\mathcal{S}_{XY}$, respectively. The number of subjects with observations of $X(\cdot)$, that is, in $\mathcal{S}_X \cup \mathcal{S}_{XY}$, is $n_1 = n_x + n_{xy}$. The number of subjects with observations of $Y(\cdot)$, that is, in $\mathcal{S}_Y \cup \mathcal{S}_{XY}$, is $n_2 = n_y + n_{xy}$. The total number of subjects is $n = n_x + n_y + n_{xy} = n_1 + n_2 - n_{xy}$. For simplicity, we assume that our observations of $X(\cdot)$ and $Y(\cdot)$ are made at $k(n)$ distinct time points $t_1 < \cdots < t_{k(n)}$. Our results can be generalized directly to cases of $X(\cdot)$ and $Y(\cdot)$ observed at different time points, but at the expense of more complex notation. The observations for $\{X(t) : t \in \mathcal{T}\}$, $\{Y(t) : t \in \mathcal{T}\}$, and $\{(X(t), Y(t))^T : t \in \mathcal{T}\}$ are given by

$$\begin{cases} \mathbb{X}_{n_1} = \{X_{i,j} = X_i(t_j) + \epsilon_i(t_j) : i = 1, \ldots, n_1; j = 1, \ldots, k(n)\}, \\ \mathbb{Y}_{n_2} = \{Y_{i,j} = Y_i(t_j) + \xi_i(t_j) : i = n_x + 1, \ldots, n_x + n_2; j = 1, \ldots, k(n)\}, \\ (\mathbb{X}_{n_{xy}}, \mathbb{Y}_{n_{xy}}) = (X_{i,j}, Y_{i,j})^T : i = n_x + 1, \ldots, n_x + n_{xy}; j = 1, \ldots, k(n)\}, \end{cases} \tag{2.2}$$

where $\epsilon_i(\cdot)$ and $\xi_i(\cdot)$ are measurement noises independent of $X_i(\cdot)$ and $Y_i(\cdot)$, $\epsilon_i(\cdot)$ is independent and identically distributed (i.i.d.) $\epsilon(\cdot)$, and $\xi_i(\cdot)$ is i.i.d. $\xi(\cdot)$, with the typical assumption that $E[\epsilon_i(t)] = E[\xi_i(t)] = 0$, for all $t \in \mathcal{T}$. Clearly, the subjects in $\left(\mathbb{X}_{n_{xy}}, \mathbb{Y}_{n_{xy}}\right)$ are the subjects in both $\mathbb{X}_{n_1}$ and $\mathbb{Y}_{n_2}$.

Let $X_i(t)$ and $Y_i(t)$ be the unknown subject-specific curves of $X(t)$ and $Y(t)$, respectively, for the ith subject. To define the curve observations of $X_i(t)$ and $Y_i(t)$ at any $t_1 \le t \le t_{k(n)}$ based on $\{\mathbb{X}_{n_1}, \mathbb{Y}_{n_2}\}$ of (2.2), we consider those observations based on the linear interpolations

$$\begin{cases} \mathcal{X}_{n_1} = \left\{X_{i,k(n)}(t) : i = 1, \ldots, n_1\right\}, \\ \mathcal{Y}_{n_2} = \left\{Y_{i,k(n)}(t) : i = n_x + 1, \ldots, n_x + n_2\right\}, \\ \left(\mathcal{X}_{n_{xy}}, \mathcal{Y}_{n_{xy}}\right) = \left\{\left(X_{i,k(n)}(t), Y_{i,k(n)}(t)\right)^T : i = n_x + 1, \ldots, n_x + n_{xy}\right\}, \end{cases} \tag{2.3}$$

where $X_{i,k(n)}(t)$ is the linear interpolation based on $X_{i,j}$, defined by

$$\begin{cases} X_{i,k(n)}(t) = X_{i,1}, \text{if } t \le t_1 \text{ and } t \in \mathcal{T}, \\ X_{i,k(n)}(t) = X_{i,k(n)}, \text{if } t \ge t_{k(n)} \text{ and } t \in \mathcal{T}, \\ X_{i,k(n)}(t) = ((t_{j+1} - t)X_{i,j} + (t - t_j)X_{i,j+1})/(t_{j+1} - t_j), \text{ if } t_j \le t \le t_{j+1}, \end{cases} \tag{2.4}$$

and $Y_{i,k(n)}(t)$ is the linear interpolation based on $Y_{i,j}$, which is defined similarly. Then, $X_{i,k(n)}(t)$ and $Y_{i,k(n)}(t)$ are the observed subject-specific curves of $X_i(t)$ and $Y_i(t)$, respectively. Although the curve observations can also be constructed using other smoothing methods, such as splines, the linear interpolation is computationally simple and does not depend on the choice of smoothing parameters.

2.2. Curve estimates

The mean curves $\mu(t)$ and $\eta(t)$ of $X(t)$ and $Y(t)$, respectively, can be naturally estimated using the sample means

$$\hat{\mu}_{n_1}(t) = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{i,k(n)}(t) \text{ and } \hat{\eta}_{n_2}(t) = \frac{1}{n_2} \sum_{i=n_x+1}^{n_x+n_2} Y_{i,k(n)}(t), \tag{2.5}$$

respectively. To evaluate the covariance curve of $X(t)$ and $Y(t)$, we need to first consider the moments and their estimators. Let $\mu_{(r_1,r_2)}(t) = E\left[X^{r_1}(t)Y^{r_2}(t)\right]$, and let $\mu_{(r_1,r_2,r_3,r_4)}(s,t) = E\left[X^{r_1}(s)X^{r_2}(t)Y^{r_3}(s)Y^{r_4}(t)\right]$ be the moments of $\left(X(t), Y(t)\right)^T$ with nonnegative integers r, k, and r_l, $l = 1, \ldots, 4$. The estimators of $\mu_{(k,r)}(t)$ and $\mu_{(r_1,r_2,r_3,r_4)}(s,t)$ constructed using $\left(X_{i,k(n)}(t), Y_{i,k(n)}(t)\right)^T$ are given by

$$\mu_{n,(r_1,r_2)}(t) = \frac{1}{n_{xy}} \sum_{i=n_x+1}^{n_x+n_{xy}} X_{i,k(n)}^{r_1}(t) Y_{i,k(n)}^{r_2}(t)$$

and

$$\mu_{n,(r_1,r_2,r_3,r_4)}(t) = \frac{1}{n_{xy}} \sum_{i=n_x+1}^{n_x+n_{xy}} X_{i,k(n)}^{r_1}(s) X_{i,k(n)}^{r_2}(t) Y_{i,k(n)}^{r_3}(s) Y_{i,k(n)}^{r_4}(t),$$

respectively. By (2.1), the estimator of $R(t)$ based on $\left(X_{i,k(n)}(t), Y_{i,k(n)}(t)\right)^T$ is

$$R_n(t) = \frac{\mu_{n,(1,1)}(t) - \mu_{n,(1,0)}(t)\mu_{n,(0,1)}(t)}{\sqrt{\left[\mu_{n,(2,0)}(t) - \mu_{n,(1,0)}^2(t)\right]\left[\mu_{n,(0,2)}(t) - \mu_{n,(0,1)}^2(t)\right]}}. \tag{2.6}$$

Let $\boldsymbol{u}(t) = \left(\mu_{(1,0)}(t), \mu_{(0,1)}(t), \mu_{(2,0)}(t), \mu_{(0,2)}(t), \mu_{(1,1)}(t)\right)^T$ be the vector of moments. Then, the estimator of $\boldsymbol{u}(t)$ is

$$\boldsymbol{u}_n(t) = \left(\mu_{n,(1,0)}(t), \mu_{n,(0,1)}(t), \mu_{n,(2,0)}(t), \mu_{n,(0,2)}(t), \mu_{n,(1,1)}(t)\right)^T, \tag{2.7}$$

and we can evaluate the asymptotic distribution of $R_n(t)$ using the asymptotic properties of $\boldsymbol{u}_n(t)$.

2.3. Hypotheses and test statistics

2.3.1. Testing the equivalence of mean curves

As illustrated in Duan, Keerthi and Poo (2003) and Fang et al. (2012), a primary scientific question is to test whether two stochastic processes $\{X(t) : t \in \mathcal{T}\}$ and $\{Y(t) : t \in \mathcal{T}\}$ have the same mean profiles; that is, $\mu(t) = \eta(t)$, for all $t \in \mathcal{T}$. A well-known approach for testing the equivalence of two mean curves is the Kolmogorov–Smirnov test statistic, $n \sup_{t\in\mathcal{T}} \left|\hat{\mu}_{n_1}(t) - \hat{\eta}_{n_2} n(t)\right|$. However, this is not robust because the supremum may be influenced by outliers.

We propose two test statistics, based on 'two-sided' and 'one-sided' alternative hypotheses. Our 'two-sided' null and alternative hypotheses are

$$H_0 : \mu(t) = \eta(t) \text{ for all } t \in \mathcal{T}, \text{ vs.} H_1 : \mu(t) \neq \eta(t) \text{ on some } \mathcal{A} \subset \mathcal{T}. \tag{2.8}$$

The 'one-sided' null and alternative hypotheses are

$$H_0 : \mu(t) = \eta(t) \text{ for all } t \in \mathcal{T}, \text{ vs.} H_1 : \mu(t) > \eta(t) \text{ for all } t \in \mathcal{T}, \tag{2.9}$$

where $\mathcal{A}$ has a positive Lebesgue measure. To test H_0 vs. H_1 in (2.8), we use the test statistic

$$L_n = \frac{1}{|\mathcal{T}|} \frac{n_1 n_2}{n} \int_{\mathcal{T}} \left[\hat{\mu}_{n_1}(t) - \hat{\eta}_{n_2}(t)\right]^2 dt, \tag{2.10}$$

where $|\mathcal{T}|$ is the length of $\mathcal{T}$. To test H_0 vs. H_1 in (2.9), we use the test statistic

$$D_n = \frac{1}{|\mathcal{T}|} \sqrt{\frac{n_1 n_2}{n}} \int_{\mathcal{T}} \left[\hat{\mu}_{n_1}(t) - \hat{\eta}_{n_2}(t)\right] dt. \tag{2.11}$$

Let $d_n(t_j) = \hat{\mu}_{n_1}(t_j) - \hat{\eta}_{n_2}(t_j)$. Because $\hat{\mu}_{n_1}(t) - \hat{\eta}_{n_2}(t)$ is obtained through a linear interpolation on $T_n = \{t_1, \ldots, t_{k(n)}\}$,

$$d_n(t) = \frac{(t - t_j)d_n(t_j) + (t_{j+1} - t)d_n(t_{j+1})}{t_{j+1} - t_j}, \quad t \in [t_j, t_{j+1}].$$

If we define $\hat{\mu}_{n_1}(t) = \hat{\eta}_{n_2}(t) \equiv 0$ for $t \le t_1$ or $t \ge t_{k(n)}$, then L_n and D_n simplify to

$$L_n = \frac{1}{3|\mathcal{T}|} \frac{n_1 n_2}{n} \sum_{j=1}^{k(n)-1} \Big(d_n^2(t_{j+1}) + d_n^2(t_j) + d_n(t_j)d_n(t_{j+1}) \Big)(t_{j+1} - t_j) \quad (2.12)$$

and

$$D_n = \frac{1}{|\mathcal{T}|} \sqrt{\frac{n_1 n_2}{n}} \sum_{j=1}^{k(n)-1} \frac{d_n(t_j) + d_n(t_{j+1})}{2}(t_{j+1} - t_j), \quad (2.13)$$

respectively. The approximate critical values for L_n and D_n are derived from their asymptotic distributions in Section 3.

2.3.2. Testing the correlation function

The second objective is to test whether $X(t)$ and $Y(t)$ are correlated for $t \in \mathcal{T}$, based on the correlation curve $R(t)$ defined in (2.1). A natural formulation is to test the null hypothesis that $X(t)$ and $Y(t)$ are uncorrelated for all time points in $\mathcal{T}$ versus the 'two-sided' alternative

$$H_0 : R(t) = 0 \text{ for all } t \in \mathcal{T}, \text{ vs. } H_1 : R(t) \neq 0 \text{ on some } \mathcal{A} \subset \mathcal{T}, \quad (2.14)$$

or that $X(t)$ and $Y(t)$ are uncorrelated versus the one-sided alternative

$$H_0 : R(t) = 0 \text{ for all } t \in \mathcal{T}, \text{ vs. } H_1 : R(t) > 0 \text{ (or } R(t) < 0) \text{ for all } t \in \mathcal{T}. \quad (2.15)$$

The 'one-sided' alternative in (2.15) suggests that the processes $X(t)$ and $Y(t)$ are always positively (or negatively) correlated, for all t in $\mathcal{T}$. In this case, we only consider $n = n_1 = n_2 = n_{xy}$, because a correlation inference requires paired data sets. Although other cases can also be handled using missing data methods, they are not our main goal here. Based on the empirical correlation curve $R_n(t)$ in (2.6), and in the definition of $\mu_{n,(k,r)}(t)$ and $\mu_{n,(r_1,r_2,r_3,r_4)}(t)$, replacing the summation sign $\sum_{i=n_x+1}^{n_x+n_{xy}}$ with $\sum_{i=1}^{n}$, we consider the test statistics

$$S_n = \frac{n}{|\mathcal{T}|} \int_{\mathcal{T}} R_n^2(t)dt \text{ and } W_n = \frac{\sqrt{n}}{|\mathcal{T}|} \int_{\mathcal{T}} R_n(t)dt \quad (2.16)$$

for the hypotheses in (2.14) and (2.15), respectively, where, as in (2.10), $|\mathcal{T}|$ is the length of $\mathcal{T}$. The approximate critical values for S_n and W_n are derived in Section 3.

3. Asymptotic Distributions of Test Statistics

Here, we derive the asymptotic distributions of the test statistics presented in Section 2. These asymptotic distributions lead to the critical values of the test statistics. The proofs of the theorems in this section are given in the online Supplementary Material.

3.1. Asymptotic distributions of mean test statistics

We first consider the hypotheses given in (2.8) and (2.9) and their testing statistics L_n and D_n, as defined in (2.12) and (2.13), respectively. Let

$$d(t) = \mu(t) - \eta(t) \text{ and } d_n(t) = \hat{\mu}_{n_1}(t) - \hat{\eta}_{n_2}(t)$$

be the real and estimated differences, respectively, of the mean curves. Let $R_{xx}(s,t) = Cov\big[X(s), X(t)\big]$, $R_{yy}(s,t) = Cov\big[Y(s), Y(t)\big]$, $R_{xy}(s,t) = Cov\big[X(s), Y(t)\big]$, $R_{yx}(s,t) = Cov\big[Y(s), X(t)\big]$, $R_{\epsilon}(s,t) = Cov\big[\epsilon(s), \epsilon(t)\big]$, and $R_{\xi}(s,t) = Cov\big[\xi(s), \xi(t)\big]$ be the covariance curves of the corresponding stochastic processes. Denote

$$\begin{aligned} R_{11}(s,t) &= R_{xx}(s,t) + R_{\epsilon}(s,t), \qquad & R_{12}(s,t) &= R_{xy}(s,t), \\ R_{21}(s,t) &= R_{yx}(s,t), & R_{22}(s,t) &= R_{yy}(s,t) + R_{\xi}(s,t). \end{aligned}$$

For the asymptotic properties described below, we denote by $\ell^{\infty}(\mathcal{T})$ the space of functions on $\mathcal{T}$ equipped with the supreme norm, and $\overset{D}{\Rightarrow}$ as the uniform weak convergence in $\ell^{\infty}(\mathcal{T})$. The following theorem shows that, under some mild conditions, $d_n(t) - d(t)$ can be uniformly weakly approximated by a Gaussian process, in the sense that the process $[(n_1 n_2)/n]^{1/2}[d_n(\cdot) - d(\cdot)]$ converges weakly to a Gaussian process. This result of weak convergence as a whole process on $\mathcal{T}$, instead of pointwise convergence on some $t \in \mathcal{T}$, allows us to study the weak limit of the test statistics constructed using the functional of $\widehat{\mu}_n(\cdot)$ and $\widehat{\eta}_n(\cdot)$, such as L_n and D_n. Let $\delta_{k(n)} = \max\{t_{j+1} - t_j : j = 0, 1, \ldots, k(n)\}$.

Before starting Theorem 1, we provide an outline of the proof. Note that

$$\begin{aligned} &\sqrt{\frac{n_1 n_2}{n}} \Big\{ \big[\widehat{\mu}_{n_1}(t) - \mu(t)\big] - \big[\widehat{\eta}_{n_2}(t) - \eta(t)\big] \Big\} \\ &= \sqrt{\frac{n_1 n_2}{n}} \Big\{ \big[\widehat{\mu}_{n_1}(t) - \mu_{k(n)}(t)\big] - \big[\widehat{\eta}_{n_2}(t) - \eta_{k(n)}(t)\big] \Big\} \\ &\quad + \sqrt{\frac{n_1 n_2}{n}} \Big\{ \big[\mu_{k(n)}(t) - \mu(t)\big] - \big[\eta_{k(n)}(t) - \eta(t)\big] \Big\}. \end{aligned}$$

We show that under the given conditions, the second term on the right-hand side is $o(2.1)$ uniformly in t. Then, we deal with the first term using empirical process theory, check the conditions for a uniform Donsker class, and identify the weak limit process.

Let $\mathcal{P}$ and $\mathcal{Q}$ be collections of all probability measures (P, Q) of the random processes $(X(\cdot)+\epsilon(\cdot), Y(\cdot)+\eta(\cdot))$. Denote $G = \sup_{t\in\mathcal{T}}\big([X(t)+\epsilon(t)]^2 + [Y(t)+\eta(t)]^2\big)^{1/2}$.

Theorem 1. *Assume that $\mathcal{T}$ is bounded, $\lim_{n\to\infty} n_j/n = \gamma_j$, for $j = 1, 2$, $\lim_{n\to\infty} n_{xy}/n = \gamma_{12}$, $\sqrt{n}\delta_{k(n)} \to 0$, $E_{(P,Q)}\big(G^2\big) < \infty$, and, for $\delta_n \to 0$, $\sup_{|t-s|\le\delta_n} E_{(P,Q)}\Big(\big[X(t)+\epsilon(t)-X(s)-\epsilon(s)\big]^2+\big[Y(t)+\eta(t)-Y(s)-\eta(s)\big]^2\Big) \to 0$. Then, as $n_1n_2/n \to \infty$,*

$$\sqrt{\frac{n_1n_2}{n}}\Big\{\big[\hat{\mu}_{n_1}(\cdot)-\mu(\cdot)\big]-\big[\hat{\eta}_{n_2}(\cdot)-\eta(\cdot)\big]\Big\} \overset{D}{\Rightarrow} W(\cdot), \tag{3.1}$$

where $W(\cdot)$ is the mean zero Gaussian process on $\mathcal{T}$, with covariance function

$$R(s,t) = E[W(s)W(t)] = \gamma_2 R_{11}(s,t) - \gamma_{12}[R_{12}(s,t)+R_{21}(s,t)] + \gamma_1 R_{22}(s,t).$$

The following remarks illustrate several special cases of Theorem 1.

Remark 1. In practice, the condition $\sup_{(P,Q)\in(\mathcal{P},\mathcal{Q})} E_{(P,Q)}(G^2) < \infty$ in Theorem 1 is not stringent, because $\mathcal{T}$ is a bounded set.

Remark 2. Because $\gamma_{12} \le \min\{\gamma_1,\gamma_2\}$, if $n_1 > 0$ and $\gamma_1 = 0$, then $\gamma_{12} = 0$, $\gamma_2 = 1$, and $R(s,t) = R_{11}(s,t)$. Similarly, if $n_2 > 0$ and $\gamma_2 = 0$, then $R(s,t) = R_{22}(s,t)$. For these two cases (i.e., $n_1 = o(n)$ or $n_2 = o(n)$), we cannot test the equivalence of the mean curves $\mu(t)$ and $\eta(t)$. If $n_{12} = 0$ or $\gamma_{12} = 0$, the samples for $X(t)$ and $Y(t)$ are independent or asymptotically independent; therefore, $R(s,t) = \gamma_2 R_{11}(s,t) + \gamma_1 R_{22}(s,t)$.

Remark 3. If $n_1 = n_2 = n_{xy}$, we observe only the paired sample $\big(X(t), Y(t)\big)^T$. In this case, $R(s,t) = \big[R_{11}(s,t) - R_{12}(s,t) - R_{21}(s,t) + R_{22}(s,t)\big]/2$, and we modify Theorem 1 by replacing $\sqrt{n_1n_2/n}$ with $\sqrt{n} = \sqrt{n_2} = \sqrt{n_1}$.

3.2. Rejection regions for mean test statistics

By definition, an $R(s,t)$ is symmetric and square integrable on $\mathcal{T}\times\mathcal{T}$, and there are eigenvalues λ and an associated eigenfunction $h(t)$, such that

$$\int_{s\in\mathcal{T}} R(s,t)h(s)ds = \lambda h(t),\ t \in \mathcal{T}. \tag{3.2}$$

Following the results of Theorem 1, the next theorem summarizes the asymptotic distributions of the test statistics L_n and D_n given in (2.10) and (2.11), respectively.

Theorem 2. *Assume that the conditions of Theorem* 1 *are satisfied, and that* $R(t,s)$ *defined in* (3.1) *is continuous. Then, under the null hypothesis* H_0 *of* (2.8) *and* (2.9),

(i) $L_n \to |\mathcal{T}|^{-1}\sum_{k=1}^{\infty}\lambda_k Z_k^2$ *in distribution as* $n \to \infty$, *and*

(ii) $D_n \to N\big(0,\sigma^2\big)$ *in distribution as* $n \to \infty$,

where $\sigma^2 = |\mathcal{T}|^{-2}\int_{\mathcal{T}\times\mathcal{T}} R(s,t)dsdt$, *the* λ_j*'s are the eigenvalues of* $R(s,t)$, *and* Z_j *are i.i.d.* $N(0,1)$ *random variables.*

From Theorem 2, the asymptotic variances of L_n and D_n depend on the unknown eigenvalues λ_k and $\sigma^2 = |\mathcal{T}|^{-2}\int_{\mathcal{T}\times\mathcal{T}} R(s,t)dsdt$. Therefore, a consistent estimator of $R(s,t)$ is needed to compute the approximate critical value for the test statistic. In practice, $R_{11}(s,t)$ can be estimated as

$$\begin{aligned}\widehat{R}_{11}(s,t) = {}& \frac{1}{n_1}\sum_{i=1}^{n_1}\big[X_{i,k(n)}(t_s^*)X_{i,k(n)}(t_t^*)\big] \\ & - \frac{1}{n_1^2}\left[\sum_{i=1}^{n_1}X_{i,k(n)}(t_s^*)\right]\left[\sum_{i=1}^{n_1}X_{i,k(n)}(t_t^*)\right],\end{aligned} \tag{3.3}$$

where, for $a = s$ and t, $t_a^* = \arg\min\big\{|t_j - a| : j = 1,\ldots,k(n)\big\}$. Similarly, $R_{12}(s,t)$, $R_{21}(s,t)$, and $R_{22}(s,t)$ can be estimated using their corresponding estimates $\widehat{R}_{12}(s,t)$, $\widehat{R}_{21}(s,t)$ and $\widehat{R}_{22}(s,t)$. Consequently, $R(s,t)$ can be estimated as

$$\widehat{R}(s,t) = \frac{n_1}{n}\widehat{R}_{11}(s,t) - \frac{n_{xy}}{n}[\widehat{R}_{12}(s,t) + \widehat{R}_{21}(s,t)] + \frac{n_2}{n}\widehat{R}_{22}(s,t).$$

Given that the function $R(t,s)$ does not have a closed-form expression, the eigenvalues λ_k are unknown and have to be estimated from the data. Let $\widehat{\mathbf{R}}_{k(n)}$ be the $k(n)\times k(n)$ matrix $\mathbf{R}_{k(n)} = \big(\widehat{R}(t_i,t_j)\big)$ of the estimators of $R(s,t)$ at the observed times $T_n = \{t_1,\ldots,t_{k(n)}\}$. As typical in practice, we compute the eigenvalues $\widehat{\lambda}_1,\ldots,\widehat{\lambda}_{k(n)}$ of $\widehat{\mathbf{R}}_{k(n)}$, and approximate the limit distribution of L_n by the distribution of

$$\widehat{L}_n = \frac{1}{|\mathcal{T}|}\sum_{j=1}^{k(n)}\widehat{\lambda}_j Z_j^2. \tag{3.4}$$

For a given significance level α, the rejection region for the null hypothesis in (2.8) is

$$L_n > Q_n(1-\alpha), \tag{3.5}$$

where the approximate critical value $Q_n(1-\alpha)$ is the $(1-\alpha)$th upper quantile of the distribution of $\widehat{L}_n$.

Remark 4. Results for eigenvalue estimations can be found elsewhere. For example, let $\lambda_1, \ldots, \lambda_p$ be the p largest eigenvalues, and let $\hat{\lambda}_j$ be the estimates. Then, by Theorem 2.7 in Horváth and Kokoszka (2012, p. 30), $E(\hat{\lambda}_j - \lambda_j)^2 = O(n^{-1})$, for all $1 \leq j \leq p$, for any fixed p. In practice, we only need the first p largest eigenvalues, for some fixed p, such as $p = 10$. Recall that in basis expansions, often only the first $k(\leq 10)$ bases are required for a good approximation; here, the situation is similar.

For the rejection region of the test statistic D_n, we replace $\int_{\mathcal{T}\times\mathcal{T}} R(s,t)dsdt$ with a summation of $\widehat{R}(s,t)$, and estimate σ^2 as

$$\widehat{\sigma}_n^2 = \frac{1}{|\mathcal{T}|^2} \sum_{j_1=1}^{k(n)} \sum_{j_2=1}^{k(n)} \Big[\widehat{R}(t_{j_1}, t_{j_2})(t_{j_1+1} - t_{j_1})(t_{j_2+1} - t_{j_2})\Big]. \tag{3.6}$$

Given that D_n is approximately normal when n is sufficiently large, the rejection region for the null hypothesis in (2.9) is approximated by

$$D_n > \widehat{\sigma}_n \Phi^{-1}(1-\alpha), \tag{3.7}$$

for a given significance level α, where $\Phi(\cdot)$ is the cumulative distribution function of the $N(0,1)$ distribution.

The power for D_n to detect the difference $\Delta = |\mathcal{T}|^{-1} \int_T \big[\mu(t) - \eta(t)\big] dt$ under H_1 of (2.9) is

$$\beta(\Delta) = P[D_n > \sigma\Phi^{-1}(1-\alpha)|H_1] \approx 1 - \Phi\left[\Phi^{-1}(1-\alpha) - \frac{\sqrt{n}\Delta}{\sigma_n}\right],$$

and the estimated power is

$$\widehat{P}[D_n > \sigma\Phi^{-1}(1-\alpha)|H_1] = 1 - \Phi\left[\Phi^{-1}(1-\alpha) - \frac{\sqrt{n}\Delta}{\widehat{\sigma}_n}\right].$$

3.3. Asymptotic distributions of moment curves

To obtain the asymptotic distributions of the correlation test statistics S_n and W_n of (2.16), we first develop the asymptotic properties of the moment estimates $\boldsymbol{u}_n(t)$ in (2.7). For notational simplicity, our derivation is for the case of $n_{xy} = n$ only. However, the results in this section also hold for the case $n_{xy} < n$ by replacing n with n_{xy}. The next theorem shows that the stochastic process $\sqrt{n}\big[\boldsymbol{u}_n(\cdot) - \boldsymbol{u}(\cdot)\big]$ converges weakly to a R^5-valued mean zero Gaussian process on $\mathcal{T}$. This result is used to derive the asymptotic distributions of the test statistics S_n and W_n in (2.16).

Theorem 3. *Assume the conditions in Theorem* 1 *hold. Then, as* $n \to \infty$,

$$\sqrt{n}\big[\boldsymbol{u}_n(\cdot) - \boldsymbol{u}(\cdot)\big] \overset{D}{\Rightarrow} \mathbf{W}(\cdot),$$

where $\mathbf{W}(\cdot)$ *is the* R^5*-valued mean zero Gaussian process on* $\mathcal{T}$ *with matrix covariance function* $\Omega(s,t) = Cov\big[\mathbf{Z}(s), \mathbf{Z}(t)\big]$ *and* $\mathbf{Z}(t) = \big(X(t) + \epsilon(t), Y(t) + \xi(t), [X(t) + \epsilon(t)]^2, [Y(t) + \xi(t)]^2, [X(t) + \epsilon(t)][Y(t) + \xi(t)]\big)^T$.

Because the test statistics S_n and W_n are functions of the moment process $\boldsymbol{u}(t)$, Theorem 3 suggests that the asymptotic distributions of S_n and W_n can be derived from the distribution of the Gaussian process $\mathbf{W}(t)$ and the covariance structure $\Omega(s,t)$ of $\mathbf{Z}(t)$.

3.4. Rejection regions for correlation test statistics

Based on the results of Theorem 3, we first derive the asymptotic distributions of the test statistics S_n and W_n, and then present their approximate rejection regions for the tests given in (2.14) and (2.15).

For any R^5-valued vector $\mathbf{z} = (z_1, \ldots, z_5)$, we define

$$H(\mathbf{z}) = \frac{z_5 - z_1 z_2}{\sqrt{(z_3 - z_1^2)(z_4 - z_2^2)}}, \tag{3.8}$$

and its derivative $\dot{H}(\mathbf{z}) = \big(\partial H(\mathbf{z})/\partial z_1, \ldots, \partial H(\mathbf{z})/\partial z_5\big)$, where

$$\frac{\partial H(\mathbf{z})}{\partial z_1} = \frac{z_1 z_5 - z_2 z_3}{z_3 - z_1^2}\frac{\partial H(\mathbf{z})}{\partial z_5}, \quad \frac{\partial H(\mathbf{z})}{\partial z_2} = \frac{z_2 z_5 - z_1 z_4}{z_4 - z_2^2}\frac{\partial H(\mathbf{z})}{\partial z_5},$$
$$\frac{\partial H(\mathbf{z})}{\partial z_3} = \frac{z_1 z_2 - z_5}{2(z_3 - z_1^2)}\frac{\partial H(\mathbf{z})}{\partial z_5}, \quad \frac{\partial H(\mathbf{z})}{\partial z_4} = \frac{z_1 z_2 - z_5}{2(z_4 - z_2^2)}\frac{\partial H(\mathbf{z})}{\partial z_5},$$
$$\frac{\partial H(\mathbf{z})}{\partial z_5} = \frac{1}{\sqrt{(z_4 - z_2^2)(z_3 - z_1^2)}}.$$

Based on the results of Theorem 3, the next theorem describes shows the asymptotic distributions of the statistics S_n and W_n.

Theorem 4. *Assume that the conditions of Theorem* 3 *are satisfied. Under the null hypothesis* H_0 *of* (2.14) *and* (2.15)*, we have that, as* $n \to \infty$,

$$S_n \to |\mathcal{T}|^{-1} \sum_{j=1}^{\infty} \lambda_j Z_j^2 \ \textit{in distribution} \tag{3.9}$$

and

$$W_n \to N\big(0, \sigma_2^2\big) \ \textit{in distribution,} \tag{3.10}$$

where $\sigma_2^2 = |\mathcal{T}|^{-2} \int_{\mathcal{T}\times\mathcal{T}} Q(s,t) ds dt$, $Q(s,t) = \dot{H}\big[\boldsymbol{u}(s)\big]\Omega(s,t)\dot{H}^T\big[\boldsymbol{u}(t)\big]$, *with*

$\Omega(s,t)$ defined in Theorem 3, the λ_j's are the eigenvalues of $Q(s,t)$, and Z_j are i.i.d. $N(0,1)$ random variables.

From Theorem 4, for a given nominal level α, we obtain the asymptotic rejection regions for H_0 of (2.14) and (2.15), based on S_n and W_n, respectively, as

$$S_n > G^{-1}(1-\alpha), \quad \text{and} \quad W_n > \widehat{\sigma}_{2,n}\Phi^{-1}(1-\alpha),$$

where $G^{-1}(1-\alpha)$ is the $(1-\alpha)$th upper quantile of the distribution on the right-hand side of (3.9), and the empirical estimator $\widehat{\sigma}_{2,n}$ is a consistent estimate of σ_2.

The power for S_n to detect the difference $\Delta = |\mathcal{T}|^{-1}\int_T R^2(t)dt$ under H_1 of (2.14) is

$$\beta(\Delta) = P\big[S_n > \chi^2(1-\alpha) - n\Delta\big|H_1\big] \approx 1 - G\big[\chi^2(1-\alpha) - n\Delta\big],$$

where $G(\cdot)$ is the distribution function of the right-hand side of (3.9).

The power for W_n to detect the difference $\Delta = |\mathcal{T}|^{-1}\int_T R(t)dt$ under H_1 of (2.15) is

$$\beta(\Delta) = P\big[W_n > \sigma_2\Phi^{-1}(1-\alpha)|H_1\big] \approx 1 - \Phi\left[\Phi^{-1}(1-\alpha) - \frac{\sqrt{n}\Delta}{\widehat{\sigma}_{2,n}}\right].$$

4. Application to TGE Data

Here, we apply the proposed testing procedures of Sections 2 and 3 to the TGE data analyzed by Fang et al. (2012). The data set contains repeated observations from high-throughput gene expressions of 18 genes in *P. aeruginosa*, expressed in 24 different biological conditions, with various antibiotics (e.g., *AMp, kam, Cm, Tc, etc.*) at different concentrations. Here, the 24 biological conditions can be viewed as 24 independent samples. The gene expression outcome was measured as log-scaled counts per second (CPS) every 30 minutes for 24 hours, yielding 48 equally spaced observation time points. For further details on the design and biological objectives of the experiment, refer to Duan, Keerthi and Poo (2003) and Fang et al. (2012). Two of the main statistical objectives of the experiment are to test whether two selected genes of interest (a) have the same mean expression curves, and (b) are correlated over time.

For the purpose of illustration, we apply our testing methods of Sections 2 and 3 to the gene expression data on three genes, PA3897 (narL), PA2997 (nqrC), and PA0649 (trpG); see Table 1 of Fang et al. (2012). PA2997 and PA0649 are important genes related to energy metabolism, and PA3897 is related to the two-

Table 1. Hypotheses, test statistics, and the corresponding p-values for two selected pairs of genes from the Temporal Gene Expression Study.

Gene pair	Hypotheses	Proposed			Loess		Spline	
		Statistic	Value	P-value	Value	P-value	Value	P-value
PA2997 vs PA3897	$\mu = \eta$ vs $\mu \geq \eta$	D_n	4.0621	< 0.0001	4.0684	< 0.0001	4.0618	< 0.0001
	$\mu = \eta$ vs $\mu \neq \eta$	L_n	19.6430	< 0.0001	19.6964	< 0.0001	19.6082	< 0.0001
	$R = 0$ vs $R > 0$	W_n	0.2852	0.3016	0.3324	0.2721	0.1224	0.4120
	$R = 0$ vs $R \neq 0$	S_n	0.9652	0.1907	1.0568	0.1559	1.0011	0.1686
PA0649 vs PA2997	$\mu = \eta$ vs $\mu \geq \eta$	D_n	-0.2261	0.7209	-0.2246	0.7196	-0.2261	0.7209
	$\mu = \eta$ vs $\mu \neq \eta$	L_n	0.5424	0.0852	0.5423	0.0896	0.5417	0.0796
	$R = 0$ vs $R > 0$	W_n	2.3851	< 0.0001	2.3978	< 0.0001	1.0017	0.0040
	$R = 0$ vs $R \neq 0$	S_n	6.5203	0.0011	6.5846	0.0007	6.5297	0.0008

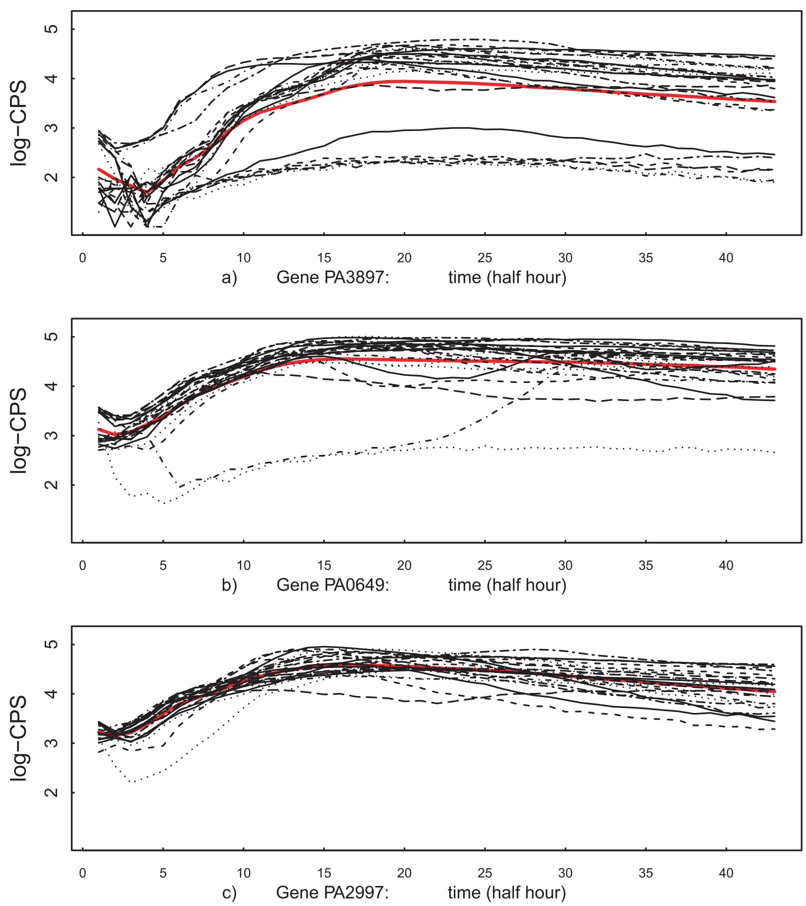

Figure 1. The observed expressions of three genes, PA3897 (narL), PA2997 (nqrC), and PA0649 (trpG), in 24 "biological conditions" over 21 hours. The red lines are the corresponding estimated mean functions of the expression profiles.

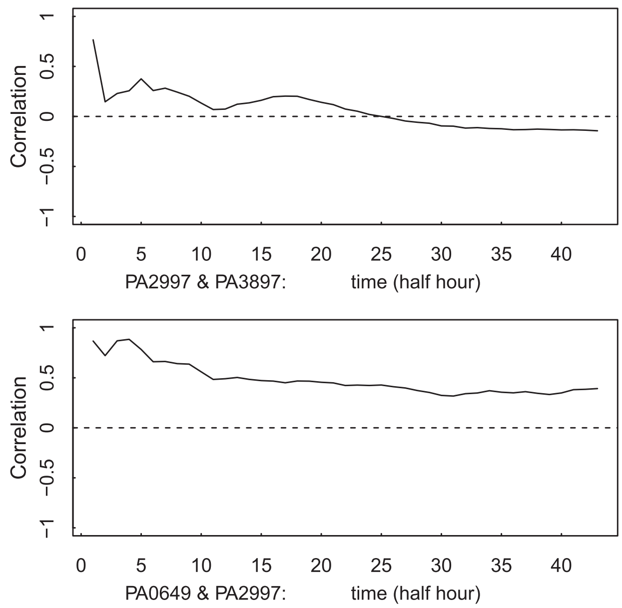

Figure 2. The estimated correlation coefficient curves of two pairs of genes over 21 hours. Top panel: PA2997 vs. PA3897. Bottom panel: PA0649 vs. PA2997.

component response regulator NarL. Figure 1 shows the observed gene expression trajectories of these three genes among the 24 "biological conditions."

We computed the test statistics L_n and D_n for testing the equality of two log-scaled CPS mean curves in (2.8) and (2.9), and the test statistics S_n and W_n for testing the log-scaled CPS correlation curve of two genes in (2.14) and (2.15). Table 1 shows the values of the test statistics for the genes "PA2997 vs. PA3897" and "PA0649 vs. PA2997," the corresponding hypotheses, and the approximated p-values of the test statistics, computed using the asymptotic distributions of Sections 3.2 and 3.4. The results show that the mean expression functions of gene PA2997 and gene PA3897 are significantly different, with $p < 0.001$ for the test statistic L_n, and that there is no significant difference between the mean expression functions of gene PA2997 and gene PA0694, with $p = 0.284$ and $p = 0.319$ for the test statistics L_n and D_n, respectively. The mean expressions of gene PA2997 are higher across the 24-hour period than that of gene PA3897, with $p < 0.001$ for the test statistic D_n.

Figure 2 shows the estimated correlation coefficient functions of the gene pairs (PA2997, PA3897) and (PA0649, PA2997) across the 24-hour period. The results of Table 1 for hypotheses (2.14) and (2.15) show that genes PA3897 and

Table 2. Test statistic L_n, empirical Type-I error, and powers under different values of C and effective differences Δ_1 for the two-sided test of the mean curves. All results are based on $N = 5{,}000$ replicates.

Time	Constant	Effective	Proposed		Loess		Spline	
point k	C	difference Δ_1	L_n	Power	Statistic	Power	Statistic	Power
10	0	0	0.0928	0.0394	0.0910	0.0398	0.0870	0.0382
	0.24	0.0076	0.2348	0.1410	0.2332	0.1298	0.1903	0.0878
	0.28	0.0103	0.2880	0.2186	0.2847	0.2118	0.2327	0.1304
	0.32	0.0135	0.3479	0.3658	0.3441	0.3536	0.2838	0.2076
	0.36	0.0170	0.4148	0.5738	0.4115	0.5580	0.3440	0.3514
	0.40	0.0210	0.4899	0.7832	0.4868	0.7772	0.4138	0.5636
30	0	0	0.0964	0.0590	0.0860	0.0472	0.0809	0.0460
	0.24	0.0076	0.3181	0.3784	0.3061	0.3404	0.2286	0.1616
	0.28	0.0103	0.4013	0.6740	0.3856	0.6254	0.2910	0.2956
	0.32	0.0135	0.4938	0.9086	0.4774	0.8802	0.3662	0.5622
	0.36	0.0170	0.6026	0.9892	0.5814	0.9836	0.4544	0.8342
	0.40	0.0210	0.7171	0.9994	0.6976	0.9996	0.5554	0.9722
50	0	0	0.0948	0.0530	0.0812	0.0482	0.0794	0.0470
	0.24	0.0076	0.3272	0.4192	0.3042	0.3408	0.2436	0.1920
	0.28	0.0103	0.4111	0.7324	0.3847	0.6380	0.3118	0.3658
	0.32	0.0135	0.5079	0.9432	0.4776	0.9026	0.3932	0.6668
	0.36	0.0170	0.6176	0.9942	0.5829	0.9858	0.4876	0.9152
	0.40	0.0210	0.7403	1.0000	0.7006	0.9998	0.5951	0.9896
100	0	0	0.0952	0.0594	0.0781	0.0508	0.0768	0.0508
	0.24	0.0076	0.3313	0.4522	0.3055	0.3688	0.2586	0.2354
	0.28	0.0103	0.4159	0.7824	0.3878	0.6862	0.3326	0.4604
	0.32	0.0135	0.5142	0.9668	0.4827	0.9304	0.4198	0.7920
	0.36	0.0170	0.6256	0.9994	0.5903	0.9964	0.5202	0.9688
	0.40	0.0210	0.7502	1.0000	0.7106	1.0000	0.6337	0.9992

PA2997 are not significantly correlated across the 24-hour period with $p = 0.280$ and 0.5048 for the test statistics W_n and S_n, respectively. Furthermore, PA2997 and PA0694, which are related to energy metabolism, have a significant and positive correlation across the 24-hour period with $p < 0.001$ for the one-sided test based on W_n, and $p = 0.045$ for the two-sided test based on S_n.

5. Simulation

To investigate the finite-sample properties of the proposed methods, and to compare them with those of the commonly used local linear smoothing (Loess) and spline methods, we present here several simulation studies, which are designed to mimic practical situations with moderate sample sizes. We consider

Table 3. Test statistic D_n, empirical Type-I errors, and powers under different values of C^* and effective differences Δ_2 for the one-sided test of the mean curves. All results are based on $N = 5{,}000$ replicates.

Time	Constant	Effective	Proposed		Loess		Spline	
point k	C	difference Δ_2	D_n	Power	Statistic	Power	Statistic	Power
10	0	0	0.0002	0.0970	0.0003	0.1046	0.0002	0.0964
	0.10	0.0315	0.1059	0.3917	0.1183	0.4423	0.1067	0.3959
	0.15	0.0472	0.1587	0.6075	0.1772	0.6649	0.1598	0.6107
	0.20	0.0629	0.2116	0.7972	0.2362	0.8472	0.2129	0.8028
	0.25	0.0786	0.2644	0.9252	0.2951	0.9512	0.2659	0.9262
30	0	0	0.0010	0.0764	0.0009	0.0862	0.0009	0.0766
	0.10	0.0315	0.0994	0.3399	0.1050	0.3637	0.1006	0.3435
	0.15	0.0472	0.1486	0.5534	0.1571	0.5840	0.1503	0.5606
	0.20	0.0629	0.1978	0.7611	0.2092	0.7845	0.2000	0.7669
	0.25	0.0786	0.2470	0.9083	0.2612	0.9193	0.2496	0.9125
50	0	0	−0.0038	0.0624	−0.0039	0.0708	−0.0038	0.0632
	0.10	0.0315	0.0949	0.3206	0.0992	0.3396	0.0957	0.3232
	0.15	0.0472	0.1442	0.5277	0.1508	0.5541	0.1454	0.5311
	0.20	0.0629	0.1935	0.7397	0.2023	0.7582	0.1950	0.7435
	0.25	0.0786	0.2429	0.9001	0.2539	0.9083	0.2445	0.9035
100	0	0	−0.0001	0.0644	−0.0002	0.0730	−0.0001	0.0648
	0.10	0.0315	0.0983	0.3156	0.1006	0.3292	0.0987	0.3170
	0.15	0.0472	0.1476	0.5274	0.1510	0.5414	0.1481	0.5302
	0.20	0.0629	0.1968	0.7565	0.2014	0.7623	0.1975	0.7567
	0.25	0.0786	0.2460	0.9099	0.2518	0.9087	0.2468	0.9105

separately the tests for the equality of two mean curves and the tests for the correlation function between two stochastic processes. For each case, the simulation is based on 5,000 replications, and the mean values of the estimates over the replications are reported. A detailed description of the simulation is given in the online Supplementary Material; here, we only describe the results.

5.1. Testing the equality of mean curves

5.1.1. Simulation for test with two-sided alternatives

For a series of alternatives determined by the C-values, Table 2 shows the corresponding differences $\Delta_1 = |\mathcal{T}|^{-1} \int_{\mathcal{T}} \left[\mu(t) - \eta(t)\right]^2 dt$, values of the test statistic L_n, empirical Type-I errors and empirical powers under different values of Δ_1. These results suggest that the Type-I error of the test is close to the nominal level of 0.05, and that the power of the test is greater than 80% if $C \geq 0.50$ or $\Delta_1 \geq 0.035$.

Table 4. Test statistics, empirical Type-I errors, and empirical powers under different values of ρ and effective differences Δ_3 for testing a zero correlation function with two-sided alternatives. All results are based on $N = 5{,}000$ replicates.

Time	Constant	Effective	Proposed		Loess		Spline	
point k	ρ	difference Δ_3	S_n	Power	Statistic	Power	Statistic	Power
10	0	0	1.1160	0.0724	1.1160	0.0724	1.1156	0.0734
	0.12	0.0069	1.4678	0.2138	1.4678	0.2138	1.2849	0.1378
	0.15	0.0108	1.6694	0.3166	1.6694	0.3166	1.3832	0.1780
	0.18	0.0156	1.9166	0.4574	1.9166	0.4574	1.5040	0.2314
	0.20	0.0192	2.1069	0.5566	2.1069	0.5566	1.5979	0.2766
30	0	0	1.0277	0.0460	1.0360	0.0540	1.0340	0.0432
	0.12	0.0069	1.3653	0.3178	1.3640	0.2814	1.1613	0.1000
	0.15	0.0108	1.5553	0.5280	1.5491	0.4484	1.2349	0.1436
	0.18	0.0156	1.7877	0.7426	1.7755	0.6442	1.3242	0.2022
	0.20	0.0192	1.9662	0.8630	1.9494	0.7530	1.3922	0.2520
50	0	0	1.0207	0.0456	1.0208	0.0570	1.0211	0.0394
	0.12	0.0069	1.3529	0.4124	1.3436	0.2720	1.1558	0.1010
	0.15	0.0108	1.5401	0.6792	1.5254	0.4148	1.2307	0.1476
	0.18	0.0156	1.7690	0.8952	1.7478	0.5886	1.3229	0.2064
	0.20	0.0192	1.9449	0.9620	1.9187	0.7018	1.3945	0.2566
100	0	0	1.0096	0.0442	1.0122	0.0592	1.0106	0.0630
	0.12	0.0069	1.3390	0.6514	1.3334	0.2690	1.2165	0.1570
	0.15	0.0108	1.5243	0.9162	1.5136	0.4088	1.3331	0.2158
	0.18	0.0156	1.7509	0.9922	1.7338	0.5862	1.4747	0.2942
	0.20	0.0192	1.9249	0.9994	1.9029	0.7034	1.5841	0.3592

5.1.2. Simulation for test with one-sided alternatives

Table 3 shows the values of the test statistic D_n, empirical Type-I errors and empirical powers under a series of C^*-values and their corresponding values of $\Delta_2 = |\mathcal{T}|^{-1} \int_{\mathcal{T}} \big(\mu(t) - \eta^*(t)\big)dt$. These results show that the Type-I error for D_n is close to the nominal level of 0.05, and that the power is more than 80% if $C^* \geq 0.30$ or $\Delta_2 \geq 0.096$.

5.2. Testing correlation functions

5.2.1. Simulation for test with two-sided alternatives

Table 4 shows the values of the test statistic S_n, empirical Type-I errors and empirical powers under a series of ρ-values and their corresponding values of $\Delta_3 = |\mathcal{T}|^{-1} \int_{\mathcal{T}} R^2(t)dt$. These results show that the Type-I error for S_n is close to the nominal level of 0.05, and that power is above 80% if $\rho \geq 0.115$ or $\Delta_3 \geq 0.0066$.

Table 5. Test statistics, empirical Type-I errors, and empirical powers under different values of ρ and effective differences Δ_4 for testing a zero correlation function with one-sided alternatives. All results are based on $N = 5{,}000$ replicates.

Time point	Constant	Effective difference	Proposed		Loess		Spline	
k	ρ	Δ_4	W_n	Power	Statistic	Power	Statistic	Power
10	0	0	0.0096	0.0500	0.0096	0.0500	0.0119	0.0580
	0.05	0.0306	0.2225	0.1688	0.2225	0.1688	0.1710	0.1124
	0.07	0.0428	0.3077	0.2364	0.3077	0.2364	0.2352	0.1436
	0.08	0.0489	0.3503	0.2780	0.3503	0.2780	0.2671	0.1618
	0.10	0.0611	0.4355	0.3748	0.4355	0.3748	0.3317	0.2062
30	0	0	0.0020	0.0556	−0.0004	0.0536	−0.0045	0.0552
	0.05	0.0306	0.2162	0.3402	0.2136	0.2408	0.1320	0.1326
	0.07	0.0428	0.3018	0.5130	0.2991	0.3574	0.1865	0.1736
	0.08	0.0489	0.3446	0.6060	0.3419	0.4264	0.2141	0.2006
	0.10	0.0611	0.4264	0.9234	0.4275	0.5660	0.2684	0.2576
50	0	0	−0.0014	0.0568	0.0059	0.0636	0.0047	0.0590
	0.05	0.0306	0.2125	0.4736	0.2192	0.2208	0.1449	0.1392
	0.07	0.0428	0.2980	0.7000	0.3046	0.3266	0.2012	0.1842
	0.08	0.0489	0.3408	0.7910	0.3473	0.3848	0.2293	0.2122
	0.10	0.0611	0.4264	0.9234	0.4326	0.5140	0.2854	0.2736
100	0	0	−0.0001	0.0530	0.0052	0.0648	0.0039	0.0626
	0.05	0.0306	0.2138	0.7188	0.2184	0.2258	0.1790	0.1500
	0.07	0.0428	0.2994	0.9230	0.3037	0.3338	0.2491	0.1960
	0.08	0.0489	0.3421	0.9678	0.3464	0.3944	0.2842	0.2210
	0.10	0.0611	0.4277	0.9962	0.4317	0.5282	0.3545	0.2754

5.2.2. Simulation for test with one-sided alternatives

To test the hypotheses in (2.15) using the test statistic W_n, our samples are generated in the same way as in Section 5.2.1, except that $\rho(t)$ in (29) in the Supplementary material is replaced with $\rho^*(t) = \rho\big|\sin(8 - t/10)\big|$, so that $\rho^*(t)$ does not change signs. Here, ρ determines the difference of the correlation curve $R(t)$ from zero.

Table 5 shows the values of the test statistic W_n, empirical Type-I errors, and empirical powers under a series of ρ-values and their corresponding values of $\Delta_4 = |\mathcal{T}|^{-1} \int_{\mathcal{T}} R(t)dt$. These results confirm that, the Type-I error for W_n is close to the nominal level of 0.05, and that the power is above 80% if $\rho \geq 0.08$ or $\Delta_4 \geq 0.0575$.

6. Discussion

We have developed a class of simple nonparametric procedures based on linear interpolations of observed functional data for testing multiple mean and correlation curves. As a direct response to a practical need in biological studies, such as the TGE study of Section 4, our testing procedures serve as an alternative to existing methods, such as basis expansions, splines, smoothing, and so on, and is simple to use. Our testing procedures also have the advantage of being computationally simple and having straightforward biological interpretations. The asymptotic properties of the test statistics suggest that our testing procedures can be justified theoretically under minimal assumptions, which is crucial in many biological studies. However, actual use still requires approximate eigenvalues, similarly to other methods such as the FPCA. In addition, inference inaccuracy is expected owing to the sparcity of the data. Our simulation results demonstrate that the approximate rejection regions obtained from the asymptotic distributions of the test statistics exhibit satisfactory performance, in general, under practical settings. Further extensions of our method may include testing procedures for structured nonparametric models with time-varying covariates, or for functional linear models.

Supplementary Material

Ming T. Tan is the corresponding author. The online Supplementary Material provides a detailed description of the simulations, including the test for the equality of mean curves and the test for correlation functions, and the proofs of Theorems 1–4.

Acknowledgements

The research of Yuan, Fang, and Tan was partially supported by the National Cancer Institute (NCI) grant R01CA164717 and P30CA 051008.

References

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

Cai, T. and Yuan, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: phase transition. *Annals of Statistics*,**39**(5), 2330–2355.

Cao, Y., Han, Y., Meng, D., Li, D., Jin, Q., Lin, Y. and Cai, Y. (2016). Structural, evolutionary, and functional analysis of the class III peroxidase gene family in Chinese pear (Pyrus bretschneideri). *Frontier Plant Science* **9**, `https://doi.org/10.3389/fpls.2016.01874`.

Chiang, C. T., Rice, J. A. and Wu, C. O. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables, *Journal of the American Statistical Association* **96**, 605–619.

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

Degras, D. (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statistica Sinica* **21**, 1735–1765.

Degras, D. (2017). Simultaneous confidence bands for the mean of functional data. *WIREs Computational Statistics*, e1397. doi:10.1002/wics.1397

Duan, K., Keerthi, S. S. and Poo, A. N. (2003). Evaluation of simple performance measures for tuning SVM hyperparameters, *Neurocomputing* **51**, 41–59.

Estévez-Pérez, G. and Vilar, J. A. (2008). Functional anova starting from discrete data: an application to air quality data. *Environmental and Ecological Statistics* **20**, 495–515.

Fan, J. and Zhang, J. T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society, B (Statistical Methodology)* **62**, 303–322.

Fang, H-B, Deng, D., Tian, G-L, Shen, L., Duan, K. and Song, J. (2012). Analysis for temporal gene expressions under multiple biological conditions, *Statistics in Biosciences* **4**, 282–299.

Fremdt, S., Steinebach, J. G., Horváth, L. and Kokoszka, P. (2012). Testing the equality of covariance operators in functional samples. *Scandinavian Journal of Statistics* **40**, 38–152.

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

Hall, P. and van Keilegom, I. (2007). Two-sample tests in functional data analysis starting from discrete data, *Statistica Sinica* **17**, 1511–1531.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *Elements of Statistical Learning*, Springer.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*, Springer, New York.

Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L. P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data, *Biometrika* **85**, 809–822.

Huang, J. Z., Wu, C. O. and Zhou, L. (2002). Varying coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89**, 111–128.

Kim, S. and Zhao, Z. (2013). Unified inference for sparse and dense longitudinal models. *Biometrika* **100**, 203–212.

Kraus, D. and Panaretos, V. M. (2012). Disperson operators and resistant second-order functional data analysis. *Biometrika* **99**, 813–832.

Li, Y., Wang, N. and Carroll, R. J. (2010). Generalized functional linear models with semiparametric single-index interactions. *Journal of the American Statistical Association* **105**, 621–633.

Müller, H-G. and Yao, F. (2008). Functional additive models. *Journal of the American Statistical Association* **103**, 1534–1544.

Pomann, G. M., Staicu, A. M. and Gosh, M. (2016). Two-sample distribution-free test for functional data with application to a diffusion tensor imaging study of multiple sclerosis. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* **65**, 395–414.

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

Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves, *Biometrics* **57**, 253–259.

Shi, M., Weiss, R. E. and Taylor, J. M. C. (1996). An analysis of paediatric AIDS CD4 counts using flexible random curves, *Applied Statistics* **45**, 151–163.

Wang, H. L., Zhong, P.-S., Cui, Y. H. and Li, Y. (2017). Unified empirical likelihood ratio tests for functional concurrent linear models and the phase transition from sparse to dense functional data. *Journal of the Royal Statistical Society, B (Statistical Methodology)*, to appear.

Wu, C. O. and Chiang, C. T. (2000). Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Statist Sinica* **10**, 433–456.

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

Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University Medical Center, Washington DC 20057 USA.

E-mail: ay312@georgetown.edu

Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University Medical Center, Washington DC 20057 USA.

E-mail: hf183@georgetown.edu

Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University Medical Center, Washington DC 20057 USA.

E-mail: hl662@georgetown.edu

Office of Biostatistics Research, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892 USA.

E-mail: wuc@nhlbi.nih.gov

Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University Medical Center, Washington DC 20057 USA.

E-mail: mtt34@georgetown.edu

(Received June 2017; accepted July 2018)