



# Evaluation of reproducibility for paired functional data

Runze Li and Mosuk Chow\*

*Department of Statistics, Pennsylvania State University, University Park, PA 16802-2111, USA*

Received 24 June 2002

---

## Abstract

Evaluation of reproducibility is important in assessing whether a new method or instrument can reproduce the results from a traditional gold standard approach. In this paper, we propose a measure to assess measurement agreement for functional data which are frequently encountered in medical research and many other research fields. Formulae to compute the standard error of the proposed estimator and confidence intervals for the proposed measure are derived. The estimators and the coverage probabilities of the confidence intervals are empirically tested for small-to-moderate sample sizes via Monte Carlo simulations. A real data example in physiology study is used to illustrate the proposed statistical inference procedures. © 2004 Elsevier Inc. All rights reserved.

*AMS 2000 subject classifications:* Primary 62F12; Secondary 62P10

*Keywords:* Concordance correlation coefficient; Functional data; Image data; Kappa coefficient

---

## 1. Introduction

Evaluation of reproducibility is needed for many scientific research problems. For example, when a new instrument is developed, it is of interest to assess whether the new instrument can reproduce the results obtained by using a traditional gold standard criterion. Indeed, the need to quantify agreement arises in many research fields when two approaches or two raters simultaneously evaluate a response. There are some traditional criteria for measuring agreement between two rating approaches, such as Pearson's correlation coefficient and paired *t*-test when the

---

\*Correspondence author. Fax: 814-863-7114.

*E-mail address:* [mchow@stat.psu.edu](mailto:mchow@stat.psu.edu) (M. Chow).

responses are continuous. Even though these criteria had been used in practice, they fail to detect poor agreement in some situations (see, for example, [12]). Thus, the topic of assessing agreement for measurements by two approaches has become an interesting research topic. Lin et al. [13] gives a review and comparison of various measures of recent developments in this area. Some measures to evaluate reproducibility include intraclass correlation [6,14] and within-subject coefficient of variation [10]. Lin [12] introduced the concordance correlation coefficient which assesses the linear relationship between two variables under the constraint that the intercept is zero and the slope is one. This measure is more appropriate for assessing reproducibility of continuous outcomes. King and Chinchilli [9] proposed a generalized concordance correlation coefficient for categorical data and continuous data. They also show that the generalized concordance correlation coefficient is actually equivalent to the weighted kappa coefficient [4] for ordinal data and equivalent to the kappa coefficient [3] for binary data.

The concordance correlation coefficient has been widely applied to various research fields since its introduction. Several extensions have been proposed to address different problems recently. Extending Lin's ideas, Chinchilli et al. [2] suggested a weighted concordance correlation coefficient for repeated measures design. Vonesh et al. [18] used the concordance correlation coefficient to assess goodness-of-fit for generalized nonlinear mixed-effects models. King and Chinchilli [8] developed a robust version of the concordance correlation coefficient. To accommodate covariate adjustment, Barnhart and Williamson [1] proposed a generalized estimating equations approach to model the concordance correlation coefficient via three sets of estimating equations.

This paper deals with the problem when paired data of repeated measures are curves, and proposes a measure to evaluate the reproducibility of repeated-paired curve data. Such data are also called functional data. Image data and growth curve data are special cases thereof. Analysis of functional data is becoming an important topic in the statistical literature. Many interesting applications can be found in the excellent book by Ramsey and Silverman [15] and references therein. The method in this paper is developed to analyze an actual data set collected in the Noll Physiological Research Center at The Pennsylvania State University (courtesy of Dr. W. L. Kenney). This study was designed to assess agreement of measurement by two different approaches to measure body core temperature.

Body core temperature refers to the temperature of tissues located at sufficient depth within the body so that it is minimally affected by the environmental temperature. The core temperature may be monitored by measuring the temperature at different locations in the body. Both the esophageal temperature ( $T_{es}$ ) and rectal temperature ( $T_{re}$ ) have been used for measuring indices of core temperature for research purposes. The  $T_{es}$  probe is a thermometer sealed inside a plastic feeding tube, which is inserted into the esophagus via the nostril to the level of the right atrium. The  $T_{re}$  probe is a thermometer sealed in soft vinyl tubing and is inserted into the rectum 8–10 cm beyond the anal sphincter. It is known that the  $T_{es}$  probe includes the temperature of the saliva swallowed by the subject and changes with depth of the probe in the esophagus and  $T_{re}$  has a slower response time than  $T_{es}$  in

the case of rapid storage of heat during intense thermal stress of a short duration. During this study, temperatures for each subject were observed by the two approaches every minute over 90 min of an experiment period. The researchers would like to have an overall index of agreement over the whole time period. They also want to have an index which can summarize the degree of agreement during different conditions. The experiment consists of using the two methods to measure core temperature when the subjects will enter a chamber set at 36°C and 50 humidity. The subjects will sit quietly for 10 min and then exercise on a motor-driven treadmill for 20 min. The experiment will end when the subject has undergone 3 cycles of 10-min rest/20-min exercise. To address the researcher's need, we propose a measure of agreement for repeated-paired curve data. We name it concordance correlation coefficient since it is motivated by the concordance correlation coefficient proposed by Lin [12] which is a measure of agreement at a single time point. When the observations are taken over different time points, the concordance correlation coefficient proposed by Lin [12] cannot be used for the data set. The weighted concordance correlation coefficient for repeated measures may not be appropriate for this situation because an observation for each subject is very dense, more like a pair of curves over time rather than a multi-dimensional vector or a set of repeated measurements.

Since functional data arise frequently, there is strong need for a generalization of concordance correlation coefficient for such data. In this paper, we deal with this problem and propose an estimate for it. (Since Pearson correlation coefficient is such a frequently encountered coefficient, we also provide a generalization of Pearson correlation coefficient for functional data and propose an estimate for it.) The consistency and asymptotic normality of the proposed estimator are established. Based on the asymptotic normality, we provide a formula to compute the standard error of the resulting estimate. The small sample performance of the proposed standard error formula is investigated via Monte Carlo simulation, and we found that it is good for practical use. Statistical inferences on the concordance correlation coefficient are also discussed. A physiological data set is used to illustrate the proposed methodology.

This paper is organized as follows. In Section 2, we give the motivation and introduce a concordance correlation coefficient for curve data and image data. We then propose an estimator for the coefficient. The consistency and asymptotic normality of the proposed estimator are established. Section 3 contains simulation results and illustration of the proposed method by analyzing the physiology data set aforementioned. Conclusions are given in Section 4. Regularity conditions, technical lemmas and proof of Theorem 2.1 are given in the appendix.

## 2. Concordance correlation coefficient for functional data

Let  $x$  and  $y$  denote scores from two raters or measurements from two instruments. Let us first consider that both  $x$  and  $y$  are univariate. Suppose that  $(x_1, y_1), \dots, (x_n, y_n)$  are independent and identically distributed paired observations

from  $(x, y)$ . Denote  $\mathbf{x} = (x_1, \dots, x_n)$  and  $\mathbf{y} = (y_1, \dots, y_n)$ . It is said that  $\mathbf{x}$  and  $\mathbf{y}$  are in perfect agreement if  $x_i = y_i$  for  $i = 1, \dots, n$ . Therefore, if  $\mathbf{x}$  and  $\mathbf{y}$  are in perfect agreement, then the angle between  $\mathbf{x}$  and  $\mathbf{y}$ , denoted by  $\theta(\mathbf{x}, \mathbf{y})$ , is 0, which implies that  $\cos \theta = 1$ . It is well known that

$$\cos \theta = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|},$$

where  $\langle \cdot, \cdot \rangle$  denotes the inner product, and  $\|\cdot\|$  is the Euclidean norm in  $R^n$ . In other words,  $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i$ , and  $\|\mathbf{x}\|^2 = \sum_{i=1}^n x_i^2$  for any  $n$ -dimensional vectors  $\mathbf{x}$  and  $\mathbf{y}$ . However,  $\cos \theta$  is not a good measure for agreement because for any positive constant  $c$ ,

$$\cos \theta(c\mathbf{x}, \mathbf{y}) = \cos \theta(\mathbf{x}, c\mathbf{y}) = \cos \theta(\mathbf{x}, \mathbf{y}).$$

This implies that  $\cos \theta$  cannot detect a scale change on either  $\mathbf{x}$  or  $\mathbf{y}$ .

The sample correlation coefficient between  $\mathbf{x}$  and  $\mathbf{y}$  can be written as

$$\rho(\mathbf{x}, \mathbf{y}) = \frac{\langle \mathbf{x} - \bar{x}, \mathbf{y} - \bar{y} \rangle}{\|\mathbf{x} - \bar{x}\| \|\mathbf{y} - \bar{y}\|},$$

where  $\bar{x}$  and  $\bar{y}$  are the sample means of  $\mathbf{x}$  and  $\mathbf{y}$ , respectively. If  $\mathbf{x}$  and  $\mathbf{y}$  are in perfect agreement, then  $\rho(\mathbf{x}, \mathbf{y}) = 1$ . However, it cannot detect a location shift on either  $\mathbf{x}$  or  $\mathbf{y}$  and a scale change on either  $\mathbf{x}$  or  $\mathbf{y}$  because for any constants  $a > 0$  and  $b$

$$\rho(a\mathbf{x} + b, \mathbf{y}) = \rho(\mathbf{x}, a\mathbf{y} + b) = \rho(\mathbf{x}, \mathbf{y}).$$

The degree of concordance between  $x$  and  $y$  can be characterized by the expected value of the squared difference  $E(x - y)^2$ . Using this characterization, Lin [12] proposed the concordance correlation coefficient

$$\rho_c(x, y) = \frac{2 \operatorname{cov}(x, y)}{\operatorname{var}(x) + \operatorname{var}(y) + \{E(x) - E(y)\}^2}. \tag{2.1}$$

The concordance correlation coefficient strikes a balance between a correlation measure insensitive to location differences and a measure of location discrepancy. It can be further written as the product of the accuracy and the precision coefficients  $\rho_c = \rho C_b$ . The accuracy coefficient  $C_b$  is a bias correction factor that measures how far the best-fit line deviates from the identity line.  $C_b = \frac{2}{w+1/w+v^2}$ , where  $w = \sqrt{\operatorname{var}(y)/\operatorname{var}(x)}$  and  $v^2 = \{E(x) - E(y)\}^2 / \sqrt{\operatorname{var}(x) \operatorname{var}(y)}$ . Note that  $0 < C_b \leq 1$ . The further  $C_b$  is from 1, the greater the deviation from the identity line. The precision coefficient is the Pearson correlation coefficient  $\rho$ . It measures how far are the observations from the best-fit line. See [12] for more detailed discussion.

Evaluation of reproducibility for functional data is a frequently encountered practical problem. In this section, we propose a concordance correlation coefficient for functional data to address the problem. To get more insights into the concordance correlation coefficient and for ease of presentation, we first concentrate on the case when data are collected over an interval of a real line.

Suppose that  $X(t)$  and  $Y(t)$ ,  $t \in \mathcal{I}$ , a finite closed real interval, are measurements from two instruments. Regard  $X(\cdot)$  and  $Y(\cdot)$  as two random elements in some

probability functional space  $\mathcal{S}$ . For the probability functional space  $\mathcal{S}$ , define an inner product

$$\langle X(\cdot), Y(\cdot) \rangle = E \int_{\mathcal{S}} X(t)Y(t) w(t) dt, \tag{2.2}$$

where  $w(\cdot)$  is a weight function and takes non-negative values over  $\mathcal{S}$ .

Using the notion of inner product, define the correlation coefficient for two random elements  $X$  and  $Y$  in  $\mathcal{S}$  as

$$\rho(X, Y) = \frac{\langle X - E(X), Y - E(Y) \rangle}{\|X - E(X)\| \|Y - E(Y)\|}, \tag{2.3}$$

where  $\|X\| = \sqrt{\langle X, X \rangle}$ , both  $E(X) = E X(t)$  and  $E(Y) = E Y(t)$  are functions of  $t$ .

Define a concordance correlation coefficient for  $X(\cdot)$  and  $Y(\cdot)$  to be

$$\rho_c(X, Y) = \frac{2\langle X - E(X), Y - E(Y) \rangle}{\|E(X) - E(Y)\|^2 + \|X - E(X)\|^2 + \|Y - E(Y)\|^2}. \tag{2.4}$$

Both  $\rho$  and  $\rho_c$  depend on the weight function  $w(\cdot)$ . The weight function allows one to assign importance to different parts of  $t$ . The choice of the weight function will be discussed in Section 2.2. No matter what the weight function is, the concordance correlation coefficient possesses the following characteristics, same as those for  $\rho_c$  defined on two random variables (see [11]):

- (a)  $|\rho_c| \leq |\rho| \leq 1$ . Also,  $\rho_c$  and  $\rho$  have the same sign.
- (b)  $\rho_c = \rho$  if and only if  $\|E(X) - E(Y)\| = 0$  and  $\|X - E(X)\| = \|Y - E(Y)\|$ .
- (c)  $\rho_c = 0$  if and only if  $\rho = 0$ .
- (d)  $\rho_c = \pm 1$  if and only if  $\rho = \pm 1$ ,  $\|X - E(X)\| = \|Y - E(Y)\|$ , and  $\|E(X) - E(Y)\| = 0$ .

Using the Cauchy–Schwarz inequality and the definition of  $\rho_c$ , the proofs of these characteristics are straightforward and are omitted here.

### 2.1. Statistical inferences

In this section, we propose statistical inference procedures for  $\rho$  and  $\rho_c$  defined in (2.3) and (2.4). Without loss of generality, we assume that the interval  $\mathcal{S} = [0, 1]$ . For ease of presentation, suppose that for subject  $i$ ,  $i = 1, \dots, n$ ,  $(X_i(t), Y_i(t))$  was observed at  $t = t_j, j = 1, \dots, N$  with  $0 \leq t_1 < \dots < t_N \leq 1$ . This implies that all subjects were observed at the same  $t_j, j = 1, \dots, N$ . In practice, different subjects may be observed at different time  $t$ . In such a situation, one may use interpolation to compute the sample means in (2.5) and (2.6) defined below. Thus, the proposed estimation procedure is still applicable. See, for instance, [5] for implementation with linear interpolation. When sampling rate is high, the bias caused by the approximation is small and asymptotically negligible.

Denote  $\Delta_j = t_{j+1} - t_j$ , the gap size between  $t_{j+1}$  and  $t_j$ . Using the sample counterparts to estimate  $\rho$  and  $\rho_c$ , we have

$$\hat{\rho} = \frac{\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^N \{X_i(t_j) - \bar{X}(t_j)\} \{Y_i(t_j) - \bar{Y}(t_j)\} w(t_j) \Delta_j}{\left\{ \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^N (X_i(t_j) - \bar{X}(t_j))^2 w(t_j) \Delta_j \right\}^{1/2} \left\{ \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^N (Y_i(t_j) - \bar{Y}(t_j))^2 w(t_j) \Delta_j \right\}^{1/2}} \tag{2.5}$$

and

$$\hat{\rho}_c = \frac{\frac{2}{n} \sum_{i=1}^n \sum_{j=1}^N \{X_i(t_j) - \bar{X}(t_j)\} \{Y_i(t_j) - \bar{Y}(t_j)\} w(t_j) \Delta_j}{\sum_{j=1}^N \{ \bar{X}(t_j) - \bar{Y}(t_j) \}^2 w(t_j) \Delta_j + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^N [ \{X_i(t_j) - \bar{X}(t_j)\}^2 + \{Y_i(t_j) - \bar{Y}(t_j)\}^2 ] w(t_j) \Delta_j}, \tag{2.6}$$

where  $\bar{X}(t_j) = n^{-1} \sum_{i=1}^n X_i(t_j)$  and  $\bar{Y}(t_j) = n^{-1} \sum_{i=1}^n Y_i(t_j)$  are the sample means of  $X(t_j)$  and  $Y(t_j)$ , respectively. We establish the asymptotic normality of the newly proposed estimator in the following theorem, whose proof is given in the appendix.

**Theorem 2.1.** *Suppose that  $\{(X_i(t), Y_i(t)), t = t_j, j = 1, \dots, N, i = 1, \dots, n\}$  is a random realization from  $\{(X(t), Y(t))\}$  at  $t = t_j, j = 1, \dots, N$ . If conditions (A)–(D) given in the appendix hold, then  $\hat{\rho}_c$  is a consistent estimator for  $\rho_c$ , and when  $|\rho_c| < 1$ ,  $\sqrt{n}(\hat{\rho}_c - \rho_c)$  has an asymptotic normal distribution with zero mean and variance  $\sigma_{\rho_c}^2 = a^T \Sigma a$ , where*

$$a = \frac{(2, -\rho_c, -\rho_c, 2\rho_c)^T}{E \int_{\mathcal{J}} X^2(t) w(t) dt + E \int_{\mathcal{J}} Y^2(t) w(t) dt - 2E \int_{\mathcal{J}} X(t) Y(t) w(t) dt} \tag{2.7}$$

and

$$\Sigma = \text{cov} \left\{ \begin{pmatrix} \int_{\mathcal{J}} \{X(t) - EX(t)\} \{Y(t) - EY(t)\} w(t) dt \\ \int_{\mathcal{J}} X^2(t) w(t) dt \\ \int_{\mathcal{J}} Y^2(t) w(t) dt \\ \int_{\mathcal{J}} \{X(t) - EX(t)\} EY(t) + Y(t) EX(t) w(t) dt \end{pmatrix} \right\}. \tag{2.8}$$

To estimate the standard error of  $\hat{\rho}_c$ , we estimate  $\sigma_{\rho_c}^2$  by its sample counterparts:

$$\hat{\sigma}_{\rho_c}^2 = \hat{a}^T \hat{\Sigma} \hat{a},$$

where  $\hat{a}$  and  $\hat{\Sigma}$  are the corresponding sample counterparts. Thus, using Theorem 2.1, standard error formulae for  $\hat{\rho}_c$  is

$$\text{SE}(\hat{\rho}_c) = \hat{\sigma}_{\rho_c} / \sqrt{n - 3}, \tag{2.9}$$

where we use the factor  $1/\sqrt{n - 3}$  rather than  $1/\sqrt{n - 2}$  for small sample correction in practice. The accuracy of these two standard error formulae for small-to-moderate sample sizes will be examined in Section 3, and we find that they perform well even for small sample sizes.

Since the range of  $\rho_c$  is  $[-1, 1]$ , one can improve upon the normal approximation by using Fisher’s  $Z$ -transformation. That is,

$$\hat{Z}_c = \frac{1}{2} \ln \frac{1 + \hat{\rho}_c}{1 - \hat{\rho}_c}.$$

Using the  $\delta$ -method, it can be shown that  $\sqrt{n}(\hat{Z}_c - \frac{1}{2} \ln \frac{1 + \rho_c}{1 - \rho_c})$  has asymptotic normal distributions with zero means and variance

$$\sigma_{z_c}^2 = \sigma_{\rho_c}^2 / (1 - \rho_c^2)^2.$$

Thus,

$$\hat{\sigma}_{z_c}^2 = \hat{\sigma}_{\rho_c}^2 / (1 - \hat{\rho}_c^2)^2.$$

Furthermore, asymptotic  $100(1 - \alpha)\%$  confidence intervals for  $Z_c = 2^{-1} \ln\{(1 + \rho_c)/(1 - \rho_c)\}$  is

$$\hat{Z}_c \pm t_{n-3}(1 - \alpha/2)\hat{\sigma}_{z_c} / \sqrt{n - 3}. \tag{2.10}$$

Using (2.10), we may construct asymmetric confidence intervals for  $\rho_c$ . The performance of the standard error formula will be examined in Section 3.

**Remark.** The consistence and asymptotic normality of  $\hat{\rho}$  can be derived in the same manner as those of  $\hat{\rho}_c$ . Furthermore, standard error formula can be derived and confidence interval for  $\rho$  can be constructed in the same approach as those for  $\rho_c$ . See [11] for details.

### 2.2. Choice of weight function

The definition of inner product (2.2) over the probability space  $\mathcal{S}$  involves a weight function. In this section, we briefly discuss how to determine the weight function for the proposed procedures.

The weight function allows us to assign importance to different time  $t$ . Therefore, when some prior information on the importance of different time intervals is available, one may determine the weight function accordingly. We call this kind of approach subjective approach because it is based on experience of data analysts or prior information.

In some pilot studies, there is no prior information, and hence a data-driven approach to choosing the weight function is desirable. We refer to this type of approach as objective approach. The points  $t_1, \dots, t_N$  at which data were collected are regarded as realizations of a random variable. In other words, regard  $t$  as a random variable defined on  $\mathcal{S}$ . We choose the density function of  $t$  as the weight function. Thus, the inner product defined in (2.2) is indeed an expectation of  $XY$ , which is consistent with the definition of an inner product in a probability space. Thus, to estimate the weight function, we need to estimate the density of  $t$  based on the data  $t_1, \dots, t_N$ . There are a large amount of literature on the topic of density estimation. See the reference book, for example, Silverman [17] for details. Here we

estimate the density function via a kernel estimator:

$$\hat{w}(t) = \frac{1}{Nh} \sum_{j=1}^N K\{(t_j - t)/h\},$$

where  $K(\cdot)$  is a kernel density function, such as the Gaussian density function, and  $h$  is a bandwidth to be chosen. A rule of thumb suggests taking  $h = 1.06s_t N^{-1/5}$  for the Gaussian kernel, where  $s_t$  is the sample standard deviation of  $t_1, \dots, t_N$ .

In many case studies, such as the example in Section 3.2, the data were evenly collected over time. Without subjective information, the objective approach implies taking a uniform weight function. This will be implemented in Section 3.

### 2.3. Concordance correlation coefficient for image data

Suppose that  $X(\mathbf{t})$  and  $Y(\mathbf{t})$ ,  $\mathbf{t} \in \mathbf{I}$ , a closed rectangle of the plane  $R^2$ , are image data produced by two instruments. Let  $X(\mathbf{t})$  and  $Y(\mathbf{t})$  be two elements in a probability functional space  $\mathcal{S}$ , and define an inner product for  $\mathcal{S}$  as

$$\langle X, Y \rangle = E \int_{\mathbf{I}} X(\mathbf{t}) Y(\mathbf{t}) w(\mathbf{t}) d\mathbf{t}.$$

Thus, we can define a correlation coefficient and a concordance correlation coefficient in the same manner as (2.3) and (2.4).

$$\rho(X, Y) = \frac{\langle X - E(X), Y - E(Y) \rangle}{\|X - E(X)\| \|Y - E(Y)\|} \tag{2.11}$$

and

$$\rho_c(X, Y) = \frac{2 \langle X - E(X), Y - E(Y) \rangle}{\|X - E(X)\|^2 + \|Y - E(Y)\|^2 + \|E(X) - E(Y)\|^2}. \tag{2.12}$$

The characteristics of  $\rho_c$  in Section 2 are still valid.

Here we focus on the estimation of  $\rho$  and  $\rho_c$ . Without loss of generality, it is assumed that the rectangle  $\mathbf{I} = [0, 1]^2$ . Suppose that  $X_i(\mathbf{t}), Y_i(\mathbf{t})$  was observed at  $\mathbf{t}_{kl} = (t_{1k}, t_{2l})$ ,  $k = 1, \dots, K$  and  $l = 1, \dots, L$  with  $0 \leq t_{11} < \dots < t_{1K} \leq 1$  and  $0 \leq t_{21} < \dots < t_{2L} \leq 1$ . Define  $N = KL$  and  $\Delta_{kl} = (t_{1,k+1} - t_{1,k})(t_{2,j+1} - t_{2,j})$ . Using the sample counterparts to estimate  $\rho$  and  $\rho_c$ , we have

$$\hat{\rho} = \frac{\frac{1}{n} \sum_{i,k,l} \{X_i(\mathbf{t}_{kl}) - \bar{X}(\mathbf{t}_{kl})\} \{Y_i(\mathbf{t}_{kl}) - \bar{Y}(\mathbf{t}_{kl})\} w(\mathbf{t}_{kl}) \Delta_{kl}}{\left\{ \frac{1}{n} \sum_{i,k,l} (X_i(\mathbf{t}_{kl}) - \bar{X}(\mathbf{t}_{kl}))^2 w(\mathbf{t}_{kl}) \Delta_{kl} \right\}^{1/2} \left\{ \frac{1}{n} \sum_{i,k,l} (Y_i(\mathbf{t}_{kl}) - \bar{Y}(\mathbf{t}_{kl}))^2 w(\mathbf{t}_{kl}) \Delta_{kl} \right\}^{1/2}} \tag{2.13}$$



and

$$\hat{\rho}_c = \frac{\frac{2}{n} \sum_{i,k,l} \{X_i(\mathbf{t}_{kl}) - \bar{X}(\mathbf{t}_{kl})\} \{Y_i(\mathbf{t}_{kl}) - \bar{Y}(\mathbf{t}_{kl})\} w(\mathbf{t}_{kl}) \Delta_{kl}}{\frac{1}{n} \sum_{i,k,l} [\{X_i(\mathbf{t}_{kl}) - \bar{X}(\mathbf{t}_{kl})\}^2 + \{Y_i(\mathbf{t}_{kl}) - \bar{Y}(\mathbf{t}_{kl})\}^2] w(\mathbf{t}_{kl}) \Delta_{kl} + \sum_{k,l} \{\bar{X}(\mathbf{t}_{kl}) - \bar{Y}(\mathbf{t}_{kl})\}^2 w(\mathbf{t}_{kl}) \Delta_{kl}}. \tag{2.14}$$

Similar to Theorem 2.1,  $\hat{\rho}_c$  is consistent and has an asymptotic normal distribution.

**Theorem 2.2.** *Suppose that  $\{(X_i(\mathbf{t}_{kl}), Y_i(\mathbf{t}_{kl}))\}, i = 1, \dots, n, k = 1, \dots, K, l = 1, \dots, L$  is a random sample from  $\{X(\mathbf{t}), Y(\mathbf{t})\}$  at  $t_{kl}$  for  $k = 1, \dots, K, l = 1, \dots, L$ . If conditions (A), (B), (C') and (D') given in the appendix hold, then the  $\hat{\rho}_c$ , defined in (2.14), is a consistent estimator for  $\rho_c$ , and  $\sqrt{n}(\hat{\rho}_c - \rho_c)$  has an asymptotic normal distribution with zero mean and variance  $\sigma_{\rho_c}^2 = b^T \Sigma_2 b$ , where*

$$b = \frac{(2, -\rho_c, -\rho_c, 2\rho_c)^T}{E \int_{\mathbf{1}} X^2(\mathbf{t})w(\mathbf{t}) dt + E \int_{\mathbf{1}} Y^2(\mathbf{t})w(\mathbf{t}) dt - 2E \int_{\mathbf{1}} X(\mathbf{t})Y(\mathbf{t})w(\mathbf{t}) dt} \tag{2.15}$$

and

$$\Sigma_2 = \text{cov} \left\{ \begin{pmatrix} \int_{\mathbf{1}} \{X(\mathbf{t}) - EX(\mathbf{t})\} \{Y(\mathbf{t}) - EY(\mathbf{t})\} w(\mathbf{t}) dt \\ \int_{\mathbf{1}} X^2(\mathbf{t})w(\mathbf{t}) dt \\ \int_{\mathbf{1}} Y^2(\mathbf{t})w(\mathbf{t}) dt \\ \int_{\mathbf{1}} \{X(\mathbf{t}) - EX(\mathbf{t})\} EY(\mathbf{t}) + Y(\mathbf{t}) EX(\mathbf{t}) w(\mathbf{t}) dt \end{pmatrix} \right\}. \tag{2.16}$$

**Proof.** Using Lemma A.2 in the appendix, Theorem 2.2 follows by similar arguments in the proof of Theorem 2.1.  $\square$

As discussed in Section 2, we may apply the Fisher Z-transformation for  $\hat{\rho}$  and  $\hat{\rho}_c$  to obtain a better normal approximation. Furthermore, one may construct an asymptotic confidence interval using the asymptotic properties in Theorem 2.1.

### 3. Simulation study and applications

In this section, we investigate the small sample performance of the proposed estimators and test the accuracy of proposed standard error formulae in (2.9) when the sample size is small. We also investigate the performance of the confidence interval in (2.10) in terms of coverage probability via Monte Carlo simulations. We then apply the proposed method to a real data set. In our simulations and analysis of the real data set, the weight function  $w(t)$  involved in the definition of  $\rho$  and  $\rho_c$  is taken to be 1.

### 3.1. Simulation study

To assess the performance of proposed estimators in (2.5) and (2.6) and their standard error formulae, a Monte Carlo simulation was conducted for four underlying  $K$ -dependent Gaussian processes (see [11] for an algorithm to generate the simulated data) and four underlying  $K$ -dependent non-Gaussian processes  $(X(t), Y(t))$  with mean  $(\mu_x(t), \mu_y(t))$ , variance  $(\sigma_x^2(t), \sigma_y^2(t))$  and covariance  $\text{cov}(X(t), Y(t)) = \sigma_{xy}(t)$ . For the non-Gaussian processes, both  $X(t)$  and  $Y(t)$  are generated from a  $t$ -distribution with 10 degrees of freedom with a location shift and a scale change. For each of the  $K$ -dependent process, we take  $K = 20, 40$  and set the sample size  $n = 10, 20$ . We generate  $N = 50, 100$  sample points over a period of time  $[0, 1]$ . Thus, for each case of 64 ( $= 2 \times 4 \times 2 \times 2 \times 2$ ) situations, we conduct 1000 Monte Carlo simulations using MATLAB.

*Case 1:* In this case,  $\mu_x(t) = \mu_y(t) = 0$ ,  $\sigma_x^2(t) = \sigma_y^2(t) = 1$  and  $\sigma_{xy} = 0.95$ . In this setting,  $\rho = \rho_c = 0.95$  with no difference in location and scale parameters.

*Case 2:* Let  $\mu_x(t) = -\sqrt{0.05}t$ ,  $\mu_y(t) = \sqrt{0.05}t$ ,  $\sigma_x^2 = \sigma_y^2(t) = 1$  and  $\sigma_{xy}(t) = 0.95$ . In this case, both the pointwise correlation coefficient and the pointwise concordance correlation coefficient are close to 1. Moreover,  $\rho = 0.95$  and  $\rho_c = 0.9048$  with a slight location shift.

*Case 3:* Let  $\mu_x(t) = -\sqrt{0.1}/2$ ,  $\mu_y(t) = \sqrt{0.1}/2$ ,  $\sigma_x^2(t) = 1.1^2$ ,  $\sigma_y^2(t) = 0.9^2$ ,  $\sigma_{xy}(t) = 1.1 \times 0.9 \times \{\sin(2\pi t) + 1\}/2$ . In this example,  $\rho = 0.5$  and  $\rho_c = 0.4670$  with slight constant differences in both locations and variance and a varying covariance function. Furthermore, both the pointwise correlation coefficient and the pointwise concordance correlation coefficient vary between 0 to 1.

*Case 4:* Let  $\mu_x(t) = -\sqrt{0.05}t$ ,  $\mu_y(t) = \sqrt{0.05}t$ ,  $\sigma_x^2(t) = 1.1^2$ ,  $\sigma_y^2(t) = 0.9^2$ , and  $\sigma_{xy} = 1.1 \times 0.9 \times \{\sin(2\pi t) + 3\}/4$ . In this case, the difference in location and covariance between  $X(t)$  and  $Y(t)$  are varying in  $t$ , but the difference in variance is a constant. In the example, we have  $\rho = 0.75$  and  $\rho_c = 0.7005$ . The pointwise correlation coefficient and the pointwise concordance correlation coefficient vary between 0.5 and 1.

These four simulation settings were motivated by actual situations that arise in practice. Cases 1 and 2 correspond to highly correlated and highly agreed paired functional data and were motivated from the real example studied in [11]; and the means and variances in Cases 3 and 4 were taken the same as those of Cases 3 and 4 in [12]. The pointwise correlation function in Cases 3 and 4 periodically changes over time, which was also motivated by the actual data example in Section 3.2. Table 1 summarizes the simulation results of the case in which the  $(X(t), Y(t))$  is  $K$ -Gaussian process with  $K = 20$ ,  $N = 50$  and  $n = 10$ . The simulation results for some other cases are similar. See [11] for more simulation results.

From Table 1, we can see that the averages of estimates of  $\hat{\rho}$ ,  $\hat{Z}$ ,  $\hat{\rho}_c$  and  $\hat{Z}_c$  are very close to the true values, although we notice that  $\hat{\rho}_c$  slightly underestimate  $\rho_c$ . The average of estimated standard errors (SE) is very close to the standard deviation of 1000 estimates, which can be regarded as the true value of the standard deviation of

Table 1  
Simulation results

	mean(std)	SE(std)		mean(std)	SE(std)	CP
Case 1:	$\rho = \rho_c = 0.95$			$Z = Z_c = 1.832$		
$\hat{\rho}$	0.950(0.017)	0.016(0.007)	$\hat{Z}$	1.853(0.169)	0.159(0.042)	0.954
$\hat{\rho}_c$	0.943(0.018)	0.017(0.007)	$\hat{Z}_c$	1.791(0.162)	0.157(0.042)	0.954
Case 2:	$\rho = 0.95, \rho_c = 0.905$			$Z = 1.832, Z_c = 1.498$		
$\hat{\rho}$	0.950(0.017)	0.016(0.007)	$\hat{Z}$	1.853 (0.169)	0.159(0.042)	0.954
$\hat{\rho}_c$	0.899(0.030)	0.029(0.011)	$\hat{Z}_c$	1.491 (0.158)	0.155(0.040)	0.954
Case 3:	$\rho = 0.5, \rho_c = 0.467$			$Z = 0.549, Z_c = 0.506$		
$\hat{\rho}$	0.497(0.139)	0.125(0.041)	$\hat{Z}$	0.563 (0.187)	0.169(0.043)	0.935
$\hat{\rho}_c$	0.442(0.128)	0.121(0.036)	$\hat{Z}_c$	0.486 (0.161)	0.153(0.041)	0.949
Case 4:	$\rho = 0.75, \rho_c = 0.701$			$Z = 0.973, Z_c = 0.868$		
$\hat{\rho}$	0.747(0.084)	0.074(0.031)	$\hat{Z}$	0.993(0.191)	0.169(0.044)	0.939
$\hat{\rho}_c$	0.678(0.091)	0.082(0.030)	$\hat{Z}_c$	0.845(0.172)	0.154(0.041)	0.938

The column labeled “mean(std)” presents the mean and standard deviation of the 1000 estimates of the coefficients in the 1000 simulations. The column labeled “SE(std)” lists the mean and standard deviation of the 1000 estimates of standard errors defined in (2.9). CP stands for the coverage probability of the confidence interval in (2.10) at the significance level  $\alpha = 0.05$ .

$\hat{\rho}$  and  $\hat{\rho}_c$ . The difference between the average of the estimated standard errors and the true value is less than half standard deviation of the estimated standard errors. This implies that the standard error formulae proposed in (2.9) are fairly accurate even for small sample sizes. We also notice that the SE is always less than the true standard deviation. This may imply that the standard error formulae proposed in (2.9) slightly underestimate the true standard error. Note that the Monte Carlo standard error for the coverage probability is 0.0069. It can be seen from the last column of Table 1 that most of the coverage probabilities are very close to the true probability 0.95. This implies that the proposed confidence interval formula in (2.10) work well even for small sample sizes.

### 3.2. An application

**Example.** In this example, we illustrate the proposed method in Section 2 by application to an actual data set collected in the Noll Physiological Research center at the Pennsylvania State University. The researchers of study is interested in measuring body core temperature and to compare agreement of measurement by two different methods under the active heating condition. Data had been collected to compare the temperature recorded by the esophageal temperature (Tes) approach and the rectal temperature (Tre) approach. There are 12 subjects in this study. During this study, temperatures for each subject were observed by the two approaches every minute over one and half hour of an experiment period. The

subjects will sit quietly for 10 min and then exercise on a motor-driven treadmill for 20 min. The subjects will alternate rest with exercise on a treadmill for 3 cycles in an environmental chamber. Figs. 1(a) and (b) display the sample mean and standard deviation curves of subjects under the active heating condition, from which we can see that the two means and standard deviation functions do not agree with each other well. The sample mean curves have tended to increase over time since the experiment was performed under the active heating condition. The effect of the three rest and exercise cycles was shown clearly in the sample mean curve. The pointwise concordance correlation coefficient, depicted in Fig. 1(c), changes over time dramatically. That coincides with the researchers' expectation since Tre has a slower response time than Tes during the heating condition. Furthermore, the time points where the concordances are high or low roughly correspond to the changing points in the sample mean curve. Again this is as expected since the response time to temperature change of the two instruments are different.

We further computed the windowed version of concordance correlation coefficient at time  $t$ , using all datum points between  $t - h$  and  $t + h$ . The idea is similar to that of moving average, and  $h$  is referred to as the window size. Fig. 1(d) depicts a windowed version concordance correlation coefficient, using  $h = 9$ . Such a window size is decided such that we use about 20% datum points around time  $t$  to estimate concordance correlation coefficient at time  $t$ . We also computed the concordance correlation coefficient using about 10% and 30% datum points around  $t$ . The shape of concordance correlation coefficient is similar to that in Fig. 1(d). We further construct a confidence interval for the windowed version concordance correlation coefficient using bootstrap method. Here we bootstrap 1000 subject-based samples rather than observations-based samples. The dotted lines are the 2.5th and 97.5th percentiles of the 1000 bootstrap estimates and provide us a 95% confidence interval. One may construct an asymptotic pointwise confidence interval for pointwise concordance correlation coefficient (see [12]), but the confidence interval depicted in Fig. 1(c) is the bootstrap confidence interval due to the small sample size.

Fig. 1(e) and (f) display the scatter plot at two typical time points:  $t = 10$  and 80. To assess overall agreement of these two approaches, we compute  $\hat{\rho}$  and  $\hat{\rho}_c$  and obtain that  $\hat{\rho} = 0.8637$  with standard error 0.0574, and  $\hat{\rho}_c = 0.6162$  with standard error 0.1222. Although the correlation coefficient is pretty high, the low value of  $\hat{\rho}_c$  implies the measurements by the two approaches do not agree with each other. This is also evidenced by Figs. 1(a), (b), (e) and (f). This example shows that  $\rho_c$  is more effective than  $\rho$  in detecting non-agreement of measurement.

We now employ bootstrap samples to estimate the sampling distributions of  $\hat{\rho}$  and  $\hat{\rho}_c$ . Fig. 2 depicts the densities of  $\hat{\rho}$  and  $\hat{\rho}_c$  using kernel density estimation with Gaussian kernel and bandwidth chosen using the plug-in method proposed by Sheather and Jones [16]. The sample mean and standard deviation of the 1000 bootstrap estimate for  $\hat{\rho}_c$  are 0.5819 and 0.1140, respectively. They approximately equal to the estimate and standard error of  $\rho_c$ . Result for  $\hat{\rho}$  is similar. Based on the 1000 bootstrap estimates, a 95% confidence interval for  $\hat{\rho}_c$  and  $\hat{\rho}$  is [0.3261, 0.7803] and [0.7011, 0.9481], respectively. Since the sampling distributions for  $\hat{\rho}_c$  and  $\hat{\rho}$  are skew, the sample means are not the center of the confidence intervals.

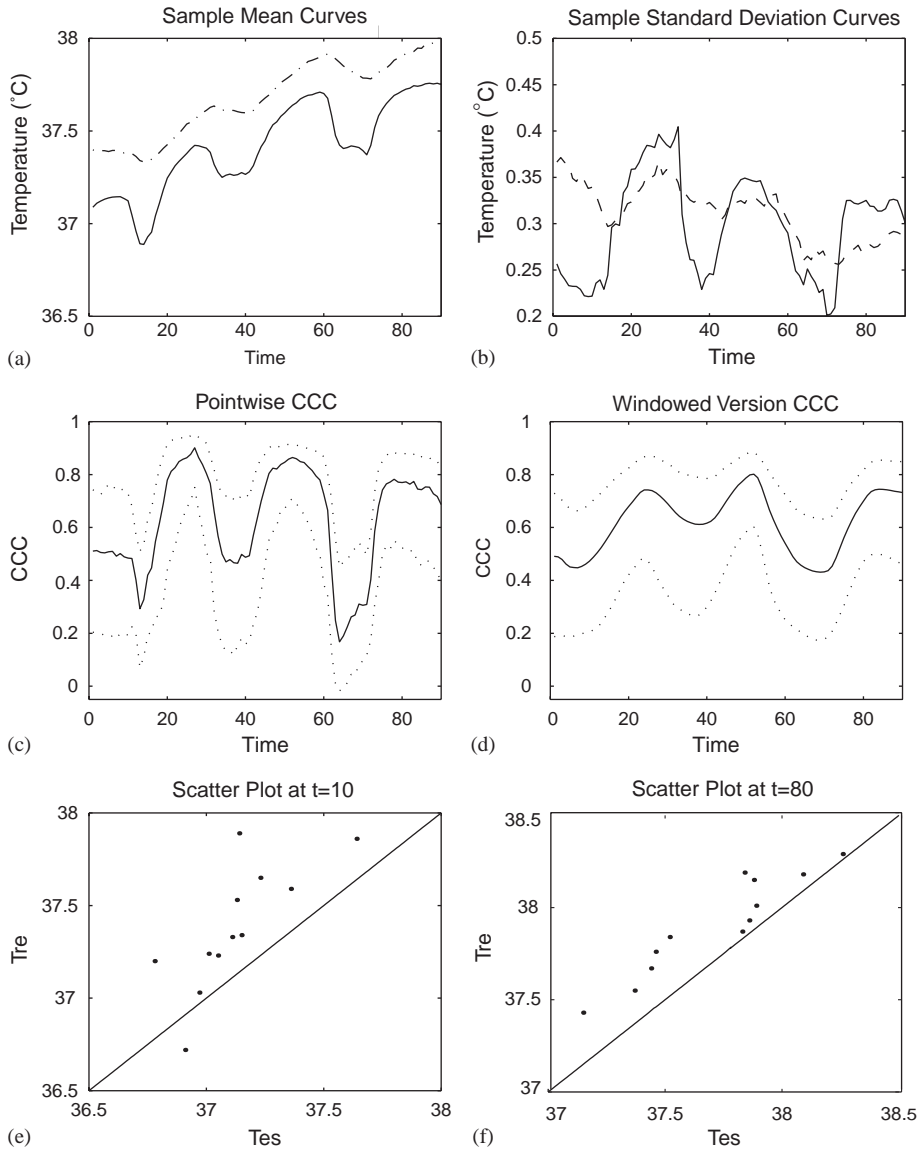


Fig. 1. Plots for Example 1. (a) and (b) are plots of the sample mean and standard deviation curves, respectively. The solid line stands for the  $T_{es}$ , and the dash-dotted line for the  $T_{re}$ . In (c) and (d), the solid line is plot of pointwise and windowed version concordance correlation coefficient, respectively, and dotted lines are the 2.5th and 97.5th percentiles of 1000 bootstrap estimates, respectively. (e) and (f) are scatter plots at  $t = 10$  and 80, respectively.

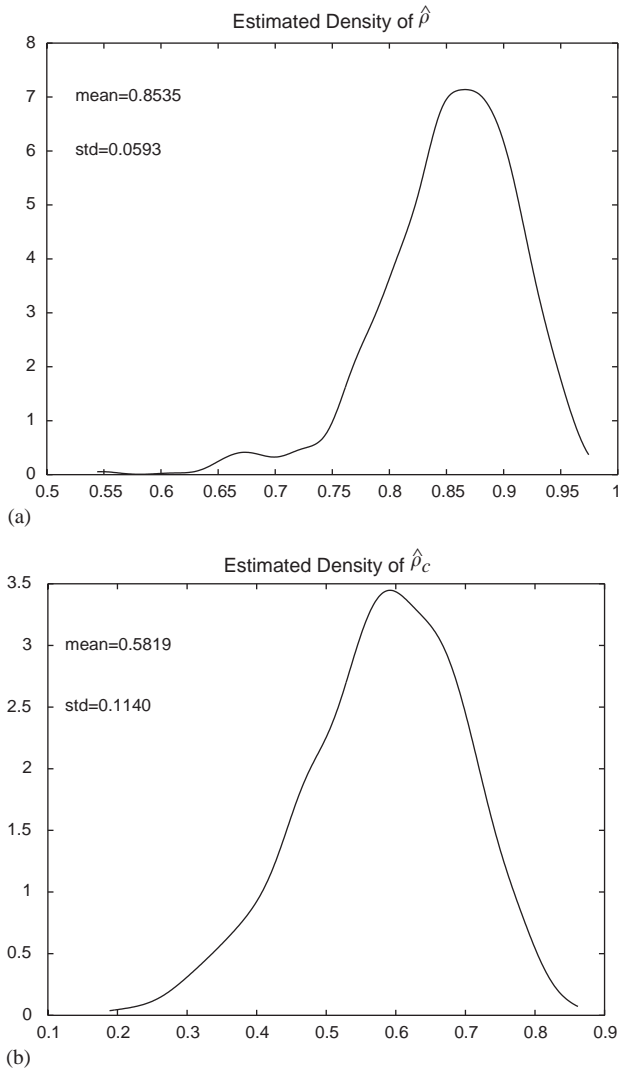


Fig. 2. Plot of Estimated Densities of  $\hat{\rho}$  and  $\hat{\rho}_c$  Based on 1000 Bootstrap Estimates.

#### 4. Conclusions

In this paper, we proposed a concordance correlation coefficient for curve data and image data. Its characteristics have been investigated. We proposed an estimator for the concordance correlation coefficient and established the asymptotic normality of the proposed estimator. A standard error formula for the resulting estimate is derived and empirically tested. An application to an actual data set illustrates the proposed methodology.

### Acknowledgments

The authors thank the referees and Professor Vernon M. Chinchilli for their constructive comments and suggestions that substantially improved an earlier draft. The authors are grateful to Drs. W. L. Kenney and Jane Piergza for providing the background of the physiological data set illustrated in Section 3. The services provided by the General Clinical Research Center of the Pennsylvania State University are appreciated. Li’s research was supported by National Science Foundation (NSF) grants DMS-0102505 and DMS-0348869, and a National Institute on Drug Abuse (NIDA) Grant 1-P50-DA10075. Chow’s research was supported by a NIH Grant M01-RR10732.

### Appendix

We first present the regularity conditions for Theorem 2.1.

**Definition A.1.** Let  $0 = t_0 < \dots < t_l = 1$  be any partition  $\mathcal{P}$  of  $[0, 1]$ , and  $f(t)$  be a real function defined on  $[0, 1]$ . If the variation

$$V_{\mathcal{P}} = \sum_{l=1}^l |f(t_{j+1}) - f(t_j)|$$

has an upper bound which is independent of the choice of  $\mathcal{P}$ , then  $f$  is called a function of bounded variation. The least upper bound of  $V_{\mathcal{P}}$  is called the total variation of  $f$  and is denoted by  $V(f)$ .

For a function of bounded variation, it has been shown that

$$\left| \int_0^1 f(t) dt - \sum_{j=1}^N f(t_j) \Delta_j \right| \leq V(f) \max_{0 \leq j \leq N-1} |t_{j+1} - t_j|. \tag{A.1}$$

See, for example, Theorem 5.3 of Hua and Wang [7].

Eq. (A.1) tells us how to impose assumptions on  $V(f)$  and  $\max_{0 \leq j \leq N-1} |t_{j+1} - t_j|$  in a natural way. Let  $\mathcal{X}$  and  $\mathcal{Y}$  denote the space consisting of paths  $X(\cdot)$  and  $Y(\cdot)$ . Further, define

$$\mathcal{L}_1 = \{X(\cdot)Y(\cdot) : X \in \mathcal{X} \text{ and } Y \in \mathcal{Y}\},$$

$$\mathcal{L}_2 = \{X^2(\cdot) : X \in \mathcal{X}\},$$

$$\mathcal{L}_3 = \{Y^2(\cdot) : Y \in \mathcal{Y}\}.$$

**Conditions.** (A) For  $\mathcal{F} = \mathcal{X}, \mathcal{Y}, \mathcal{L}_i, i = 1, 2$  and  $3$ ,

$$\sup_{f \in \mathcal{F}} V(f) < \infty \quad \text{a.s.} \tag{A.2}$$

(B) The weight function  $w(t)$  is a function of bounded variation.

(C) There exists a constant  $\Delta$  such that

$$\max_{0 \leq j \leq N-1} |t_{j+1} - t_j| \leq \frac{\Delta}{N}$$

and  $\sqrt{n}/N \rightarrow 0$  as  $n \rightarrow \infty$ .

(D)

$$E \left[ \int_{\mathcal{J}} X^2(t)w(t) dt \right]^2 < \infty,$$

$$E \left[ \int_{\mathcal{J}} Y^2(t)w(t) dt \right]^2 < \infty,$$

$$E \left[ \int_{\mathcal{J}} X(t)Y(t)w(t) dt \right]^2 < \infty.$$

To prove Theorem 2.1, we need the following lemma.

**Lemma A.1.** *Under the conditions of Theorem 2.1, we have*

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^N X_i(t_j) Y_i(t_j) w(t_j) \Delta_j = \frac{1}{n} \sum_{i=1}^n \int_{\mathcal{J}} X_i(t) Y_i(t) w(t) dt + O_P\left(\frac{1}{N}\right), \tag{A.3}$$

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^N X_i^2(t_j) w(t_j) \Delta_j = \frac{1}{n} \sum_{i=1}^n \int_{\mathcal{J}} X_i^2(t) w(t) dt + O_P\left(\frac{1}{N}\right), \tag{A.4}$$

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^N Y_i^2(t_j) w(t_j) \Delta_j = \frac{1}{n} \sum_{i=1}^n \int_{\mathcal{J}} Y_i^2(t) w(t) dt + O_P\left(\frac{1}{N}\right), \tag{A.5}$$

$$\sum_{j=1}^N \bar{X}^2(t_j) w(t_j) \Delta_j = \int_{\mathcal{J}} \bar{X}^2(t) w(t) dt + O_P\left(\frac{1}{N}\right), \tag{A.6}$$

$$\sum_{j=1}^N \bar{Y}^2(t_j) w(t_j) \Delta_j = \int_{\mathcal{J}} \bar{Y}^2(t) w(t) dt + O_P\left(\frac{1}{N}\right), \tag{A.7}$$

$$\sum_{j=1}^N \bar{X}(t_j) \bar{Y}(t_j) w(t_j) \Delta_j = \int_{\mathcal{J}} \bar{X}(t) \bar{Y}(t) w(t) dt + O_P\left(\frac{1}{N}\right). \tag{A.8}$$

**Proof.** Denote

$$\mathcal{G} = \{X(\cdot)Y(\cdot)w(\cdot) : X \in \mathcal{X}, Y \in \mathcal{Y}\}.$$

Since  $w(t)$  is a function of bounded variation, using (A.2), it can be shown that

$$\sup_{f \in \mathcal{G}} V(f) < \infty, \quad \text{a.s.}$$



Thus, by condition (C), it follows that

$$\begin{aligned} & \left| \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^N X_i(t_j) Y_i(t_j) w(t_j) \Delta_j - \frac{1}{n} \sum_{i=1}^n \int_{\mathcal{J}} X_i(t) Y_i(t) w(t) dt \right| \\ & \leq \frac{1}{n} \sum_{i=1}^n \left| \sum_{j=1}^N X_i(t_j) Y_i(t_j) w(t_j) \Delta_j - \int_{\mathcal{J}} X_i(t) Y_i(t) w(t) dt \right| \\ & \leq \frac{1}{n} \sum_{i=1}^n V(X_i Y_i w) \max_{0 \leq j \leq N-1} |t_{j+1} - t_j| \\ & \leq \sup_{f \in \mathcal{G}} V(f) \frac{\Delta}{N}. \end{aligned}$$

So (A.3) follows. Along the same lines as the proof of (A.3), it can be shown that (A.4)–(A.8) hold. This completes the proof of Lemma A.1.  $\square$

**Proof of Theorem 2.1.** Denote

$$A_n = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^N \{X_i(t_j) - \bar{X}(t_j)\} \{Y_i(t_j) - \bar{Y}(t_j)\} w(t_j) \Delta_j.$$

Using Lemma A.1 in the appendix, it follows that

$$\begin{aligned} A_n &= \frac{1}{n} \sum_{i=1}^n \int_{\mathcal{J}} X_i(t) Y_i(t) w(t) dt - \int_{\mathcal{J}} \bar{X}(t) \bar{Y}(t) w(t) dt + O_P\left(\frac{1}{N}\right) \\ &= \frac{1}{n} \sum_{i=1}^n \int_{\mathcal{J}} \{X_i(t) - EX(t)\} \{Y_i(t) - EY(t)\} w(t) dt \\ &\quad - \int_{\mathcal{J}} \{\bar{X}(t) - EX(t)\} \{\bar{Y}(t) - EY(t)\} w(t) dt + O_P\left(\frac{1}{N}\right). \end{aligned}$$

Since

$$E \int_{\mathcal{J}} \{\bar{X}(t) - EX(t)\} \{\bar{Y}(t) - EY(t)\} w(t) dt = O_P\left(\frac{1}{n}\right)$$

and

$$\text{var} \left\{ \int_{\mathcal{J}} \{\bar{X}(t) - EX(t)\} \{\bar{Y}(t) - EY(t)\} w(t) dt \right\} = O_P\left(\frac{1^2}{n}\right),$$

we have

$$\begin{aligned} A_n &= \frac{1}{n} \sum_{i=1}^n \int_{\mathcal{J}} \{X_i(t) - EX(t)\} \{Y_i(t) - EY(t)\} \\ &\quad \times w(t) dt + O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{N}\right) \end{aligned} \tag{A.9}$$

which tends to  $E \int_{\mathcal{J}} \{X(t) - EX(t)\} \{Y(t) - EY(t)\} w(t) dt$  in probability by the weak law of large numbers and  $N \rightarrow \infty$ . Similarly, we can show that all other terms involved in  $\hat{\rho}_c$  are convergent in probability, and therefore  $\hat{\rho}_c$  is consistent.

To establish the asymptotic normality of  $\hat{\rho}_c$ , we rewrite  $\hat{\rho}_c$  as

$$\hat{\rho}_c = \frac{\frac{2}{n} \sum_{i=1}^n \sum_{j=1}^N \{X_i(t_j) - \bar{X}(t_j)\} \{Y_i(t_j) - \bar{Y}(t_j)\} w(t_j) A_j}{\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^N X_i^2(t_j) w(t_j) A_j + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^N Y_i^2(t_j) w(t_j) A_j - 2 \sum_{j=1}^N \bar{X}(t_j) \bar{Y}(t_j) w(t_j) A_j}.$$

Denote

$$B_n = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^N X_i^2(t_j) w(t_j) A_j,$$

$$C_n = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^N Y_i^2(t_j) w(t_j) A_j$$

and

$$D_n = \sum_{j=1}^N \bar{X}(t_j) \bar{Y}(t_j) w(t_j) A_j.$$

By straightforward calculation, we have

$$B_n = \frac{1}{n} \sum_{i=1}^n \int_{\mathcal{J}} X_i^2(t) w(t) dt + O_P\left(\frac{1}{N}\right),$$

$$C_n = \frac{1}{n} \sum_{i=1}^n \int_{\mathcal{J}} Y_i^2(t) w(t) dt + O_P\left(\frac{1}{N}\right).$$

Next we deal with  $D_n$ . It is not difficult to show that

$$D_n = \int_{\mathcal{J}} \bar{X}(t) \bar{Y}(t) w(t) dt + O_P\left(\frac{1}{N}\right)$$

which equals

$$\int_{\mathcal{J}} \{\bar{X}(t) - EX(t)\} \{\bar{Y}(t) - EY(t)\} w(t) dt + \int_{\mathcal{J}} \{(\bar{X}(t) - EX(t))EY(t) + EX(t)\bar{Y}(t)\} w(t) dt + O_P\left(\frac{1}{N}\right).$$

This is equal to

$$D_n = \frac{1}{n} \sum_{i=1}^n \int_{\mathcal{J}} \{(X_i(t) - EX(t))EY(t) + EX(t)Y_i(t)\} w(t) dt + O_P\left(\frac{1}{n}\right) + O_P\left(\frac{1}{N}\right).$$

Define  $V_n = (A_n, B_n, C_n, D_n)^T$  and

$$V = \begin{pmatrix} E \int_{\mathcal{J}} \{X(t) - EX(t)\} \{Y(t) - EY(t)\} w(t) dt \\ E \int_{\mathcal{J}} X^2(t) w(t) dt \\ E \int_{\mathcal{J}} Y^2(t) w(t) dt \\ \int_{\mathcal{J}} EY(t)EX(t)w(t) dt \end{pmatrix}.$$

By condition (C), the multivariate central limit theorem and the Slutsky theorem,  $\sqrt{n}(V_n - V)$  has an asymptotic normal distribution with mean zero and variance  $\Sigma$  given by (2.8).

Define  $h(v_1, v_2, v_3, v_4) = 2v_1/(v_2 + v_3 - 2v_4)$ . Thus,  $\hat{\rho}_c = h(A_n, B_n, C_n, D_n)$ . Using the delta method,  $\sqrt{n}(\hat{\rho}_c - \rho_c)$  has an asymptotic normal distribution with mean zero and variance  $a^T \Sigma a$ , where  $a$  and  $\Sigma$  are defined in (2.7) and (2.8), respectively. This completes the proof.  $\square$

Let us present the regularity conditions for Theorem 2.2.

**Definition A.3.** Let  $0 = t_{10} < \dots < t_{1m_1} = 1$  and  $0 = t_{20} < \dots < t_{2m_2} = 1$  be any partition  $\mathcal{P}$  of  $[0, 1]^2$ , and  $f(t_1, t_2)$  be a real function defined on  $[0, 1]^2$ . Define

$$\Delta_{10}f(t_{1i}, t_2) = f(t_{1(i+1)}, t_2) - f(t_{1i}, t_2),$$

$$\Delta_{01}f(t_1, t_{2j}) = f(t_1, t_{2(j+1)}) - f(t_1, t_{2j}),$$

$$\Delta_{11}f(t_{1i}, t_{2j}) = f(t_{1i}, t_{2j}) - f(t_{1(i+1)}, t_{2j}) - f(t_{1i}, t_{2(j+1)}) + f(t_{1(i+1)}, t_{2(j+1)}).$$

If the variation

$$V_{\mathcal{P}} = \sum_{i=0}^{m_1-1} \sum_{j=0}^{m_2-1} |\Delta_{11}f(t_{1i}, t_{2j})| + \sum_{i=0}^{m_1-1} |\Delta_{10}f(t_{1i}, 1)| + \sum_{j=0}^{m_2-1} |\Delta_{01}f(1, t_{2j})|$$

has an upper bound which is independent of the choice  $\mathcal{P}$ , then  $f$  is called a function of bounded variation (see, for example, [7]). The least upper bound of  $V_{\mathcal{P}}$  is called the total variation of  $f$  and is denoted by  $V(f)$ .

**Condition.** (C') There exists  $\Delta_1$  and  $\Delta_2$  such that

$$\max_{0 \leq k \leq K} |t_{1(k+1)} - t_{1k}| \leq \frac{\Delta_1}{K}$$

and

$$\max_{0 \leq l \leq L} |t_{2(l+1)} - t_{2l}| \leq \frac{\Delta_2}{L}.$$

Further,  $\sqrt{n}/N \rightarrow 0$  as  $n \rightarrow \infty$ , where  $N = KL$ .

(D')

$$E \left[ \int_{\mathbf{I}} X^2(\mathbf{t})w(\mathbf{t}) dt \right]^2 < \infty,$$

$$E \left[ \int_{\mathbf{I}} Y^2(\mathbf{t})w(\mathbf{t}) dt \right]^2 < \infty,$$

$$E \left[ \int_{\mathbf{I}} X(\mathbf{t})Y(\mathbf{t})w(\mathbf{t}) dt \right]^2 < \infty.$$

**Lemma A.2.** Under the conditions of Theorem 4.1, we have

$$\frac{1}{n} \sum_{i,k,l} X_i(\mathbf{t}_{kl})Y_i(\mathbf{t}_{kl})w(\mathbf{t}_{kl})\Delta_{kl} = \frac{1}{n} \sum_{i=1}^n \int_{\mathbf{I}} X_i(\mathbf{t})Y_i(\mathbf{t})w(\mathbf{t}) dt + O_P\left(\frac{1}{N}\right), \tag{A.10}$$

$$\frac{1}{n} \sum_{i,k,l} X_i^2(\mathbf{t}_{kl})w(\mathbf{t}_{kl})\Delta_{kl} = \frac{1}{n} \sum_{i=1}^n \int_{\mathbf{I}} X_i^2(\mathbf{t})w(\mathbf{t}) dt + O_P\left(\frac{1}{N}\right), \tag{A.11}$$

$$\frac{1}{n} \sum_{i,k,l} Y_i^2(\mathbf{t}_{kl})w(\mathbf{t}_{kl})\Delta_{kl} = \frac{1}{n} \sum_{i=1}^n \int_{\mathbf{I}} Y_i^2(\mathbf{t})w(\mathbf{t}) dt + O_P\left(\frac{1}{N}\right), \tag{A.12}$$

$$\sum_{k,l} \bar{X}^2(\mathbf{t}_{kl})w(\mathbf{t}_{kl})\Delta_{kl} = \int_{\mathbf{I}} \bar{X}^2(\mathbf{t})w(\mathbf{t}) dt + O_P\left(\frac{1}{N}\right), \tag{A.13}$$

$$\sum_{k,l} \bar{Y}^2(\mathbf{t}_{kl})w(\mathbf{t}_{kl})\Delta_{kl} = \int_{\mathbf{I}} \bar{Y}^2(\mathbf{t})w(\mathbf{t}) dt + O_P\left(\frac{1}{N}\right), \tag{A.14}$$

$$\sum_{k,l} \bar{X}(\mathbf{t}_{kl})\bar{Y}(\mathbf{t}_{kl})w(\mathbf{t}_{kl})\Delta_{kl} = \int_{\mathbf{I}} \bar{X}(\mathbf{t})\bar{Y}(\mathbf{t})w(\mathbf{t}) dt + O_P\left(\frac{1}{N}\right). \tag{A.15}$$

**Proof.** The proof is similar to that of Lemma lemA.1. We omit it here. A rigorous proof is given in [11].  $\square$

**References**

[1] H.X. Barnhart, J.M. Williamson, Modeling concordance correlation via GEE to evaluate reproducibility, *Biometrics* 57 (2001) 931–940.  
 [2] V.M. Chinchilli, J.K. Martel, S. Kumanyika, T. Lloyd, A weighted concordance correlation coefficient for repeated measurement designs, *Biometrics* 52 (1996) 341–353.  
 [3] J. Cohen, A coefficient of agreement for nominal scales, *Educational Psychol. Measurement* 20 (1960) 37–46.  
 [4] J. Cohen, Weighted kappa: normal scale agreement with provision for scaled disagreement or partial credit, *Psychol. Bull.* 70 (1968) 213–220.

- [5] J. Fan, J. Zhang, Two-step estimation of functional linear models with applications to longitudinal data, *J. Roy. Statist. Soc. B* 62 (2000) 303–322.
- [6] J.L. Fleiss, *The Design and Analysis of Clinical Experiments*, Wiley, New York, 1986.
- [7] K.L. Hua, Y. Wang, *Applications of Number Theory to Numerical Analysis*, Springer, New York, 1981.
- [8] T.S. King, V.M. Chinchilli, Robust estimators of the concordance correlation coefficient, *J. Biopharm. Statist.* 31 (2001a) 83–105.
- [9] T.S. King, V.M. Chinchilli, A generalized concordance correlation coefficient for continuous and categorical data, *Statist. Med.* 20 (2001b) 2131–2147.
- [10] J. Lee, D. Koh, C.N. Ong, Statistical evaluation of agreement between two methods for measuring a quantitative variable, *Comput. Bio. Med.* 19 (1989) 61–70.
- [11] R. Li, M. Chow, Evaluation of reproducibility when the data are curves, Technical Report 01-07, Department of Statistics, The Pennsylvania State University, 2001, available at [http://www.stat.psu.edu/~rli/research/reprod\\_tech.pdf](http://www.stat.psu.edu/~rli/research/reprod_tech.pdf).
- [12] L.I. Lin, A concordance correlation coefficient to evaluate reproducibility, *Biometrics* 45 (1989) 255–268.
- [13] L. Lin, A.S. Hedayat, B. Sinha, M. Yang, Statistical methods in assessing agreement: models, issues and tools, *J. Amer. Statist. Assoc.* 97 (2002) 257–270.
- [14] H. Quan, W.J. Shih, Assessing reproducibility by the within-subject coefficient of variation with random effects models, *Biometrics* 52 (1996) 1195–1203.
- [15] J.O. Ramsey, B.W. Silverman, *Functional Data Analysis*, Springer, New York, 1997.
- [16] S.J. Sheather, M.C. Jones, A reliable data-based bandwidth selection method for kernel density estimation, *J. Royal Statist. Soc. B* 53 (1991) 683–690.
- [17] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman Hall, London, 1986.
- [18] E.F. Vonesh, V.M. Chinchilli, K. Pu, Goodness-of-fit in generalized nonlinear mixed-effects models, *Biometrics* 52 (1996) 572–587.