Statistica Sinica Preprint No: SS-2025-0404	
Title	Kernel-based Method for Detecting Structural Break in Distribution of Functional Data
Manuscript ID	SS-2025-0404
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0404
Complete List of Authors	Peijun Sang and Bing Li
Corresponding Authors	Peijun Sang
E-mails	peijun.sang@uwaterloo.ca
Notice: Accepted author version.	

Kernel-based Method for Detecting Structural Break in Distribution of Functional Data

Peijun Sang[1,*] and Bing Li [2]

1. University of Waterloo; 2. Pennsylvania State University

Abstract:

We propose a novel method to detect and date structural breaks in the entire distribution of functional data. Theoretical guarantees are developed for our procedure under fewer assumptions than in the existing work. In particular, we establish the asymptotic null distribution of the test statistic, which enables us to test the null hypothesis at a certain significance level. Additionally, the limiting distribution of the estimated structural break date is developed under two situations of the break size: fixed and shrinking towards 0 at a specified rate. We further propose a unified bootstrap procedure to construct a confidence interval for the true structural break date for these two situations. These theoretical results are justified through comprehensive simulation studies in finite samples. We apply the proposed method to Australian temperature data for detecting structural beaks.

Key words and phrases:

Change point analysis; Functional principal component analysis; Mean embedding; Reproducing kernel Hilbert space; Temperature data

1 Introduction

In this article we consider structural break detection for functional data. In functional data analysis (FDA), there is a large literature on functional principal component analysis (FPCA) and functional regression models. Recent developments in these two main topics can be found in comprehensive monographs such as Ramsay and Silverman (2005) and and Hsing and Eubank (2015). Structural break detection has been extensively studied in multivariate data. However, in FDA, detecting the structural break in distribution has received less attention compared with the two main topics.

Early work on structural break detection for functional data mainly deals with random samples of independent functional data that can be written as a mean function plus identically and independently distributed errors. Without loss of generality, the random samples are assumed to be elements of $L^2[0,1]$, and are generated as

$$X_i(t) = \mu(t) + \mathbb{1}(i \geq k^*)\delta(t) + \varepsilon_i(t), \quad i = 1, \ldots, n. \tag{1.1}$$

Here $k^* \in \{1, \ldots, n\}$ denotes the break date, $\delta(t)$ denotes the difference in the mean functions of X_i before and after k^*, and ε_i's are assumed to be i.i.d random errors with mean 0. Berkes et al. (2009) developed a procedure for testing the null hypothesis of no structural breaks against the alternative of a single break in the mean function under model (1.1). They established the consistency of the estimator of k^*/n under the alternative, but did not develop a method to quantify the uncertainty of the estimator. Aue et al. (2009) filled this gap by establishing the limiting distribution of the estimator, and Aston and Kirch (2012a) and Aston and Kirch (2012b) relaxed the independence assumption of ε_i's in model (1.1), and developed a similar procedure to detect and estimate k^* in the presence of certain serial correlation in ε_i's.

It should be noted that the aforementioned procedures rely on dimension reduction techniques such as FPCA. On the one hand, the use of dimension reduction techniques converts this infinite-dimensional structural break detection problem to a finite-dimensional one. On the other hand, dimension reduction would result in a loss of information. In particular, as noted in Aue et al. (2018), when there exists a structural break and δ in model (1.1) lies outside the space spanned by the retained functional principal components, a consistent FPCA-based test or break date estimator would be unavailable. To address this issue, Aue et al. (2018) developed a *fully functional* tool to detect the structural break based on the L^2-norm of the functional cumulative sum (CUSUM) statistic. Moreover, they investigated the limiting distribution of the estimated break date and developed confidence intervals for k^* based on the limiting distribution under the regime of a shrinking break size. Li et al. (2024) proposed a procedure for detecting structural breaks in the mean function of high-dimensional functional time series, which are allowed to be cross-sectionally correlated over subjects and temporally dependent over time.

Procedures for detecting structural breaks in the second moment of functional data have also been developed in the literature. In these procedures, functional data are still assumed to be generated from model (1.1) with a structural break in the covariance function of ε_i's but with $\delta(t) = 0$. For example, Aue et al. (2020) considered the structural break of the trace of the covariance operator. More general structural breaks of the covariance operator, e.g., structural breaks in the eigenvalues and/or eigenfunctions of the covariance operator, were studied in Horváth et al. (2022) and Dette and Kutta (2021). A summary of these methods for detecting the structural break in the mean or covariance function of functional data can be found in the monograph Horváth and Rice (2024).

However, structural breaks need not occur only through the means or the covariance operators of the functional data—indeed, it is easy to construct a series of random functions with no structural breaks in either their means or their covariance operators, but with a structural break in their distributions. In this paper, we develop a novel procedure to detect the structural break in the underlying distributions of functional data, which has broader applications and is more powerful than the existing methods in detecting the structural break of functional data. This procedure is built upon the concept of *mean embedding* (Muandet et al., 2017) and the structure of nested Hilbert spaces. Roughly speaking, we assume that the stochastic process, denoted as X, lies in a general Hilbert space, and then use a positive kernel κ to construct a reproducing kernel Hilbert space (RKHS) defined on the first-level Hilbert space. Under mild conditions, the function $\mathrm{E}\left[\kappa(X,\cdot)\right]$, which is known as the mean embedding of X, can characterize the entire distribution of X. We make use of this fact to develop a procedure to detect the structural break in the distributions of a sequence of independently distributed random functions. In particular, we propose a specific functional CUSUM statistic to detect and estimate the structural break date in the spirit of fully functional approach in Aue et al. (2018).

In addition to developing this procedure, we make three more important contributions. First, in contrast to the procedures mentioned above for detecting the structural break in the mean or covariance function, our framework does not require specific assumptions on the distribution of the sequence of stochastic processes as in (1.1). Second, we develop appealing theoretical properties for our procedure. More specifically, under mild conditions, we establish the limiting distribution of the proposed test statistic under the null hypothesis that there is no structural break. This allows us to detect the existence of a structural break in the distribution with theoretical guarantees. Additionally, we construct a unified bootstrap confidence interval for the true structural break date. Third, we conduct extensive simulation studies to demonstrate the performance of our procedure compared with some alternative procedures. In particular, these studies show that the size and power of the proposed test statistic compare favorably with these alternatives under various simulation settings. Additionally, the true coverage probability of the bootstrap confidence interval is close to the nominal level. In contrast, the confidence interval developed in Aue et al. (2018) is excessively conservative.

Though the MMD-based statistic has been employed in change point detection in multivariate data, our methodological contributions beyond the existing literature on MMD-based change-point detection for multivariate data are summarized as follows. First, unlike multivariate data, in the functional setting we do not observe the data fully, but can only make discretized observations that may be contaminated with random noise rather than the entire trajectory for functional data. Therefore, to implement our method, we first need to recover the trajectory from noisy and irregularly spaced observations for each subject. Second, unlike the functional data setting, change-point detection methods for multivariate data may suffer from the curse of dimensionality. Specifically, the performance of multivariate methods often deteriorates as the dimension of the data increases. In contrast, dimensionality is typically a blessing rather than a curse for functional data. As the observation grid becomes denser, the underlying smoothness of functional trajectories can be exploited more effectively, leading to improved estimation and inference. This distinction was identified and discussed in Section 6 of Wynne and Duncan (2022). Although our paper does not investigate this issue, as our primary focus is on detecting distributional changes, we believe that the same distinction applies in our setting and warrants further investigation in future research. Third, the limiting behavior of the MMD of two distributions defined on stochastic processes is different from that defined on multivariate data. See Proposition 2 of Wynne and Duncan (2022) for more details. In the context of change point detection, we studied the asymptotic behavior of the estimated break date under two situations: the fixed break size situation where the break size δ is independent of n, and the shrinking break size situation where $\delta = \delta_n$ converges to 0 at a specified rate. To the best of our knowledge, the second situation has not been studied yet when applying MMD to change-point detection for multivariate data. Fourth, as pointed out by Aue et al. (2018), a direct application of the limiting distribution at the shrinking break size situation leads to an excessively conservative confidence interval for dating the break in the mean function. This motivates us to devise a bootstrap procedure that does not depend on this limiting distribution to construct a valid and useful confidence interval.

Throughout the paper, $\xrightarrow{d}$ denotes weak convergence of random variables taking values in $\mathbb{R}^p$ for $p \in \mathbb{N}$, while $\xrightarrow{\mathcal{L}}$ denotes weak convergence of random functions taking values in an infinite-dimensional metric space. For $i < j$, we use the notation $a_{i:j}$ to denote the vector $(a_i, \ldots, a_j)$. For a Hilbert-Schmidt operator A, $\|A\|_{\text{HS}}$ denotes its Hilbert-Schmidt norm. For any $x \in \mathbb{R}$, let $\lfloor x \rfloor$ be the largest integer that is smaller than or equal to x.

The remainder of the article is organized as follows. In Section 2, we introduce relevant mathematical concepts for our method. In Section 3 we develop the procedure to detect the structural break and estimate the break date. Theoretical properties of the proposed procedure are further investigated in Section 3. Section 4 provides additional details to implement our procedure. We carry out extensive simulation studies to demonstrate the performance of our procedure in Section 5. In Section 6 we apply our procedure to one real-world example. Section 7 concludes the article. All theoretical proofs, some additional details for numerical implementations and extra empirical results are relegated to the Supplementary Material.

2 Problem Setup

In this section, we introduce the structure of nested Hilbert spaces and the concept of mean embedding, which play important roles in our method development.

2.1 Nested Hilbert spaces

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathbb{I}$ a bounded and closed interval in $\mathbb{R}$, and $\mathcal{H}$ a separable Hilbert space of functions on $\mathbb{I}$. Without loss of generality, we assume that $\mathbb{I} = [0, 1]$. Let $X : \Omega \to \mathcal{H}$ be a random element taking values in $\mathcal{H}$ measurable with respect to $\mathcal{F}/\mathcal{F}_X$, where $\mathcal{F}_X$ denotes the Borel σ-algebra generated by the open sets in $\mathcal{H}$. Let P_X denote the distributions of X.

Let $\kappa : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ be a positive kernel and $\mathfrak{M}_X$ be the RKHS generated by κ. We assume that κ is induced by the inner product in $\mathcal{H}$; that is, there exists a function $\rho : \mathbb{R}^3 \to \mathbb{R}^+$, such that for any $f, g \in \mathcal{H}$, $\kappa(f, g) = \rho(\langle f, f\rangle_{\mathcal{H}}, \langle f, g\rangle_{\mathcal{H}}, \langle g, g\rangle_{\mathcal{H}})$. A typical example is the Gaussian radial basis (GRB) kernel

$$\kappa(f, g) = \exp(-\gamma \|f - g\|^2_{\mathcal{H}}), \tag{2.2}$$

where $\gamma > 0$ is a tuning constant. This is an extension of the GRB function with the Euclidean norm replaced by the $\mathcal{H}$-norm. Since the kernel of $\mathfrak{M}_X$ is determined by the inner product of $\mathcal{H}$, we refer to $\mathfrak{M}_X$ as the nested RKHS via ρ. Similar structures were adopted in Li and Song (2017) for nonlinear sufficient dimension for functional data and Sang and Li (2026) for nonlinear function-on-function regression.

2.2 Maximum Mean Discrepancy

Let $\mathcal{P}$ denote the set of Borel probability measures defined on $\mathcal{F}_X$ and $\mathcal{P}_\kappa \subset \mathcal{P}$ be the set of probability measures P such that $\mathrm{E}_{X\sim P}[\kappa(X, X)] < \infty$. We characterize the distance between two probability measures in $\mathcal{P}_\kappa$ via mean embedding (Muandet et al., 2017), which is the mapping $P_\kappa \to \mathfrak{M}_X : \; P_X \mapsto \int_{\mathcal{H}} \kappa(\cdot, x) dP_X(x)$. We denote the right-hand side by μ_X. By the Cauchy-Schwarz inequality, it is straightforward to show $\|\mu_X\|_{\mathfrak{M}_X} < \infty$, which means μ_X is a member of $\mathfrak{M}_X$.

Let $P_Y \in \mathcal{P}_\kappa$ denote the distribution of another random element Y in $\mathcal{H}$.

Suppose that $\{X_1, \dots, X_n\}$ and $\{Y_1, \dots, Y_m\}$ are i.i.d samples from P_X and P_Y, respectively. The *maximum mean discrepancy* (MMD) was introduced by Gretton et al. (2012) to test $P_X = P_Y$. In general, assume that U and V are random variables taking values in a same topological space $\mathcal{S}$ and P and Q denote their distribution, respectively. Let $\mathrm{E}_U\{f(u)\} := \mathrm{E}_{U\sim P}\{f(u)\}$ and $\mathrm{E}_V\{f(v)\} := \mathrm{E}_{V\sim Q}\{f(v)\}$. For a class of functions $f : \mathcal{S} \to \mathbb{R}$, denoted by $\mathcal{F}$, the MMD is defined as

$$\mathrm{MMD}[\mathcal{F}, P, Q] = \sup_{f\in\mathcal{F}}[\mathrm{E}_U\{f(u)\} - \mathrm{E}_V\{f(v)\}]. \tag{2.3}$$

Gretton et al. (2012) suggested taking $\mathcal{F}$ as the unit ball of an RKHS.

Based on (2.3), we have $\mathrm{MMD}^2[\mathcal{F}, P_X, P_Y] = \|\mu_X - \mu_Y\|^2_{\mathfrak{M}_X}$, by Lemma 4 of Gretton et al. (2012), where $\mathcal{F}$ is the unit ball of $\mathfrak{M}_X$. If the mean embedding under the reproducing kernel κ is injective, MMD is indeed a metric over $\mathcal{P}_\kappa$. According to Sriperumbudur et al. (2011), this amounts to saying κ is *characteristic* of $\mathcal{P}_\kappa$; namely, two probability measures $P_X, P_Y \in \mathcal{P}_\kappa$ satisfy that $P_X = P_Y$ if and only if $\mu_X = \mu_Y$. By Theorem 3 of Wynne and Duncan (2022), the GRB kernel defined in (2.2) is characteristic of $\mathcal{P}_\kappa$. Interested readers may refer to Fukumizu et al. (2007) and Sriperumbudur et al. (2011) for more details about characteristic kernels.

Furthermore, it can be easily shown that MMD^2 can be rewritten as

$$\mathrm{MMD}^2[\mathcal{F}, P_X, P_Y] = \mathrm{E}\{\kappa(X, X')\} - 2\mathrm{E}\{\kappa(X, Y)\} + \mathrm{E}\{\kappa(Y, Y')\},$$

where X' and Y' are identical and independent copies of X and Y, respectively. According to Wynne and Duncan (2022), when $n = m$, an empirical version of $\mathrm{MMD}^2[\mathcal{F}, P_X, P_Y]$ is given by a U-statistic:

$$\widehat{\mathrm{MMD}}^2(X_{1:n}, Y_{1:n}) = \frac{1}{n(n-1)} \sum_{i \neq j}^{n} h(Z_i, Z_j), \tag{2.4}$$

where $Z_i = (X_i, Y_i)$ and $h(Z_i, Z_j) = \kappa(X_i, X_j) + \kappa(Y_i, Y_j) - \kappa(X_i, Y_j) - \kappa(X_j, Y_i)$.

We assume that $X_1, \ldots, X_n$ are independently distributed random functions taking value in $\mathcal{H}$. The problem of primary interest in this article is to leverage MMD^2 to test the null hypothesis that there's no structural break in the distribution of this sequence. Moreover, if this null hypothesis is rejected, we aim to develop a method to date this structural break.

2.3 Covariance operator

Assumption 1. The n random functions $X_1, \ldots, X_n$ are independently and identically distributed with a random function X.

Assumption 2. $\mathrm{E}\left[\kappa(X, X)\right] < \infty$.

Obviously, Assumption 2 is met when κ is the GRB kernel. Let $L^2(P_X)$ denote the class of all measurable functions of X such that $\mathrm{E}\left\{f^2(X)\right\} < \infty$ under P_X. Assumptions 1 and 2 ensure that the inclusion mapping $\mathfrak{M}_X \to L^2(P_X)$, $f \mapsto f$ is a bounded linear operator. Indeed, since $f(X_i) = \langle f, \kappa(X_i, \cdot)\rangle_{\mathfrak{M}_X}$ for any $f \in \mathfrak{M}_X$, the Cauchy-Schwarz inequality leads to $\mathrm{E}\left\{f^2(X_i)\right\} \leq \|f\|^2_{\mathfrak{M}_X} \mathrm{E}\left[\kappa(X_i, X_i)\right]$. Furthermore, by the same argument, the bilinear form $\mathfrak{M}_X \times \mathfrak{M}_X \to \mathbb{R}$, $(f, g) \mapsto \mathrm{cov}(f(X), g(X))$ is bounded. Therefore under Assumptions 1 and 2, there exists a linear operator $\Sigma_{XX} \in \mathcal{B}(\mathfrak{M}_X)$, the collection of bounded and linear operators on $\mathfrak{M}_X$, such that $\langle f, \Sigma_{XX} g\rangle_{\mathfrak{M}_X} = \mathrm{cov}(f(X), g(X))$.

Under Assumptions 1 and 2, we can easily show that Σ_{XX} is a trace-class operator. Then there exits a non-increasing and summable sequence of non-negative eigenvalues $\{\theta_\nu\}$ and corresponding eigenfunctions $\{\varphi_\nu\} \subset \mathfrak{M}_X$ such that $\Sigma_{XX} = \sum_{\nu=1}^{\infty} \theta_\nu (\varphi_\nu \otimes \varphi_\nu)$, where $\langle \varphi_\mu, \varphi_\nu\rangle_{\mathfrak{M}_X} = \delta_{\mu\nu}$, and $\delta_{\mu\nu}$ is the Kronecker δ function which equals 1 when $\mu = \nu$ and 0 otherwise. Further, $\kappa(X, \cdot)$ admits the following representation:

$$\kappa(X, \cdot) = \mu_X + \sum_{\nu=1}^{\infty} \theta_\nu^{1/2} \xi_\nu \varphi_\nu(\cdot), \tag{2.5}$$

where $\xi_\nu = \theta_\nu^{-1/2}\langle \kappa(X,\cdot) - \mu_X, \varphi_\nu \rangle_{\mathfrak{M}_X}$, are mean zero and uncorrelated random variables.

3 Methodology and Theoretical Results

In this section, we develop the procedure to detect the structural break in the distributions of $X_1, \ldots, X_n$ based on the MMD defined in (2.3). If our proposed detection procedure suggests that there is a structural break, we further construct a confidence interval for the break date. Theoretical properties of the procedure are also established.

The basic idea of detecting a structural break is as follows. Suppose there is a structural break in the distribution of X_i's at k. Namely, the distribution of $X_{1:k}$ is different from that of $X_{(k+1):n}$. Based on the concept of MMD and the empirical version of MMD^2 in (2.4), we may take $\|\widehat{\mu}_{1:k} - \widehat{\mu}_{(k+1):n}\|^2_{\mathfrak{M}_X}$, where $\widehat{\mu}_{1:k} = k^{-1}\sum_{i=1}^k \kappa(X_i, \cdot)$ and $\widehat{\mu}_{(k+1):n} = (n-k)^{-1}\sum_{i=k+1}^n \kappa(X_i, \cdot)$ with a characteristic kernel κ, as the test statistic. But as pointed out by Berkes et al. (2009), statistics based on this type of summations tend to be inflated artificially when the break point k is close to 1 or n, which corresponds to early break and late break. Therefore, we consider a CUSUM-type statistic instead:

$$S_k = \left\| \sum_{i=1}^{k} \kappa(X_i, \cdot) - \frac{k}{n} \sum_{i=1}^{n} \kappa(X_i, \cdot) \right\|^2_{\mathfrak{M}_X}. \tag{3.6}$$

If there is indeed a structural break at k, the value of S_k tends to be relatively large. Then for any $\tau \in (0,1)$, setting $k = \lfloor n\tau \rfloor$, we obtain a distance metric

$$d_n(\tau) = \left\| \sum_{i=1}^{\lfloor n\tau \rfloor} \kappa(X_i, \cdot) - \frac{\lfloor n\tau \rfloor}{n} \sum_{i=1}^{n} \kappa(X_i, \cdot) \right\|^2_{\mathfrak{M}_X}.$$

Remark 1. Alternatively, instead of calculating $d_n(\tau)$ directly, we can project the difference of these two empirical means onto particular directions. In the context of detecting the structural break in the mean function of independently distributed functional data, Berkes et al. (2009) chose the first d FPCs of X_i's as the projection directions. In our context, we can perform kernel FPCA on X_i's to estimate φ_ν's defined in (2.5). However, as pointed out by Aue et al. (2018), the use of a fully functional test statistic has potential advantages over the projection-based method in detecting and dating the mean break of X_i's. Therefore, we adopt $d_n(\tau)$ for detecting the structural break in the distribution of X_i's.

Let $\{B_\nu(\cdot) : \nu \in \mathbb{N}\}$ be a sequence of independent standard Brownian bridges. The following theorem establishes weak convergence for the stochastic process $\{d_n(\tau) : 0 \leq \tau \leq 1\}$ under the null hypothesis, where there is no structural break in the distribution of these independent random functions.

Theorem 3.1. *Under Assumptions 1 and 2,*

$$\frac{d_n(\tau)}{n} \xrightarrow{\mathcal{L}} \sum_{\nu=1}^{\infty} \theta_\nu B_\nu^2(\tau) \qquad \tau \in [0,1],$$

in the Skorohod topology of $D[0,1]$, a collection of functions on $[0,1]$ that are right-continuous and have left limits.

By Theorem 3.1, we consider a Kolmogorov–Smirnov-type procedure for testing structural breaks. In particular, define

$$T_n = \sup_{\tau \in [0,1]} \frac{d_n(\tau)}{n}. \tag{3.7}$$

By the continuous mapping theorem, one has, under Assumptions 1 and 2, as $n \to \infty$,

$$T_n \xrightarrow{d} \sup_{\tau \in [0,1]} \sum_{\nu=1}^{\infty} \theta_\nu B_\nu^2(\tau).$$

Then we reject the null hypothesis at significance level α if $T_n > C_\alpha$, where C_α satisfies $\mathbb{P}(T_n > C_\alpha) = \alpha$. Implementing this test requires estimation of C_α, which will be discussed in Section S.2 of the Supplementary Material.

In particular, the estimation of C_α involves the kernel functional principal component analysis (Li and Song, 2017). Under the assumption that the trajectory of functional data is fully observed, the complexity of the this procedure is $O(n^3)$ if we ignore the complexity of evaluating the inner product in $\mathcal{H}$. To improve the computational efficiency, a possible approach is to adapt the idea of using random Fourier features to approximate $\kappa(x, y)$ (Rahimi and Recht, 2007), where x, y are multivariate inputs, to to the current setting functional data. This is left for future research.

Next we investigate the asymptotic behavior of T_n under the alternative hypothesis. To avoid obscuring the main idea by technical complexities due to multiple structure break dates, we will focus on the scenarios with a single structural break date. This motivates us to consider the following scenario under the alternative hypothesis.

Assumption 3. The n random functions $X_1, \ldots, X_n$ are independently distributed. For some $k^* = \lfloor n\tau^* \rfloor$ with $\tau^* \in (0, 1)$, the probability laws of $X_1, \ldots, X_{k^*}$ are identical and denoted as P_1, while the probability laws of $X_{(k^*+1)}, \ldots, X_n$ are identical and denoted as P_2, with $P_1 \neq P_2$. Further, $\mathrm{E}\left[\{\kappa(X_i, \cdot) - \mu_1\} \otimes \{\kappa(X_i, \cdot) - \mu_1\}\right] = \mathrm{E}\left[\{\kappa(X_j, \cdot) - \mu_2\} \otimes \{\kappa(X_j, \cdot) - \mu_2\}\right]$ for any $i \in 1 : k^*, j \in (k^* + 1) : n$, where μ_1 and μ_2 denote the mean embeddings of P_1 and P_2, respectively.

Theorem 3.2. *Under Assumptions 2 and 3, $T_n \xrightarrow{\mathbb{P}} \infty$ as $n \to \infty$.*

Theorem 3.2 states that under the alternative hypothesis specified in Assumption 3, T_n would diverge in probability. If the null hypothesis is rejected, define the estimator of the structural break date as $\hat{k}_n = \min\left\{k : S_k = \max_{1 \le k' \le n} S_{k'}\right\}$. Under Assumption 3, we still use Σ_{XX} to denote the common covariance operator of P_1 and P_2. Let $\epsilon_i = \kappa(X_i, \cdot) - \mu_1$ for $1 \le i \le k^*$ and $\epsilon_i = \kappa(X_i, \cdot) - \mu_2$ for $k^* < i \le n$. Obviously, the mean embedding of ϵ_i for $1 \le i \le n$ is 0 and its covariance operator is Σ_{XX} under Assumption 3. Let $\delta = \mu_2 - \mu_1$ denote the break size. As in Aue et al. (2018), we study the asymptotic behavior of $\hat{k}_n - k^*$ under two situations: the fixed break size situation where δ is independent of n, and the shrinking break size situation where $\delta = \delta_n$ converges to 0 at a specified rate. Let $\hat{\tau}_n = \hat{k}_n / n$. The following theorem establishes the consistency for $\hat{\tau}_n$, and the limiting distribution for the structural break date estimator under Assumption 3 with a fixed break size.

Theorem 3.3. *Under Assumptions 2 and 3 with $\delta \neq 0$, we have as $n \to \infty$,*

(i) $\hat{\tau}_n \xrightarrow{\mathbb{P}} \tau^*$;

(ii)

$$\hat{k}_n - k^* \xrightarrow{d} \min\left\{k : G(k) = \sup_{k' \in \mathbb{Z}} G(k')\right\}, \tag{3.8}$$

where

$$G(k) = \begin{cases} (1-\tau^*)\|\delta\|^2_{\mathfrak{M}_X} k + \langle \delta, V_k \rangle_{\mathfrak{M}_X}, & k < 0, \\ 0, & k = 0, \\ -\tau^*\|\delta\|^2_{\mathfrak{M}_X} k + \langle \delta, V_k \rangle_{\mathfrak{M}_X}, & k > 0, \end{cases}$$

with $V_k = \sum_{i=1}^{k} \epsilon_i$ *for* $k > 0$ *and* $\sum_{i=k}^{-1} \epsilon_i$ *for* $k < 0$, *where* ϵ_{-i} *denotes an independent copy of* ϵ_i *for* $i > 0$.

According to Theorem 3.3, the limiting distribution of $\hat{k}_n - k^*$ depends on ϵ_i's and δ when the break size δ is fixed. This fact renders great difficulty in constructing confidence intervals for the break date k^*. Inspired by Aue et al. (2018), we investigate the limiting distribution of $\hat{k}_n - k^*$ in the second situation where $\delta = \delta_n$ depends on the sample size and shrinks towards 0 at a specified rate. We anticipate a more tractable limiting distribution that can facilitate the construction of confidence intervals for k^*.

Theorem 3.4. *Suppose that Assumptions 2 and 3 hold with* $0 \neq \delta = \delta_n$, *where* δ_n *satisfies* $\|\delta_n\|_{\mathfrak{M}_X} \to 0$ *but* $n\|\delta_n\|^2_{\mathfrak{M}_X} \to 0$. *Then we have, as* $n \to \infty$,

(i) $\hat{\tau}_n \xrightarrow{\mathbb{P}} \tau^*$;

(ii)

$$\|\delta_n\|^2_{\mathfrak{M}_X}(\hat{k}_n - k^*) \xrightarrow{d} \inf\left\{x : U(x) = \sup_{x' \in \mathbb{R}} U(x')\right\},$$

where

$$U(x) = \begin{cases} (1-\tau^*)x + \sigma W(x), & x < 0, \\ -\tau^* x + \sigma W(x), & x \geq 0, \end{cases}$$

with $(W(x) : x \in \mathbb{R})$ *a two-sided Brownian motion, and*

$$\sigma^2 = \lim_{n\to\infty} \frac{\langle \Sigma_{XX}\delta_n, \delta_n \rangle_{\mathfrak{M}_X}}{\|\delta_n\|^2_{\mathfrak{M}_X}}.$$

Remark 2. When replacing $\kappa(X_i, \cdot)$ with X_i in (3.6) to detect the structural break in the mean function of a sequence of random functions, a similar result to Part (i) of Theorem 3.3 was established in Berkes et al. (2009), and a similar result to Part (ii) of Theorems 3.3 and 3.4 was established in Aue et al. (2009) and Aue et al. (2018). It should be noted that these reference works impose a specific assumption on the data generation process as in model (1.1), where ε_i' are assumed to be i.i.d copies of the error process (Berkes et al., 2009; Aue et al., 2009) or dependent innovations satisfying the L^p-m approximable condition specified in Assumption 1 of Aue et al. (2018). In model (1.1) we see that the only change in the distribution of X_i before and after the change point k^* is the mean $E(X_i)$, which is μ before the change point and $\mu + \delta$ afterwards. In contrast, our method is designed to detect the change in the distribution of the random functions. Our change point problem is formulated by $P_1 = \cdots = P_{k^*} \neq P_{k^*+1} = \cdots = P_n$, which is much more general. For example, if the change before and after k^* is the volatility of X_i rather than the mean, then it cannot be detected by the mean-change-point setting of Aue et al. (2018), but can be detected by our distribution-change-point setting. The advantage of our method over these alternatives will be further demonstrated through simulation studies.

We can construct a confidence interval for k^* in the situation of a shrinking break size based on part (ii) of Theorem 3.4. Let $\hat{\sigma}$ and $\hat{\delta}_n$ denote the estimates of the nuisance parameter σ and break size δ_n, respectively. For any $\alpha \in (0,1)$, we construct an asymptotic $(1-\alpha)$-sized confidence interval for k^* as

$$\left(\hat{k}_n - \frac{q_{1-\frac{\alpha}{2}}}{\|\hat{\delta}_n\|^2_{\mathfrak{M}_X}}, \hat{k}_n + \frac{q_{\frac{\alpha}{2}}}{\|\hat{\delta}_n\|^2_{\mathfrak{M}_X}}\right), \tag{3.9}$$

where q_α denotes the qth quantile of $\Lambda = \inf\{x : U(x) = \sup_{x' \in \mathbb{R}} U(x')\}$.

4 Sample-Level Implementations

4.1 Choice of tuning parameters in the characteristic kernal

We use the GRB kernel defined in (2.2) for our numerical implementations, which require the choice of the tuning parameter γ. We extend the widely adopted median heuristic (Garreau et al., 2017) to the Gaussian kernel defined on $\mathcal{H} \times \mathcal{H}$. In particular, we choose γ as

$$\frac{1}{\gamma} = \text{median}\{\|X_i - X_j\|_{\mathcal{H}}^2 : i, j = 1, \ldots, n, i \neq j\}.$$

Theoretical properties for this particular choice of γ were investigated in Wynne and Duncan (2022) for functional data taking values in $L^2(\mathbb{I})$, the space of square-integrable functions defined on $\mathbb{I}$. In fact, in the literature of FDA, $L^2(\mathbb{I})$ is widely adopted as the ambient space within which functional data are considered; see Wang et al. (2016) for a detailed overview. In our numerical implementations, we take $\mathcal{H} = L^2(\mathbb{I})$, but the proposed framework can be applied to other choices of $\mathcal{H}$.

4.2 Confidence interval for the structural break date under alternative hypothesis

In Remark 2, we develop an asymptotically $(1-\alpha)$-sized confidence interval (3.9) for k^* under Assumption 3 with a shrinking break size δ_n. This confidence interval requires accurate estimates of the nuisance parameters σ and δ_n.

When the null hypothesis is rejected, given the estimated structural break date $\hat{k}_n$, we take the sample mean as the estimate of the two mean embeddings: $\hat{\mu}_1 = 1/\hat{k}_n \sum_{i=1}^{\hat{k}_n} \kappa(X_i,)$ and $\hat{\mu}_2 = 1/(n - \hat{k}_n) \sum_{\hat{k}_n+1}^{n} \kappa(X_i,)$. Then $\hat{\delta}_n = \hat{\mu}_2 - \hat{\mu}_1$ is a natural estimate of δ_n. Under Assumption 3, we consider another pooled sample covariance for estimating Σ_{XX} (with a slight abuse of notation):

$$\hat{\Sigma}_{XX} = \frac{1}{n}\left[\sum_{i=1}^{\hat{k}_n}\{\kappa(X_i,) - \hat{\mu}_1\} \otimes \{\kappa(X_i,) - \hat{\mu}_1\} + \sum_{\hat{k}_n+1}^{n}\{\kappa(X_i,) - \hat{\mu}_2\} \otimes \{\kappa(X_i,) - \hat{\mu}_2\}\right].$$

Then an estimate of σ^2 is immediately available. With $\hat{\sigma}$ and $\hat{\tau}_n$, we can easily estimate q_α by generating a sufficiently large number of i.i.d sample paths of the process $\{U(x) : x \in \mathbb{R}\}$.

Numerical studies demonstrate that the confidence interval developed above is excessively wide. Similar issues were identified in the confidence interval for the structural break date in the mean function in Aue et al. (2018). A plausible reason is that this confidence interval is built upon the assumption that the break size δ_n shrinks towards 0 at a specified rate, which is seriously violated in numerical studies. To address this issue, we adopt the bootstrap-based method proposed by Antoch et al. (1995), who was the first to recognize the difference in the limiting behavior of $\hat{k}_n - k^*$ for independently distributed random variables with a mean change under these two situations: fixed δ and shrinking δ_n. Antoch et al. (1995) proposed a unified bootstrap method to construct confidence intervals for k^* for these two situations. In particular, after we obtain the estimated structural break date $\hat{k}_n$, define the estimated residuals:

$$\hat{\epsilon}_i = \begin{cases} \kappa(X_i, \cdot) - \hat{\mu}_1, & i = 1, \dots, \hat{k}_n, \\ \kappa(X_i, \cdot) - \hat{\mu}_2, & i = \hat{k}_n + 1, \dots, n. \end{cases}$$

Then we take $\epsilon_1^*, \dots, \epsilon_n^*$ independently from the empirical distribution of $\hat{\epsilon}_1, \dots, \hat{\epsilon}_n$, and consider the bootstrap samples.

$$Y_i^* = \begin{cases} \epsilon_i^* + \hat{\mu}_1, & i = 1, \dots, \hat{k}_n, \\ \epsilon_i^* + \hat{\mu}_2, & i = \hat{k}_n + 1, \dots, n. \end{cases}$$

Then we carry out the procedure in Section 3 to estimate the structural break date by replacing $\kappa(X_i, \cdot)$ with Y_i^* for $i = 1, \dots, n$. We repeat this entire procedure B times, and employ the sample quanitles of these B estimates of k^* to construct confidence intervals for k^*.

Regarding the choice of the reproducing kernel κ, our framework requires κ to be characteristic. In addition, using a bounded kernel can improve robustness. Besides the GRB kernel used throughout the paper, the Laplace kernel meets both requirements.

5 Simulation Studies

In this section, we perform extensive simulation studies to demonstrate the finite-ample performance of our proposed method in terms of detecting and dating structural breaks in the distribution of functional data. We compare our method with two alternative methods: the fully functional method proposed by Aue et al. (2018) for detecting and dating structural breaks in the mean function (FF-mean for short), and the fully functional method proposed by Horváth et al. (2022) for detecting and dating structural breaks in the covariance function (FF-cov for short). While the two alternative methods focus on detecting the changes in mean and covariance function of the functional data, our method targets the entire distribution of the functional data. Our method is referred to as FF-dist for short.

5.1 Simulation designs

Following the data-generation procedure in Aue et al. (2018), we generate functional data as

$$X_i(t) = m_i(t) + \sum_{j=1}^{21} \sqrt{\lambda_{ij}} \zeta_{ij} \phi_j(t), \quad i = 1, \ldots, n,$$

where $\phi_1, \ldots, \phi_{21}$ denote the 21 Fourier basis functions on [0, 1] and $\zeta_{i1}, \ldots, \zeta_{i,21}$ denote the independent FPC scores. Each X_i is observed at 50 equally-spaced time points on [0, 1]. Let $k^* = \lfloor n\tau^* \rfloor$ denote the true structure break date. We consider the following three settings.

(a) Setting 1: $\lambda_{ij} = j^{-1}$ for $j = 1, \ldots, 21$ and $i = 1, \ldots, n$; $m_i(t) = 0$ for $i \leq k^*$, while $m_i(t) = a \cdot \sum_{j=1}^{21} \phi_j(t)/\sqrt{21}$ for $i > k^*$. Following Aue et al. (2018), different values of a are taken such that the signal-to-noise ratio, which is defined as $\text{SNR} = a\tau^*(1-\tau^*)/\sum_{j=1}^{21} \lambda_j$, varies among 0, 0.1, 0.2, 0.3, 0.5, 1 and 1.5. Again, ζ_{ij}'s are independently generated from $N(0, 1)$.

(b) Setting 2: $m_i(t) = 0$ for $i = 1, \ldots, n$; $\lambda_{ij} = j^{-1}$ for $j = 1, \ldots, 21$ and $i \leq k^*$ while $\lambda_{ij} = c \cdot j^{-1}$ for $j = 1, \ldots, 21$ and $i > k^*$. We take $c = 1, 1.6, 1.8, 2, 2.5$ and 3 to accommodate slight to notable changes in the covariance operator after the structural break date. Moreover, ζ_{ij}'s are independently generated from $N(0, 1)$.

(c) Setting 3: $m_i(t) = 0$ and $\lambda_{ij} = j^{-1}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, 21$. Moreover, ζ_{ij}'s are independently generated from $N(0, 1)$ for $i \leq k^*$, while for $i > k^*$, $c_\pi \zeta_{ij}$'s are independently generated from a mixture model of the standard normal distribution and the student t distribution with 3 degrees of freedom, i.e., $c_\pi \zeta_{ij} \sim \pi N(0,1) + (1-\pi)t_3$, where π denotes the mixing proportion and c_π is chosen such that $var(\zeta_{ij}) = 1$. We take $\pi = 0, 0.2, 0.4, 0.6, 0.8$ and 1 to accommodate slight to notable changes in the distribution of X after the structural break time.

Setting 1 is exactly the same as setting 3 in Aue et al. (2018). Settings 1 and 2 are designed to showcase the performance of our method in detecting and dating the break time when there exists a structural break in the mean function and the covariance function of independently distributed functional data, respectively. In contrast to the previous two settings, there is no change in the mean function or covariance function of X_i's in Setting 3. To accommodate a structural break in the underlying distribution of X_i's, we consider a distributional change in the functional principal component (FPC) scores of X_i's. In particular, we generate FPC scores from a mixture model. Figure S1 in the Supplementary Material displays the profiles of 10 randomly selected subjects before and after structural breaks for the three settings. We identify a change in the mean function and the covariance function in the first two panels of Figure S1. However, the difference between these groups of functional data is not discernible even with $\pi = 1$ in Setting 3 from Figure S1. These findings indicate the escalated difficulty in detecting and dating the structural break in Setting 3 compared to the first two settings.

To evaluate the performance of our method, we consider $\tau^* = 0$ (null hypothesis), and $\tau^* = 0.25$ and 0.5, and sample size $n = 100$ and 200 in the first two settings, and $n = 500$ and 800 in Setting 3 because of the escalated difficulty. 1000 independent simulation trials are carried out for each simulation scenario to calculate relevant metrics.

Setting	n	Methods		
		FF-mean	FF-cov	FF-dist
1	100	0.084		0.058
	200	0.076		0.070
2	100		0.060	0.049
	200		0.058	0.051
3	500	0.039	0.030	0.049
	800	0.041	0.032	0.050

Table 1: Empirical sizes of the detection methods under the three simulation settings with the nominal level $\alpha = 0.05$.

To demonstrate the sensitivity of our method to the choice of the tuning parameter γ and the kernel, we further carry out simulation studies with other choices of γ and kernels. It turns out the choice of the tuning parameter and kernel plays a critical role in the performance of our method. More details can be found in Section S.3.1 of the Supplementary Material.

5.2 Simulation results

Table 1 summarizes the type-I errors of the three methods under the three simulations settings under various sample sizes. As expected, the empirical sizes of our method and its competitors are close to the nominal level for under all settings. These findings indicate that our method is not inferior to its competitors in terms of controlling type I errors.

Figure 1 displays the power curves of the three detection methods under the three simulation settings for different sample sizes with $\tau^* = 0.5$. Somewhat surprisingly, though FF-mean and FF-cov were designed to detect the structural break in the mean function and the covariance function, respectively, our method shows a greater power of rejecting the null hypothesis in the first two settings. The advantage of our method is more evident when the break size in the mean function or the covariance function is moderate. All these findings demonstrate the effectiveness of our method in detecting the structural break in the first two moments of independently distributed functional data. More noticeable advantages of our method over the two competitors are found in the two right panels of Figure 1, which showcase the power curves of these three methods in Setting 3. Since there is no structural break in either the mean function or the covariance function in Setting 3, FF-mean and FF-cov cannot reject the null hypothesis. However, our proposed method can effectively detect the structural break in the the underlying distribution of functional data, especially when the mixing proportion of the t_3 distribution of the FPC scores becomes sufficiently large. To investigate whether these results are affected by the location of the structural break, we also consider $\tau^* = 0.25$ and the corresponding power curves are shown in Figure S2 in the Supplementary Material. Similar patterns in the comparison of the power curves for these detection methods are identified in Figure S2.

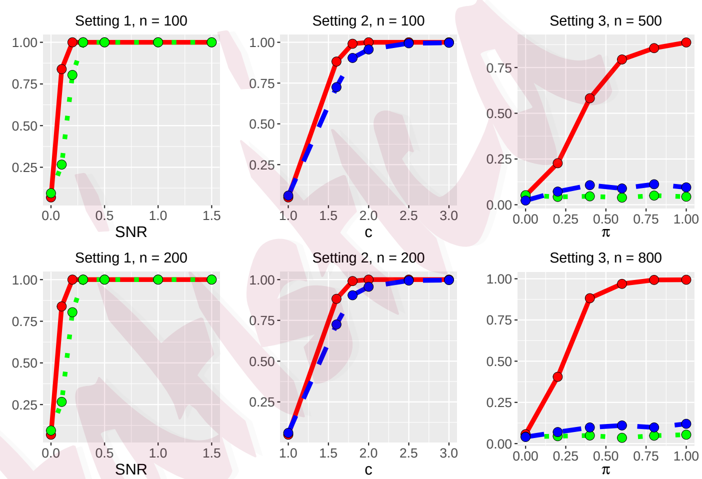

Figure 1: Power curves of the three detection methods under the three simulation settings with $\tau^* = 0.50$ for various sample sizes. In these panels, the red solid, green dotted and blue dashed lines represent the power curves of FF-dist, FF-mean and FF-cov, respectively.

Figure S4 in the Supplementary Material depicts the boxplots of the estimated structural break dates obtained from these three methods across 1000 Monte Carlo simulations under these three settings for $\tau^* = 0.5$. In Settings 1 and 2, FF-dist and its competitors can date the true structural break time reasonably well. But our proposed method is more stable regardless of sample size or the break size of the mean function or the covariance function. Given that FF-mean and FF-cov were proposed to focus on detecting the structural break in the mean function and the covariance function, respectively, our method provides a unified and powerful framework to detect and date the structural breaks in these population parameters of functional data. In Setting 3, which seems to be the most challenging scenario for dating the structural break as indicated in Figure S1, our method still shows remarkable performance in dating the structural break for relatively larger π and n. If we take a closer look at such boxplots for $\tau^* = 0.25$ in Figure S3 in the Supplementary Material, we find that, in contrast, FF-mean and FF-cov, especially the latter, cannot accurately date the structural break in this scenario.

Lastly, we examine the performance of the bootstrap confidence intervals developed in Section 4.2. Figure 2 presents the boxplots of the lengths of the 95% confidence intervals of τ^* for the three settings across 1000 simulation runs, where $\tau^* = 0.5$ and we generate $B = 500$ bootstrap resamples

Setting	n					
				SNR		
1		0.1	0.2	0.3	0.5	1
	100	0.998	0.976	0.974	0.958	0.974
	200	0.984	0.978	0.975	0.966	0.983
				c		
2		1.6	1.8	2	2.5	3
	100	0.974	0.961	0.957	0.947	0.948
	200	0.949	0.955	0.939	0.948	0.961
				π		
3		0.2	0.4	0.6	0.8	1
	500	0.982	0.963	0.942	0.931	0.940
	800	0.966	0.929	0.919	0.928	0.921

Table 2: The empirical coverage probabilities of the 95% bootstrap confidence intervals for τ^* under the three simulation settings with $\tau^* = 0.50$ based on 1000 Monte Carlo simulations.

in each simulation run.

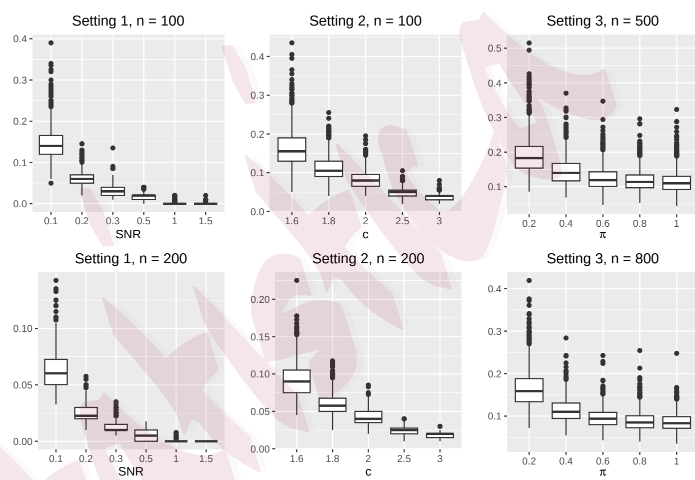

Figure 2: Boxplots of the lengths of the 95% bootstrap confidence intervals for τ^* under the three simulation settings with $\tau^* = 0.50$ across 1000 Monte Carlo simulations.

In Figure 2 we find that the confidence interval becomes shorter when the structural break size becomes larger or more samples are available in these three settings. Similar patterns are found in Figure S5 in the Supplementary Material, which presents the boxplots of the lengths of those confidence intervals with $\tau^* = 0.25$. We further assess the true coverage probability of the bootstrap confidence interval for τ^*. Table 2 summarizes the empirical coverage probabilities of these 95% bootstrap confidence intervals for $\tau^* = 0.5$ based on 1000 simulation runs for the three simulation settings. When the sample size or the structural break size is reasonably large, the empirical coverage probability is close to the nominal level.

6 Real Data Analysis

In this section, we apply our proposed method and the two alternative methods to detect and date the structural break in the profiles of the temperature data collected in Australia.

We study the temperature data collected at eight weather stations in Australia. At each site, the raw data consist of 365 daily observations of minimum temperatures over a certain time span, which may vary from site to site. The sites and the corresponding time spans during which the data were collected are shown in the first two columns of Table S8. The raw data are available from the Australian Bureau of Meteorology at `www.bom.gov.au`. Both Aue et al. (2018) and Chen et al. (2023) have considered structural breaks for this dataset from the perspective of the FDA.

Station	Time Span	$\hat{k}_{\text{FF-mean}}$	$\hat{k}_{\text{FF-dist}}$
Boulia Airport	1888-2012	1978 (1960, 1981)	1978 (1967, 1979)
Cape Otway Lighthouse	1864-2012	1970 (1920, 1976)	1970 (1946, 1972)
Gayndah Post Office	1893-2009	1962 (1956, 1964)	1962 (1959, 1963)
Gunnedah Pool	1876-2011	1965 (1929, 1974)	1965 (1958, 1966)
Hobart (Ellerslie Road)	1882-2012	1965 (1959, 1967)	1967 (1963, 1968)
Melbourne (Regional Office)	1855-2012	1958 (1955, 1959)	1958 (1956, 1958)
Robe Comparison	1884-2012	1960 (1943, 1967)	1972 (1957, 1974)
Sydney (Observatory Hill)	1859-2012	1957 (1952, 1959)	1957 (1954, 1958)

Table 3: Summary of the estimated structural breaks of the temperature curves and the corresponding 95% confidence intervals based on FF-mean and FF-dist at these eight sites.

As in Aue et al. (2018) and Chen et al. (2023), the raw observations collected in each year at each station are smoothed with 21 Fourier basis functions. Meanwhile, we remove years at which the proportion of missing values exceeds 1/3, from our study. We only compare the performance of our procedure with that of FF-mean developed in Aue et al. (2018), but do not consider the procedure proposed in Chen et al. (2023), since the latter focuses on detecting multiple structural breaks based on segmentation. Both our detection procedure and FF-mean yield tiny p-values at each of these 8 stations. These findings suggest that there exists a structural break in the mean function of the temperature profiles.

We further employ FF-mean and FF-dist to estimate the structural break in the mean function or the distribution of the temperature curves at each of these eight stations. Table 3 summarizes the estimated structural breaks and the associated 95% confidence intervals obtained from these two procedures. Except the stations Hobart and Robe Comparison, these two procedures give rise to identical estimated structural breaks, while our procedure always yields short confidence intervals, which echo those from the simulation studies.

Now we take a closer look at the estimated structure breaks obtained from these two procedures at Robe Comparison. Figure S8 in the Supplementary Material presents the profiles and mean functions of minimum temperatures after and before 1960, which is the structural break year obtained from FF-dist. The mean function of minimum temperatures after 1960 is slightly larger than that before 1960. To evaluate the brak size in the mean function compared with that in the distribution, we further calculate the functional CUSUM statistics of these two procedures, as shown in Figure S9 in the Supplementary Material. It turns out that the confidence interval obtained from our procedure, consists of years that have relatively large functional CUSUM statistics defined in (3.6). In contrast, the confidence interval from FF-mean contains years that have relatively small Z_k's, the functional CUSUM statistics defined in Aue et al. (2018). Lastly, it should be noted that our confidence interval contains 1960, but the confidence interval from FF-mean does not contain 1972, which is the estimated structural break year from our procedure. These demonstrate the effectiveness and robustness of our procedure.

7 Conclusion

In this article, we develop a fully functional method to detect and date structural breaks in the underlying distribution for independently observed functional data based on the mean embedding of functional data. Compared with existing work on structural breaks for functional data, the assumptions for our method are considerably weaker. In particular, our method does not rely on the classical functional change-point model (1.1), which is usually required by existing methods. Moreover, we establish appealing theoretical properties of the proposed method under mild conditions. These theoretical guarantees greatly broaden the applicability of the proposed method. Through extensive simulation studies, we demonstrated that our method performs notably better than the existing methods when the structural break lies in the aspects of distribution of the functional data beyond the mean and the covariance functions. Furthermore, even in the cases where the structural break can be fully characterized by the mean and covariance functions, our method is not inferior to the existing methods that are designed for such cases.

References

Antoch, J., Hušková, M., and Veraverbeke, N. (1995). Change-point problem and bootstrap. *Journaltitle of Nonparametric Statistics*, 5(2):123–144.

Aston, J. A. and Kirch, C. (2012a). Detecting and estimating changes in dependent functional data. *Journal of Multivariate Analysis*, 109:204–220.

Aston, J. A. and Kirch, C. (2012b). Evaluating stationarity via change-point alternatives with applications to fMRI data. *Annals of Applied Statistics*, 6(4):1906–1948.

Aue, A., Gabrys, R., Horváth, L., and Kokoszka, P. (2009). Estimation of a change-point in the mean function of functional data. *Journal of Multivariate Analysis*, 100(10):2254–2269.

Aue, A., Rice, G., and Sönmez, O. (2018). Detecting and dating structural breaks in functional data without dimension reduction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(3):509–529.

Aue, A., Rice, G., and Sönmez, O. (2020). Structural break analysis for spectrum and trace of covariance operators. *Environmetrics*, 31(1):e2617.

Berkes, I., Gabrys, R., Horváth, L., and Kokoszka, P. (2009). Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(5):927–946.

Chen, Y.-T., Chiou, J.-M., and Huang, T.-M. (2023). Greedy segmentation for a functional data sequence. *Journal of the American Statistical Association*, 118(542):959–971.

Dette, H. and Kutta, T. (2021). Detecting structural breaks in eigensystems of functional time series. *Electronic Journal of Statistics*, 15(1):944–983.

Fukumizu, K., Gretton, A., Sun, X., and Schölkopf, B. (2007). Kernel measures of conditional dependence. *Advances in Neural Information Processing Systems*, 20:489–496.

Garreau, D., Jitkrittum, W., and Kanagawa, M. (2017). Large sample analysis of the median heuristic. *arXiv preprint arXiv:1707.07269*.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. (2012). A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773.

Horváth, L. and Rice, G. (2024). *Change Point Analysis for Time Series.* Springer, New York.

Horváth, L., Rice, G., and Zhao, Y. (2022). Change point analysis of covariance functions: A weighted cumulative sum approach. *Journal of Multivariate Analysis*, 189:104877.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators.* John Wiley, Chichester.

Li, B. and Song, J. (2017). Nonlinear sufficient dimension reduction for functional data. *The Annals of Statistics*, 45(3):1059–1095.

Li, D., Li, R., and Shang, H. L. (2024). Detection and estimation of structural breaks in high-dimensional functional time series. *The Annals of Statistics*, 52(4):1716–1740.

Muandet, K., Fukumizu, K., Sriperumbudur, B., and Schölkopf, B. (2017). Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends® in Machine Learning*, 10(1-2):1–141.

Rahimi, A. and Recht, B. (2007). Random features for large-scale kernel machines. *Advances in Neural Information Processing Systems*, 20.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis 2nd edition.* Springer, New York.

Sang, P. and Li, B. (2026). Nonlinear function-on-function regression by RKHS. *Journal of Machine Learning Research*, 27(12):1–54.

Sriperumbudur, B. K., Fukumizu, K., and Lanckriet, G. R. (2011). Universality, characteristic kernels and RKHS embedding of measures. *Journal of Machine Learning Research*, 12(7):2389–2410.

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, 3(1):257–295.

Wynne, G. and Duncan, A. B. (2022). A kernel two-sample test for functional data. *Journal of Machine Learning Research*, 23(73):1–51.

Department of Statistics and Actuarial Science, University of Waterloo, Ontario, Canada

E-mail: peijun.sang@uwaterloo.ca

ORCID: 0000000253120204

Department of Statistics, Pennsylvania State University, Pennsylvania, USA

E-mail: bxl9@psu.edu

(Received ???? 20??; accepted ???? 20??)