ENHANCED STRUCTURAL BREAK DETECTION IN FUNCTIONAL MEANS

Shuhao Jiao[a], Ngai Hang Chan[a,b], Chun Yip Yau[b]

[a]*Department of Biostatistics, City University of Hong Kong*

[b]*Department of Statistics, the Chinese University of Hong Kong*

Abstract: A new change-point detector for structure breaks in functional means is developed in this paper. The detector is based on a novel easy-to-implement approach of dimension reduction. One major advantage of the proposed method is its efficiency in selecting the basis functions that capture the change/jump of functional means, leading to a higher detection power. We thoroughly investigate the asymptotic properties of the proposed detector when both the sample size and the incorporated dimension increase. The numerical simulation studies justify the superiority of the proposed approach compared to the existing competitors and highlight the necessity of aligning the basis functions with the change to be detected. An application to annual humidity trajectories illustrates the practical superiority of the developed approach.

Key words and phrases: Change point analysis, Change alignment, Dimension reduction, Functional Mean, Weakly dependent functional data.

1. Introduction

This paper provides a new method to tackle a popular problem in functional data analysis, detecting the change point in functional means of a sequence of functional time series. The general setting is that a single change point partitions the entire sequence into two local stationary blocks, where the functions in each block share the same mean function.

There have been a number of methods developed for functional structural breaks in mean function. Many of them are developed based on dimension reduction or projection. A typical step in projection-based approaches is to project the functions onto a finite number of basis functions, and the projection scores are employed to detect the change in mean of independent or dependent functional data sequence. See, for example, Berkes *et al.* (2009), Aue *et al.* (2009), Zhang *et al.* (2011), Aston and Kirch (2012a) and the references therein. More recently, Fremdt *et al.* (2014) consider structural break detection by using functional principal component analysis (fPCA) with an increasing number of projections. Dimension reduction is also utilized to detect change points of multivariate functions under separability assumptions (e.g., spatial temporal data or brain image data), see Aston and Kirch (2012b), Gromenko, Kokoszka and Reimherr (2017) and Stoehr, Aston and Kirch (2021). Structural break detection in the coefficient operators of functional linear models is considered in Aue *et al.* (2014). Structural break detection in spectrum and trace of

covariance operator is studied in Jaruskova (2013) and Aue, Rice and Sönmez (2020). Test for stationarity of functional time series in the spectral domain is developed in Aue and van Delft (2020). Chiou, Chen and Hsing (2019) and Chen, Chiou and Huang (2021) study the multiple change-point problem for functional data.

In the change-point analysis of functional data, one major limitation of dimension reduction is that, when the selected basis functions are not aligned with the jump of mean function, the projection-based detector fails to detect the change points. To solve this problem, an alternative fully functional approach is employed in Horváth, Kokoszka and Rice (2014), Aue, Rice and Sönmez (2018) and Jiao, Frostig and Ombao (2022), which does not rely on dimension reduction. In the fully functional detection procedure, the null distribution involves infinitely many unknown parameters and requires additional truncation step, however. To circumvent this difficulty, Sharipov, Tewes and Wendler (2016) and Bucchia and Wendler (2017) study the bootstrap procedure. In addition to the fully functional approach, Torgovitski (2015) considered aligning the leading fPC with the change.

Although the fully functional detector is guaranteed to detect the change as the sample size increases, one major limitation of the approach is that it incorporates all basis functions that span the functional space, including potentially infinitely many irrelevant (unaligned with the change) basis functions. The irrelevant basis functions do not contribute to change point detection, and can potentially lead to loss

of detection power due to their nuisance effect. In the literature, sample trajectories are often pre-smoothed with a few smooth basis functions. In that case, the fully functional approach also performs decently for the smoothed functions, since the nuisance effect is significantly attenuated by functional smoothing, and thus is not very substantial when the number of pre-smoothing basis is kept small. Extended pre-smoothing can lead to a serious loss of information, however. It will make the fully functional detector fail when the functions are smoothed with low-frequency basis but the change functions are driven by high frequencies. Therefore, it is more advantageous to select basis functions which are informative to the change of mean, than to incorporate all basis functions or smooth the functions with pre-specified basis.

In this paper, we develop a new detection method for structural breaks in functional means. The *key idea* is to align the selected basis functions with the change/jump function. To achieve this goal, we introduce a discrepancy enhanced covariance (DEC) operator, of which the eigenfunctions constitute the basis functions for dimension reduction. The DEC operator involves two parts. The first part is the long-run covariance and the second part is the enhancement term, which is calibrated to magnify the influence of the change-aligned basis functions. These basis functions have the advantage that they are aligned with the jump function. Unlike the fully functional approach, the null distribution of the developed detector only involves a

finite number of parameters, and the nuisance effect of unaligned/irrelevant basis functions is substantially reduced. Another contribution of this paper is that we investigate the asymptotic properties under more complicated settings. Specifically, we allow both the change magnitude and the incorporated dimension to change with the sample size. We thoroughly investigate the regularity conditions under which the power of the proposed detector approaches one as the sample size goes to infinity, including the cases when the change magnitude diminishes.

The rest of the article is organized as follows. In Section 2, we develop the change-aligned detection procedure and discuss the implementation details. Theoretical results are discussed in Section 3. In Section 4, we report the simulation results under different settings. In Section 5, we present the real data analysis on annual humidity trajectories. The paper is concluded in Section 6. Proofs are given in the online supplementary materials.

2. Change-aligned Detection Procedure

2.1 Projection-based Detector

For a sequence of random functions $\{X_n(t) : n = 1, \ldots, N\}$, where $X_n \in L^2[0,1]$, a

single change-point model can be formulated as

$$
X_n(t) = \begin{cases} \mu_0(t) + e_n(t), & n \le k^*, \\[2mm] \mu_1(t) + e_n(t), & n > k^*, \end{cases} \tag{2.1}
$$

where $k^* = \lfloor N\theta^* \rfloor$ and θ^* is the fixed scaled location of the change-point in $[0,1]$, and the zero-mean random functions $\{e_n(t)\colon n \in \mathbb{N}\}$ take realizations in $L^2[0,1]$ and satisfy that $E \int e_n^2(t)dt < \infty$. In the space $L^2[0,1]$, the inner product of two elements $x(t), y(t)$ are defined as $\langle x, y \rangle = \int_0^1 x(t)y(t)dt$ and the norm is defined as $\|x\| = \{\int_0^1 x^2(t)dt\}^{1/2} < \infty$. It is assumed that $\{e_n(t)\colon n \in \mathbb{N}\}$ are weakly dependent as quantified in the following assumption.

Assumption 1. There is a measurable function $f\colon S^\infty \to L^2[0,1]$, where S is a measurable space, and the $i.i.d.$ innovations $\{\epsilon_n\colon n \in \mathbb{N}\}$ take values in S, so that $e_n(t) = f(\epsilon_n, \epsilon_{n-1}, \ldots)(t)$. In addition, there exists a m-dependent sequence $\{e_{n,m}(t)\colon i \in \mathbb{N}\}$, so that $e_{n,m}(t) = f(\epsilon_n, \ldots, \epsilon_{n-m+1}, \epsilon_{n-m}^*, \epsilon_{n-m-1}^*, \ldots)(t)$, where ϵ_n^* is an independent copy of ϵ_n, such that $\sum_{m=0}^\infty \{E\|e_n(t) - e_{n,m}(t)\|^p\}^{1/p} < \infty$ for some $p > 2$.

A process is termed L^p-m approximable if it satisfies Assumption 1 (see Hörmann & Kokoszka (2010)). This is a mild assumption which is satisfied by many processes, such as auto-regressive processes and moving average processes.

In this paper, a single structural break problem is considered. When there are multiple change-points, we propose to apply some localization method to segment

the whole sequence into multiple blocks, where at most one change-point (AMOC) assumption is made for each block, and then use the proposed approach to each block to detect the change-point. More details can be found in the real data analysis of humidity trajectories. This is beyond the scope of this paper, and we do not pursue its details here.

Therein, the goal is to detect whether a change point exists and to identify the location of the change-point. Define the jump function as $\delta(t) = \mu_0(t) - \mu_1(t)$, and the following test is implemented to detect the change point,

$$H_0: \|\delta\| = 0 \quad \text{vs} \quad H_a: \|\delta\| \neq 0. \tag{2.2}$$

Remark 1. In H_a, we require that $\|\delta\| \neq 0$, say, $\delta(t) \neq 0$ for a set of t with positive Lebesgue measure.

Given a sequence of basis functions $\{b_d(t): d \geq 1\}$, suppose that $X_n(t) = \sum_{d \geq 1} \eta_{nd} b_d(t)$ and let $\boldsymbol{\eta}_n = (\eta_{nd}, \ldots, \eta_{nD})'$ where $D > 0$, then the cumulative sum (CUSUM) is defined as

$$\boldsymbol{S}_{N,\theta} = \sum_{n=1}^{\lfloor N\theta \rfloor} \boldsymbol{\eta}_n - \theta \sum_{n=1}^{N} \boldsymbol{\eta}_n.$$

The projection-based method is based on the squared (scaled) CUSUM statistic,

$$T_N(\theta) = N^{-1}\|\boldsymbol{S}_{N,\theta}\|_2^2, \tag{2.3}$$

where $\|\cdot\|_2$ denotes the ℓ^2-norm. The value of $T_N(\theta)$ should be large at the true change point θ^*, thus by convention, the following max-type quantity is employed as

the detector of change point

$$T_N(\hat{\theta}_N^*) = \max_{0 < \theta < 1} T_N(\theta),$$

and for uniqueness, the infimum of the maximizers of $T_N(\theta)$, namely

$$\hat{\theta}_N^* = \inf\{\theta : T_N(\theta) = \sup_{\theta' \in (0,1)} T_N(\theta')\}$$

is assumed to be the change point candidate. In principle, it is desirable for the selected basis functions $\{b_d(t) : d = 1, \ldots, D\}$ to capture the jump function, say, $\langle b_d, \delta \rangle \neq 0$ for some d. Otherwise, the method fails even if $\|\delta\|$ is much bigger than zero.

The following result quantifies the null distribution of the projection-based detector.

Theorem 1. *Under Assumption 1 and H_0,*

$$T_N(\hat{\theta}_N^*) \xrightarrow{d} \sup_{\theta \in (0,1)} \boldsymbol{B}'(\theta)\Sigma_D \boldsymbol{B}(\theta),$$

where $\boldsymbol{B}(\theta) = (B_1(\theta), \ldots, B_D(\theta))'$ and $\{B_d(\theta) : d \geq 1\}$ are i.i.d. Brownian bridges and $\Sigma_D = \sum_{h=-\infty}^{\infty} Cov(\boldsymbol{\eta}_n, \boldsymbol{\eta}_{n+h})$.

The theorem follows from Theorem A.1 in Aue *et al.* (2009). Theorem 1 asymptotically validates the test of H_0. Specifically, H_0 is rejected if the test statistic $T_N(\hat{\theta}_N^*)$ exceeds the corresponding quantile of the null distribution $\sup_{\theta \in (0,1)} \boldsymbol{B}'(\theta)\Sigma_D \boldsymbol{B}(\theta)$.

2.2   Selection of Basis Functions

Define the (auto)covariance function of $\{X_n\colon n \in \mathbb{N}\}$ as $C_{X,h}(t,s) = E\{e_n(t)e_{n+h}(s)\}$, and the long-run covariance as $LC_X(t,s) = \sum_{h=-\infty}^{\infty} C_{X,h}(t,s)$. Assume that, with a sequence of positive and decreasing eigenvalues $\{\tau_d\colon d \geq 1\}$ and orthonormal eigenfunctions $\{\phi_d(t)\colon d \geq 1\}$, the spectral decomposition $LC_X(t,s) = \sum_{d\geq 1} \tau_d \phi_d(t)\phi_d(s)$ is allowed.

The selection of $\{b_d(t)\colon d = 1,\ldots,D\}$ highly influences the performance of the detector. A popular way of selecting the basis functions is to employ the major eigenfunctions of the (long-run) covariance operator $\mathcal{LC}_X(\cdot)$, induced by the kernel $LC_X(t,s)$, and the resulting null distribution is $\sup_{\theta \in (0,1)} \sum_{d=1}^{D} \tau_d B_d^2(\theta)$ (see Berkes *et al.* (2009), Hörmann, Kidziński and Hallin (2015) and Torgovitski (2015)). Such basis functions are not guaranteed to align with $\delta(t)$. To solve this problem, our approach is based on the major eigenfunctions of the discrepancy enhanced covariance (DEC) operator described as follows.

To separate the jump-aligned component and other irrelevant components, first transform the functions as follows:

$$Y_n^{(\kappa)}(t) = X_n(t) - \left\langle X_n, \frac{\delta}{\|\delta\| + \kappa} \right\rangle \frac{\delta(t)}{\|\delta\| + \kappa}, \tag{2.4}$$

where κ is a small-valued positive tuning parameter shrinking to zero as $N \to \infty$. Note that $\delta(t)$ is typically unknown, and the estimation of $\delta(t)$ will be discussed in

Section 2.3.

The term $Y_n^{(\kappa)}(t)$ in (2.4) is well defined under both H_0 and H_a. The DEC is then defined as

$$K^{(\kappa)}(t,s) = LC_{Y,\kappa}(t,s) + \rho\delta(t)\delta(s),$$

where ρ is the enhancement parameter to be specified. The quantity $LC_{Y,\kappa}(t,s)$ is the long-run covariance of $Y_n^{(\kappa)}(t)$, defined as

$$LC_{Y,\kappa}(t,s) = \sum_{h=-\infty}^{\infty} C_{Y,h}^{(\kappa)}(t,s),$$

where $C_{Y,h}^{(\kappa)}(t,s) = E\{Y_n^{(\kappa)}(t) - E(Y_n^{(\kappa)})(t)\}\{Y_{n+h}^{(\kappa)}(s) - E(Y_{n+h}^{(\kappa)})(s)\}$. By Mercer's theorem, suppose that a sequence of decreasing positive eigenvalues $\{\omega_d^{(\kappa)}\colon d \geq 1\}$ and a sequence of corresponding orthonormal eigenfunctions $\{\psi_d^{(\kappa)}\colon d \geq 1\}$ can be found such that

$$K^{(\kappa)}(t,s) = \sum_{d\geq 1} \omega_d^{(\kappa)}\psi_d^{(\kappa)}(t)\psi_d^{(\kappa)}(s).$$

It is proposed to make use of $\{\psi_d^{(\kappa)}(t)\colon d = 1,\ldots,D\}$ in defining the test statistic $T_N(\theta)$.

To understand this selection procedure, first define $K(t,s) = LC_Y(t,s) + \rho\delta(t)\delta(s)$ under H_a, where

$$LC_Y(t,s) = \sum_{h=-\infty}^{\infty} C_{Y,h}(t,s) = \sum_{h=-\infty}^{\infty} E\{Y_n(t) - E(Y_n)(t)\}\{Y_{n+h}(s) - E(Y_{n+h})(s)\}$$

and $\{Y_n(t) : n \in \mathbb{N}\}$ are defined as

$$Y_n(t) = X_n(t) - \left\langle X_n, \frac{\delta}{\|\delta\|} \right\rangle \frac{\delta(t)}{\|\delta\|}. \tag{2.5}$$

The kernel function $K(t,s)$ is positive definite, thus similar to $K^{(\kappa)}(t,s)$, the spectral decomposition can be found as $K(t,s) = \sum_{d\geq 1} \omega_d \psi_d(t) \psi_d(s)$.

Evidently, under H_a and as $\kappa \to 0$, $LC_{Y,\kappa} \to LC_Y$, thus $\{(\omega_d^{(\kappa)}, \psi_d^{(\kappa)}(t)) : d \geq 1\}$ converge to $\{(\omega_d, \psi_d(t)) : d \geq 1\}$. Clearly, $\delta(t)$ is orthogonal to all the eigenfunctions of $LC_Y(\cdot)$ due to the projection (2.5), and thus $\{\rho\|\delta\|^2, \delta(t)/\|\delta\|\}$ is a pair of eigenvalue and eigenfunction of $K(t,s)$. In what follows, it is supposed that $\omega_{d*} = \rho\|\delta\|^2$ and $\psi_{d*}(t) = \delta(t)/\|\delta\|$, and denote $\omega_{d*}^{(\kappa)}$ and $\psi_{d*}^{(\kappa)}(t)$ as the counterparts of $K^{(\kappa)}(t,s)$. Observe that $\psi_{d*}(t) = \delta(t)/\|\delta\|$ is the only eigenfunction of $K(t,s)$ aligned with the jump function $\delta(t)$, and a large value of ρ leads to large eigenvalue $\rho\|\delta\|^2$.

In practice, $\delta(t)$ is typically unknown and $K(t,s)$ may be not well-defined (when $\|\delta\| = 0$ under H_0, and this leads to the non-consistency of $\hat{\delta}/\|\hat{\delta}\|$). To solve this problem, we add the tuning parameter κ and obtain $\{Y_n^{(\kappa)}(t), n \geq 1\}$. It is of major interest to enhance the influence of the jump-aligned counterpart of $K^{(\kappa)}(t,s)$, namely, $\psi_{d*}^{(\kappa)}(t)$. In the following, $\psi_{d*}^{(\kappa)}(t)$ is termed *jump-aligned basis function*.

The step (2.4) is important for the selection of dimension D (see Section 2.5). An alternative approach is to employ the major functional principal components of $LC_X(t,s) + \rho\delta(t)\delta(s)$. While this is also reasonable, we still recommend to do the projection (2.4) first. The reason is that, without projection, it is hard to find all

aligned (not orthogonal to $\delta(t)$) eigenfunctions of $LC_X(t, s) + \rho\delta(t)\delta(s)$, making the selection of D much more complicated.

2.3 Estimations

This section illustrates the estimation of basis functions. The selected basis functions are the first D eigenfunctions of $K^{(\kappa)}(t, s)$, which incorporates an unknown quantity $\delta(t)$. Thus we need to first get a reasonable estimate of $\delta(t)$, and then estimate $K^{(\kappa)}(t, s)$, and the basis functions are then obtained from the estimate of $K^{(\kappa)}(t, s)$.

Denote $\hat{k}_N^{(f)}$ as the infimum of the maximizer(s) of the squared norm of the (scaled) fully functional cumulative sum statistic

$$M(k) = \frac{1}{N} \int \left(\sum_{n=1}^{k} X_n(t) - \frac{k}{N} \sum_{n=1}^{N} X_n(t) \right)^2 dt.$$

We propose to segment the entire functional sequence at the midpoint into two separate subsequences, $\{X_n(t): n = 1, \ldots, \hat{k}_N^{(f)}\}$ and $\{X_n(t): n = \hat{k}_N^{(f)} + 1, \ldots, N\}$, and estimate $\hat{\delta}(t)$ as follows

$$\hat{\delta}(t) = \frac{1}{\hat{k}_N^{(f)}} \sum_{n=1}^{\hat{k}_N^{(f)}} X_n(t) - \frac{1}{N - \hat{k}_N^{(f)}} \sum_{n=\hat{k}_N^{(f)}+1}^{N} X_n(t).$$

We make the following assumption on $\delta(t)$.

Assumption 2. Under H_a, $N^{\alpha_\delta} = O(\|\delta\|)$, where $\alpha_\delta > -1/2$.

Then we have the following result.

Theorem 2. *Under Assumption 1, 2 and H_a, $\|\hat{\delta} - \delta\| = O_p(N^{-1/2} + (N\|\delta\|)^{-1})$.*

We make Assumption 2 so that the change point can be detected asymptotically almost surely (see Theorem 6). This theorem tells that, under some mild conditions, $\hat{\delta}(t)$ is consistent with $\delta(t)$. The proof of this theorem can be found in Aue, Rice and Sönmez (2018) (Lemma A.5).

Then we construct $Y_n^{(\kappa)}(t)$ as

$$Y_n^{(\kappa)}(t) = X_n(t) - \left\langle X_n, \frac{\hat{\delta}}{\|\hat{\delta}\| + \kappa} \right\rangle \frac{\hat{\delta}(t)}{\|\hat{\delta}\| + \kappa}.$$

Further this leads to the empirical (auto)covariance of $\{Y_n^{(\kappa)}(t) \colon n \in \mathbb{N}\}$ displayed below

$$\widehat{C}_{Y,h}^{(\kappa)}(t, s) = \frac{1}{N - h} \sum_{n=1}^{N-h} \{Y_n^{(\kappa)}(t) - \bar{Y}_n^{(\kappa)}(t)\}\{Y_{n+h}^{(\kappa)}(s) - \bar{Y}_{n+h}^{(\kappa)}(s)\}, \qquad h \geq 0,$$

$$\widehat{C}_{Y,h}^{(\kappa)}(t, s) = \frac{1}{N + h} \sum_{n=|h|+1}^{N} \{Y_n^{(\kappa)}(t) - \bar{Y}_n^{(\kappa)}(t)\}\{Y_{n+h}^{(\kappa)}(s) - \bar{Y}_{n+h}^{(\kappa)}(s)\}, \qquad h < 0,$$

where

$$\bar{Y}_n^{(\kappa)}(t) = \begin{cases} \dfrac{1}{\hat{k}_N^{(f)}} \displaystyle\sum_{j=1}^{\hat{k}_N^{(f)}} Y_j^{(\kappa)}(t), & 1 \leq n \leq \hat{k}_N^{(f)}, \\[3mm] \dfrac{1}{N - \hat{k}_N^{(f)}} \displaystyle\sum_{j=\hat{k}_N^{(f)}+1}^{N} Y_j^{(\kappa)}(t), & \hat{k}_N^{(f)} + 1 \leq n \leq N, \end{cases}$$

The estimation of $K^{(\kappa)}(t, s)$ is then given as follows,

$$\widehat{K}^{(\kappa)}(t, s) = \sum_{h=-\ell}^{\ell} W\left(\frac{h}{\ell}\right) \widehat{C}_{Y,h}^{(\kappa)}(t, s) + \rho\hat{\delta}(t)\hat{\delta}(s),$$

where $W(\cdot)$ is the kernel function, and ℓ is the bandwidth. See, e.g., Rice and Shang (2017) for the selection of the bandwidth ℓ.

The corresponding empirical eigenfunctions are defined through the eigen-equation $\int \widehat{K}^{(\kappa)}(t,s)\hat{\psi}_d^{(\kappa)}(s)ds = \hat{\omega}_d^{(\kappa)}\hat{\psi}_d^{(\kappa)}(t)$, leading to $\hat{\boldsymbol{\eta}}_n = (\hat{\eta}_{n1},\ldots,\hat{\eta}_{nD})'$, where $\hat{\eta}_{nd} = \langle X_n, \hat{\psi}_d^{(\kappa)} \rangle$ and $\widehat{\boldsymbol{S}}_{N,\theta} = \sum_{n=1}^{\lfloor N\theta \rfloor} \hat{\boldsymbol{\eta}}_n - \theta \sum_{n=1}^{N} \hat{\boldsymbol{\eta}}_n$. Similarly, Σ_D is estimated with the kernel estimator

$$\widehat{\Sigma}_D = \sum_{h=-\ell}^{\ell} W\left(\frac{h}{\ell}\right) \widehat{C}_{\eta,h},$$

where $\widehat{C}_{\eta,h}$ is defined similar to $\widehat{C}_{Y,h}^{(\kappa)}$ with $Y_n^{(\kappa)}(t)$ replaced by $\hat{\boldsymbol{\eta}}_n$. The empirical detector is then obtained as $\widetilde{T}_N(\theta) = N^{-1}\|\widehat{\boldsymbol{S}}_{N,\theta}\|_2^2$.

2.4   Selection of ρ and κ

Note that we only need to do the enhancement when H_a is true, and the enhancement term only brings more estimation uncertainty under H_0.

Therefore, it is ideal that the enhancement term lays asymptotically trivial influence under H_0. To achieve this goal, ρ should be selected so that $\rho\hat{\delta}(t)\hat{\delta}(s)$ converges to zero faster than $\widehat{LC}_{Y,\kappa}(t,s) - E\{LC_{Y,\kappa}(t,s)\}$, say,

$$\rho\|\hat{\delta}(t)\|^2 / \|\widehat{LC}_{Y,\kappa}(t,s) - E\{\widehat{LC}_{Y,\kappa}(t,s)\}\| \xrightarrow{p} 0.$$

Under Assumption 1, it can be deduced that $NE\|\widehat{LC}_{Y,\kappa}(t,s) - E\{\widehat{LC}_{Y,\kappa}(t,s)\}\|^2 < \infty$ for any fixed κ and ℓ (see Theorem 4.1 in Hörmann & Kokoszka (2010) and Lemma

4 in Jiao, Frostig and Ombao (2022)), and it can also be deduced that $NE\|\hat{\delta}(t)\|^2 < \infty$ under Assumption 1 (see Lemma 2 in Jiao, Frostig and Ombao (2022)). Therefore, $\|\hat{\delta}(t)\|^2 = O_p(N^{-1})$ and $\|\widehat{LC}_{Y,\kappa}(t,s) - E\{\widehat{LC}_{Y,\kappa}(t,s)\}\| = O_p(N^{-1/2})$, and we propose to choose $\rho = N^\beta$, with some $0 < \beta < 1/2$.

For the identifiability of the jump aligned basis function, we propose to adjust the selected ρ so that $\rho\|\hat{\delta}\|^2$ lies in the middle of the two neighboring eigenvalues of $\widehat{\mathcal{LC}}_{Y,\kappa}(\cdot)$. Suppose that

$$\widehat{LC}_{Y,\kappa}(t,s) = \sum_{d \geq 1} \hat{\lambda}_d^{(\kappa)} \hat{\nu}_d^{(\kappa)}(t) \hat{\nu}_d^{(\kappa)}(s),$$

where $\hat{\lambda}_1^{(\kappa)} > \hat{\lambda}_2^{(\kappa)} > \cdots$. If $\rho\|\hat{\delta}\|^2$ is greater than the maximal eigenvalue of $\widehat{\mathcal{LC}}_{Y,\kappa}(\cdot)$, then ρ is selected so that $\rho\|\hat{\delta}\| - \hat{\lambda}_1^{(\kappa)}$ is greater than a non-trivial positive value L_ρ, e.g., $\hat{\lambda}_1^{(\kappa)} - \hat{\lambda}_2^{(\kappa)}$. The identifiability of $\psi_{d^*}^{(\kappa)}(t)$ will be discussed in Section 3.2.

The principle of selecting κ is that the term $\hat{\delta}/(\|\hat{\delta}\| + \kappa)$ in Eq. (2.4) converges to zero in probability as $N \to \infty$ under H_0 to solve the non-consistency problem ($\hat{\delta}/\|\hat{\delta}\|$ is not consistent and not well-defined under H_0), or equivalently $\kappa^{-2} E\|\hat{\delta}\|^2 \to 0$ under H_0. Therefore, we propose that $\kappa = N^{-\alpha_\kappa}\{\int \widehat{LC}_X(t,t)\,dt\}^{1/2}$ with some $0 < \alpha_\kappa < 1/2$. We require $\alpha_\kappa > 0$, since otherwise κ would mitigate the enhancement. Here, the role of $\int \widehat{LC}_X(t,t)\,dt$ is to attenuate the effect of data variation. In the simulation section, it is shown that the detection performance is *robust* to the selection of ρ and κ.

2.5   Selection of D

Cumulative percentage of variance is a widely accepted criterion for the selection of dimension, and is adjusted for the selection of D in our new method. Here, the goal is to incorporate the jump-aligned basis function $\psi_{d^*}^{(\kappa)}(t)$, of which the corresponding eigenvalue converges to $\rho\|\delta\|^2$ under H_a as $\kappa \to 0$. In principle, D should be selected so that 1) the selected basis functions explain sufficient data variation to attenuate the nuisance effect of the estimation deviation $\hat{\delta}(t) - E\hat{\delta}(t)$ and control the type-I error, and 2) $D > d^*$ to incorporate the jump-aligned basis function. More details are now discussed.

Define

$$R_\lambda(D) = \sum_{d=1}^{D} \hat{\lambda}_d^{(\kappa)} \bigg/ \sum_{d \geq 1} \hat{\lambda}_d^{(\kappa)}, \tag{2.6}$$

γ as a positive constant taking the value, e.g., 90%∼95%, D_{pre} as the minimal value of D satisfying $R_\lambda(D) \geq \gamma$, and $\hat{d}^*$ as the minimal value of d satisfying $\hat{\lambda}_d^{(\kappa)} < \rho\|\hat{\delta}\|^2$. Two scenarios are considered:

(1) if $\rho\|\hat{\delta}\|^2 > \hat{\lambda}_{D_{pre}}^{(\kappa)}$, set $D > D_{pre}$,

(2) if $\rho\|\hat{\delta}\|^2 \leq \hat{\lambda}_{D_{pre}}^{(\kappa)}$, set $D > \hat{d}^*$.

Now we give the reasoning of the selection. Condition (1) is important in controlling the size of the test. As an extreme case, if $\hat{\psi}_{d^*}^{(\kappa)}$ is the only incorporated basis function, the type-I error can be much higher than the nominal level, since the

estimation errors falsely favor a change-point under H_0. To attenuate such "over-enhancement" effect, it is necessary to incorporate multiple basis functions that capture sufficient variation to mitigate the estimation uncertainty of $\hat{\delta}(t)$. Condition (1) and (2) substantially increases the chance that the jump-aligned basis $\psi_{d^*}^{(\kappa)}$ is selected, and is important in solving the "non-alignment" problem in projection-based detector.

3. Theoretical Results

3.1 Convergence rate $\widehat{\mathcal{LC}}_{Y,\kappa}(t,s)$

In this section, we present the convergence rate of $\widehat{\mathcal{LC}}_{Y,\kappa}(t,s)$ under both H_0 and H_a. In what follows, the quantity "const." represents a positive constant and $\|\cdot\|_{\mathcal{S}}$ signifies the Hilbert-Schmidt norm. First we introduce some notations.

Assumption 3. There exist $\alpha_c, \alpha_\kappa > 0$, so that $\|C_{X,h}\|_2 \leq$ const. $h^{-\alpha_c}$ and $\kappa = O(N^{-\alpha_\kappa})$.

Assumption 4. $c_1^{-1}|u|^{\alpha_w} \leq 1 - W(u) \leq c_1|u|^{\alpha_w}$ for $|u| \leq 1$ and some $c_1 \geq 0$, $W(0) = 1$, $0 \leq W(\cdot) \leq 1$, $W(u) = W(-u)$, $W(u) = 0$ if $|u| > 1$, and the bandwidth ℓ satisfies $\ell = O(N^{\alpha_\ell})$, where $0 < \alpha_\ell < 1/2$.

Assumption 4 is suitable for a general class of kernel functions $W(u)$. When

$\alpha_w = 1$, the triangular kernel

$$W(u) = \begin{cases} 1 - |u|, & |u| \leq 1, \\ 0, & |u| > 1, \end{cases}$$

satisfies the assumption. When $\alpha_w = 0$, uniform kernel

$$W(u) = \begin{cases} 1, & |u| \leq 1, \\ 0, & |u| > 1, \end{cases}$$

satisfies the assumptions. Other values of α_w indicate polynomial decay rates of $W(u)$.

The following theorem quantifies the convergence rate of the estimated long-run covariance operator $\widehat{\mathcal{LC}}_{Y,\kappa}(\cdot)$ induced by $\widehat{LC}_{Y,\kappa}(t,s)$.

Theorem 3. *Under Assumption 1, 3, and 4, if H_0 is true and $N^{-1}\ell\kappa^{-2} \to 0$, then for an arbitrarily small $\epsilon > 0$,*

$$\|\widehat{\mathcal{LC}}_{Y,\kappa} - \mathcal{LC}_X\|_{\mathcal{S}} \leq O_p(1) N^{\max\{\alpha_\ell - 1/2, -(\alpha_c - 1)/\alpha_\ell, -1 + 2\alpha_\kappa + \alpha_\ell\}}$$

$$\vee \begin{cases} N^{-(\alpha_c - 1)\alpha_\ell}, & \text{if } \alpha_w - \alpha_c > -1. \\ N^{-(\alpha_c - 1)\alpha_\ell + \epsilon}, & \text{if } \alpha_w - \alpha_c = -1. \\ N^{-\alpha_w \alpha_\ell}, & \text{if } \alpha_w - \alpha_c < -1. \end{cases}$$

Moreover, if H_a is true and $\ell\kappa\|\delta\|^{-1} \to 0$, then for an arbitrary small $\epsilon > 0$,

$$\|\widehat{\mathcal{LC}}_{Y,\kappa} - \mathcal{LC}_Y\|_{\mathcal{S}} \leq O_p(1)N^{\max\{\alpha_\ell - 1/2, -(\alpha_c - 1)/\alpha_\ell, \alpha_\ell - \alpha_\kappa - \alpha_\delta\}}$$

$$\vee \begin{cases} N^{-(\alpha_c - 1)\alpha_\ell}, & \textit{if } \alpha_w - \alpha_c > -1. \\[2mm] N^{-(\alpha_c - 1)\alpha_\ell + \epsilon}, & \textit{if } \alpha_w - \alpha_c = -1. \\[2mm] N^{-\alpha_w \alpha_\ell}, & \textit{if } \alpha_w - \alpha_c < -1. \end{cases}$$

Theorem 3 gives the convergence rate of $\widehat{\mathcal{LC}}_{Y,\kappa}$ under both H_0 and H_a. The convergence rate is specified through multiple parameters. To simplify notations, r_0 is denoted as the convergence rate of $\widehat{\mathcal{LC}}_{Y,\kappa}$ under H_0, and r_a is denoted as the convergence rate of $\widehat{\mathcal{LC}}_{Y,\kappa}$ under H_a.

Denote $T_N^o(\theta)$ to be the projection-based test statistics based on the eigenfunctions of $\mathcal{LC}_X(\cdot)$. We develop the following theorem.

Theorem 4. *Under Assumption 1, 3, 4, and H_0, if $D \to \infty$ and $N^{-r_0} \sum_{d=1}^{D} \delta_{\tau,d}^{-1} \to 0$, then uniformly for $\theta \in (0,1)$, $\widetilde{T}_N(\theta) \xrightarrow{d} T_N^o(\theta)$, where $\delta_{\tau,1} = \tau_1 - \tau_2$, $\delta_{\tau,d} = \max\{\tau_{d-1} - \tau_d, \tau_d - \tau_{d+1}\}$ for $d \geq 2$, where $\{\tau_d : d \geq 1\}$ is defined in Section 2.2.*

The theorem illustrates that, under some regularity conditions, the null distribution of the new change-aligned detector is asymptotically equivalent to that of the ordinary fPC-based detector.

3.2   Selection of $\psi_{d*}^{(\kappa)}$ and Identifiability

In this section, we investigate the performance of change-alignment and the identifiability of the basis functions selected by the approach in Section 2.5. Since $K^{(\kappa)}(t,s)$ is tuned by ρ, the identifiability of the eigenfunctions of $K^{(\kappa)}(t,s)$ might be violated if ρ is not selected judiciously under H_a. Since $LC_{Y,\kappa}(t,s)$ and $LC_Y(t,s)$ are positive definite, by Mercer's theorem, we can find a sequence of decreasing positive values and orthonormal basis functions for each of them so that,

$$LC_{Y,\kappa}(t,s) = \sum_{d \geq 1} \lambda_d^{(\kappa)} \nu_d^{(\kappa)}(t) \nu_d^{(\kappa)}(s), \qquad LC_Y(t,s) = \sum_{d \geq 1} \lambda_d \nu_d(t) \nu_d(s).$$

Since $LC_{Y,\kappa}(t,s) \to LC_Y(t,s)$ under H_a as $\kappa \to 0$, the identifiability of $\{\psi_d^{(\kappa)}(t) \colon d \neq d^*\}$ is asymptotically guaranteed given the identifiability of $\{\nu_d(t) \colon d \geq 1\}$. Thus the identifiability of $\psi_{d*}^{(\kappa)}(t)$ is of major interest here. Recall that the selected ρ is adjusted so that $\rho\|\hat{\delta}\|^2$ lies in the middle of the two neighboring eigenvalues of $\widehat{\mathcal{LC}}_{Y,\kappa}(\cdot)$, and if $\rho\|\hat{\delta}\|^2$ is greater than the maximal eigenvalue of $\widehat{\mathcal{LC}}_{Y,\kappa}(\cdot)$, then ρ is selected so that $\rho\|\hat{\delta}\| - \hat{\lambda}_1^{(\kappa)}$ is greater than a positive value L_ρ. It can be shown that the identifiability of $\psi_{d*}^{(\kappa)}(t)$ is asymptotically guaranteed under some mild conditions. To justify this, we first introduce the following assumptions.

Assumption 5. Under H_a, the eigenvalues $\{\lambda_d \colon d \geq 1\}$ satisfy the conditions $R^{-1} d^{-\alpha_\lambda} \leq \lambda_d \leq R d^{-\alpha_\lambda}$, where R is a positive constant, and $\lambda_d - \lambda_{d+1} \geq \text{const.} d^{-\alpha_\lambda - 1}$.

Assumption 5 quantifies the decay rate of eigenvalues $\{\lambda_d \colon d \geq 1\}$ and restricts

the spacing of eigenvalues from being overly small, which enables the identifiability

of $\{\lambda_d \colon d \geq 1\}$ (see also Cai and Hall (2006)). We define $\Delta_{\rho,D} = |\hat{\lambda}_{d^*-1} - \hat{\lambda}_{D-1}^{(\kappa)}|$,

which depends on the selection of D and ρ.

Theorem 5. *Under Assumption 1–5, and H_a, if $D^{\alpha_\lambda+1} N^{-r_a} \to 0$ and $\Delta_{\rho,D}^{-1} \ell \kappa \|\delta\|^{-1} \to$*

0, then asymptotically almost surely, the following events are true:

1) $D > d^*$,

2) For an arbitrary $\epsilon > 0$, $\max\{\hat{\omega}_{d^*-1}^{(\kappa)} - \hat{\omega}_{d^*}^{(\kappa)}, \hat{\omega}_{d^*}^{(\kappa)} - \hat{\omega}_{d^*+1}^{(\kappa)}\} \geq (\omega_{d^*-1} - \omega_{d^*+1})/(2+\epsilon)$

or $\hat{\omega}_{d^*}^{(\kappa)} - \hat{\omega}_2^{(\kappa)} \geq L_\rho/(1+\epsilon)$ as $d^* = 1$.

Remark 2. See the definition of $\{\omega_d \colon d \geq 1\}$ and $\{\omega_d^{(\kappa)} \colon d \geq 1\}$ in Section 2.2.

Theorem 5 demonstrates that the identifiability of the jump-aligned basis is

guaranteed and the basis functions selected by the adjusted variance portion criterion

contain the jump-aligned basis as $N \to \infty$.

3.3　Power Studies

This section presents the asymptotic properties and the detection power of the new

detector. It is known that, when 1) the number of projections and the jump func-

tion is fixed, and 2) some of the selected basis functions are aligned with the jump

function, the power of CUSUM-type detector approaches one as the sample size goes

to infinity (see Berkes *et al.* (2009)). However, it has not been studied under what

conditions the asymptotically perfectly performed detector (with power approaching

one) can be achieved in a more general setting. In contrast, the dimension D, the magnitude of jump $\|\delta\|$, and the tuning parameters ρ and κ are all allowed to vary with the sample size N here. To present the theoretical results under this general setting, we first introduce the function

$$V(\theta) = \begin{cases} \theta(1 - \theta^*), & 0 < \theta \leq \theta^*, \\ \\ \theta^*(1 - \theta), & \theta^* < \theta < 1. \end{cases}$$

The following theorem presents the convergence rate of $\widetilde{T}_N(\theta)$. Here we suppose $\rho = O(N^\beta)$.

Theorem 6. *Under Assumptions 1–2 and H_a, define $r_\delta = \max\{\alpha_\ell - \alpha_\kappa - \alpha_\delta - 1/2, \alpha_\delta + \beta - 1/2 + \epsilon\}$ where $\epsilon > 0$ is some arbitrary small value, and*

$$U_N = \max\{D^{1/2}(\rho\|\delta\|^2)^{-1}, N^{r_\delta - r_a}(\rho\|\delta\|^2)^{-(1+1/\alpha_\lambda)}, N^{r_\delta - 1/2}\},$$

if $D^{\alpha_\delta + 3/2}/N^{r_a} \to 0$ and $D(\rho\|\delta\|^2)^{1/\alpha_\lambda} \to \infty$, then

$$\sup_{\theta \in (0,1)} \left| N^{-1}\widetilde{T}_N(\theta) - \|\delta\|^2 V^2(\theta) \right| \leq O_p(N^{-r_\delta})\|\delta\|U_N.$$

Since it is assumed that $\alpha_\delta > -1/2$, $N\|\delta\|^2 \to \infty$ as $N \to \infty$. Note that $d^* = O((\rho\|\delta\|^2)^{-1/\alpha_\lambda})$ under Assumption 5, and the condition $D(\rho\|\delta\|^2)^{1/\alpha_\lambda} \to \infty$ ensures that the jump-aligned basis function is selected. The theorem demonstrates that the convergence rate of $N^{-1}\widetilde{T}_N(\theta)$ is uniformly bounded by $N^{-r_\delta}\|\delta\|U_N$. Therefore, if the ratio $N^{-r_\delta}\|\delta\|U_N/\|\delta\|^2$ converges to zero, then it is sufficient to conclude that $\widetilde{T}_N(\theta) \xrightarrow{p} N\|\delta\|^2 V^2(\theta)$ uniformly for $\theta \in (0,1)$, which leads to $\widetilde{T}_N(\theta) \xrightarrow{p} \infty$.

Remark 3. The first element of U_N measures the estimation uncertainty of the first D eigenfunctions of $K^{(\kappa)}(t,s)$ excluding the jump aligned basis, and it changes with rate $D^{1/2}$. The second element of U_N pertains to the enhancement term. The last element of U_N pertains to the effect of $\{e_n(t)\colon n \geq 1\}$ on the CUSUM statistic. As a special case, if all tuning parameters and the jump magnitude are assumed to be fixed, then the convergence rate degenerates to a simple version

$$\sup_{\theta \in (0,1)} \left| N^{-1}\widetilde{T}_N(\theta) - \|\delta\|^2 V^2(\theta) \right| \leq O_p(N^{-1/2}),$$

which coincides with the result in Berkes *et al.* (2009).

Based on Theorem 6, the following corollary presents the regularity conditions that guarantees the detection power approaching one.

Corollary 1. *Under Assumptions 1–5 and H_a, if the conditions in Theorem 6 hold, and $N^{-r_\delta}\|\delta\|^{-1}U_N \to 0$, then $Pr(H_0$ is rejected$|H_a) \to 1$. In addition, $\hat{\theta}_N \xrightarrow{p} \theta^*$.*

The first part of the corollary can be obtained from Theorem 6, say, $\widetilde{T}_N(\theta) \xrightarrow{p} \infty$ under H_a. The consistency of $\hat{\theta}_N$ can be obtained by the continuous mapping theorem of "argmax" function and the fact that θ^* is the unique maximizer of $V(\theta)$.

4. Simulation

4.1   General Setting

Finite sample properties are investigated in this section. First, $N = 120$ or $N = 200$ identically distributed functions are generated over the unit interval $[0, 1]$ with $B = 20$ Fourier basis functions $\{F_6(t), \ldots, F_{25}(t)\}$ specified as follows

$$F_i(t) = \begin{cases} 1, & \text{if } i = 1. \\ \sqrt{2}\cos(2\pi kt), & \text{if } i = 2k. \\ \sqrt{2}\sin(2\pi kt), & \text{if } i = 2k+1. \end{cases}$$

The change in functional means is located in the middle of the sequence and is driven by the 2nd Fourier basis, which is orthogonal to the 20 basis functions $\{F_6(t), \ldots, F_{25}(t)\}$. The final functions are simulated by the following basis expansion contaminated by random noises $\{\epsilon_n(t_j) : n = 1, \ldots, N, \ j = 1, \ldots, T\}$

$$X_n(t_j) = \begin{cases} \sum_{d=1}^{B} \xi_{nd} F_{d+5}(t_j) + \epsilon_n(t_j), & 1 \leq n \leq \lfloor N/2 \rfloor, \ j = 1, \ldots, T \\ \sum_{d=1}^{B} \xi_{nd} F_{d+5}(t_j) + \delta(t_j) + \epsilon_n(t_j), & \lfloor N/2 \rfloor + 1 \leq n \leq N, \ j = 1, \ldots, T, \end{cases}$$

where $\epsilon_n(t_j) \overset{i.i.d.}{\sim} \mathcal{N}(0, s^2)$, and $t_j = j/100$, $T = 100$. We set $\delta(t) = aF_2(t)$, where $a = 0$ under H_0 and $a > 0$ under H_a. To highlight the effect of the magnitude $\|\delta\|$, different values of a are considered. The variation of random errors $\{\epsilon_n(t_j) : t_j = 1, \ldots, T, n \geq 1\}$ is tuned through s. The obtained functions are smoothed with the first 55 Fourier basis functions.

Define $\boldsymbol{\sigma} = \text{diag}(1.2^{-2}, 1.2^{-4}, \ldots, 1.2^{-2B})$. Two distributional setups of $\boldsymbol{\xi}_n = (\xi_{n1}, \ldots, \xi_{nB})$ are considered, namely, $\{\boldsymbol{\xi}_n \colon n \geq 1\}$ are

1. (Independent case) $i.i.d.$ random vectors following the distribution $\mathcal{N}(0, \boldsymbol{\sigma})$.

2. (Dependent case) a FMA(3) process $\boldsymbol{\xi}_n = \sum_{j=1}^{3} \Phi_j \boldsymbol{e}_{n-j} + \boldsymbol{e}_n$, where $\boldsymbol{e}_n \sim \mathcal{N}(0, \boldsymbol{\sigma})$, $\Phi_1 = 0.6 \boldsymbol{I}_B$, $\Phi_2 = 0.4 \boldsymbol{I}_B$, $\Phi_3 = 0.2 \boldsymbol{I}_B$ and $\boldsymbol{I}_B$ is the $B \times B$ identity matrix.

We use the R package *sde* to simulate the null distribution, and the simulations are run with the same seed under different settings. The proposed detector (denoted by CA) is compared with two other competitors, which are representative in change-point detection, say, the fPC-based approach (denoted by fPC, see e.g., Berkes *et al.* (2009)) and the fully functional approach (denoted by FF, see e.g., Aue, Rice and Sönmez (2018)). For the fPC-based detector, the dimension is selected so that the incorporated functional principal components explain 90% of the total data variation.

4.2   Empirical Size and Power

In this section, we compare the empirical size and power of the four methods. In each setting, the simulation runs are repeated for 3000 times at nominal level 0.05. The enhancement parameters ρ considered are $\rho_1 = N^{0.25}$, $\rho_2 = N^{0.3}$, $\rho_3 = N^{0.35}$, and $\rho_4 = N^{0.40}$. The tuning parameter κ considered here is $\kappa = \kappa_1 = N^{-0.4} \{ \int \widehat{LC}_X(t, t) \, dt \}^{1/2}$.

The empirical sizes and powers are reported in Table 1 for setting 1 and Table 2 for setting 2. We also test the performance of the new detector when $\kappa = \kappa_2 = N^{-0.35}\{\int \widehat{LC}_X(t,t)\,dt\}^{1/2}$, $\kappa = \kappa_3 = N^{-0.3}\{\int \widehat{LC}_X(t,t)\,dt\}^{1/2}$ under the $i.i.d.$ setting to show its robustness to κ, see Table 1 in the supplementary materials.

Table 1: Empirical sizes and powers under different values of a and s (i.i.d., $\kappa = \kappa_1$).

a	N	s	CA				FF	fPC
			ρ_1	ρ_2	ρ_3	ρ_4		
0.00	120	0.5	0.056	0.056	0.057	0.056	0.051	0.043
		1.5	0.056	0.055	0.056	0.056	0.048	0.043
		2.5	0.059	0.059	0.060	0.059	0.037	0.032
	200	0.5	0.055	0.056	0.055	0.054	0.046	0.049
		1.5	0.059	0.058	0.059	0.058	0.050	0.042
		2.5	0.065	0.063	0.064	0.064	0.042	0.040
0.25	120	0.5	0.149	0.149	0.148	0.149	0.127	0.058
		1.5	0.146	0.145	0.145	0.146	0.122	0.112
		2.5	0.149	0.151	0.149	0.149	0.104	0.097
	200	0.5	0.281	0.271	0.272	0.269	0.227	0.062
		1.5	0.296	0.291	0.289	0.290	0.245	0.226
		2.5	0.299	0.302	0.300	0.299	0.212	0.202
0.30	120	0.5	0.225	0.223	0.223	0.224	0.194	0.070
		1.5	0.226	0.226	0.225	0.226	0.190	0.191
		2.5	0.231	0.231	0.229	0.229	0.157	0.149
	200	0.5	0.539	0.523	0.522	0.522	0.448	0.086
		1.5	0.553	0.542	0.540	0.537	0.472	0.463
		2.5	0.528	0.530	0.532	0.533	0.402	0.399
0.35	120	0.5	0.372	0.368	0.368	0.368	0.329	0.142
		1.5	0.368	0.369	0.369	0.366	0.320	0.321
		2.5	0.352	0.353	0.354	0.352	0.263	0.253
	200	0.5	0.927	0.927	0.930	0.928	0.856	0.350
		1.5	0.883	0.872	0.868	0.868	0.824	0.818
		2.5	0.785	0.789	0.789	0.787	0.675	0.675

Table 2: Empirical sizes and powers under different values of a and s (FMA,

$\kappa = \kappa_1$).

a	N	s	CA				FF	fPC
			ρ_1	ρ_2	ρ_3	ρ_4		
0.00	120	2.0	0.050	0.051	0.050	0.049	0.048	0.012
		3.0	0.046	0.046	0.047	0.047	0.041	0.009
		4.0	0.038	0.038	0.038	0.039	0.032	0.005
	200	2.0	0.060	0.057	0.058	0.059	0.052	0.020
		3.0	0.057	0.055	0.055	0.055	0.044	0.017
		4.0	0.044	0.045	0.046	0.044	0.035	0.015
0.55	120	2.0	0.140	0.142	0.141	0.139	0.133	0.035
		3.0	0.127	0.127	0.127	0.125	0.110	0.030
		4.0	0.104	0.104	0.104	0.103	0.082	0.017
	200	2.0	0.266	0.270	0.267	0.265	0.241	0.144
		3.0	0.273	0.259	0.255	0.255	0.222	0.126
		4.0	0.232	0.226	0.224	0.223	0.187	0.114
0.65	120	2.0	0.216	0.216	0.216	0.214	0.199	0.063
		3.0	0.198	0.194	0.194	0.194	0.170	0.047
		4.0	0.163	0.163	0.160	0.161	0.127	0.032
	200	2.0	0.486	0.493	0.491	0.494	0.454	0.330
		3.0	0.494	0.478	0.465	0.465	0.419	0.285
		4.0	0.422	0.413	0.412	0.412	0.358	0.250
0.75	120	2.0	0.350	0.346	0.345	0.346	0.328	0.129
		3.0	0.322	0.318	0.316	0.316	0.281	0.096
		4.0	0.266	0.264	0.265	0.262	0.216	0.061
	200	2.0	0.800	0.803	0.803	0.805	0.773	0.669
		3.0	0.786	0.773	0.754	0.752	0.715	0.598
		4.0	0.706	0.693	0.693	0.695	0.637	0.523

It is noted that as the variation of random error increases, the detection power

sometimes increases. One interpretation of this phenomenon is that the random

errors are composed of oscillations over all frequencies, including the frequency

that drives the jump function, and the jump-aligned frequency is not captured by $\{F_6, (t) \ldots, F_{25}(t)\}$. As the variation of random error increases, the jump-aligned frequency is more likely to be incorporated into the basis functions. Therefore, the power does not necessarily decrease as s increases.

4.3   Variation of the Detected Change-points

To study the variation of the detected change-points, we provide the box-plots of the detected change-points. In each figure, there are six boxes. The first four boxes pertains to the proposed detector under $\rho = N^{0.25}$, $\rho = N^{0.3}$, $\rho = N^{0.35}$, and $\rho = N^{0.4}$ respectively. The 5th box pertains to the fully functional detector and the last one pertains to the fPC-based approach. Overall, the variance of the detected change-points of the proposed detector and the fully functional detector are similar, and that of the fPC-based procedure can be sometimes much higher. For the reason of page limit, the figures are moved to the supplementary materials.

4.4   Necessity of Change Alignment

To thoroughly investigate the necessity of aligning the basis functions with the jump function, we examine more comparisons between the new approach and the fPC-based approach. We consider a variety of cases where the alignment between the major eigenfunction of $LC_X(t, s)$ and the jump function $\delta(t)$ changes. Specifically, in

addition to the 20 basis functions used to generate $\{X_n(t)\colon n \geq 1\}$, the 2nd Fourier basis, which drives the jump function, is also incorporated in the simulation here. Specifically, $\{X_n(t)\colon n \geq 1\}$ can be expressed as

$$
X_n(t) = \begin{cases} \xi_n^a F_2(t) + \sum_{d=1}^{20} \xi_{nd} F_{d+5}(t), & 1 \leq n \leq \lfloor N/2 \rfloor, \\[2mm] \xi_n^a F_2(t) + \sum_{d=1}^{20} \xi_{nd} F_{d+5}(t) + \delta(t), & \lfloor N/2 \rfloor + 1 \leq n \leq N, \end{cases}
$$

see Section 4.1 for the details of $\delta(t)$. Observe that, only $F_2(t)$ is aligned with the potential mean change. Here, $\xi_n^a \overset{i.i.d.}{\sim} \mathcal{N}(0, s_a^2)$, $s_a = 0.9, 0.45, 0.2, 0.05$, and $\boldsymbol{\xi}_n$ are simulated under the independent case as described in Section 4.1. The role of s_a is to tune the alignment between the eigenfunctions of $LC_X(t, s)$ and the jump function, and a large value of s_a leads to a high rank of the jump-aligned function in the set of eigenfunctions of $LC_X(t, s)$, which makes it easier to select the jump-aligned basis for the fPC-based approach.

Here, we set $\rho = N^{0.4}$, and $N = 200$. The sizes/powers of the two approaches are displayed in Figure 1. From the results, we conclude that

1. Our proposed approach substantially increases the power of the detection when the employed eigenfunctions of $LC_X(t, s)$ cannot explain the change function.

2. When the employed eigenfunctions of $LC_X(t, s)$ can explain the change, the new detector still produces decent detection power. Thus there is no loss to apply the new detector.

In practice, it is tricky to know if the major eigenfunctions of $LC_X(t, s)$ can suffi-
ciently explain $\delta(t)$, thus the change-aligned procedure is more reliable and likely to
detect a true change-point.

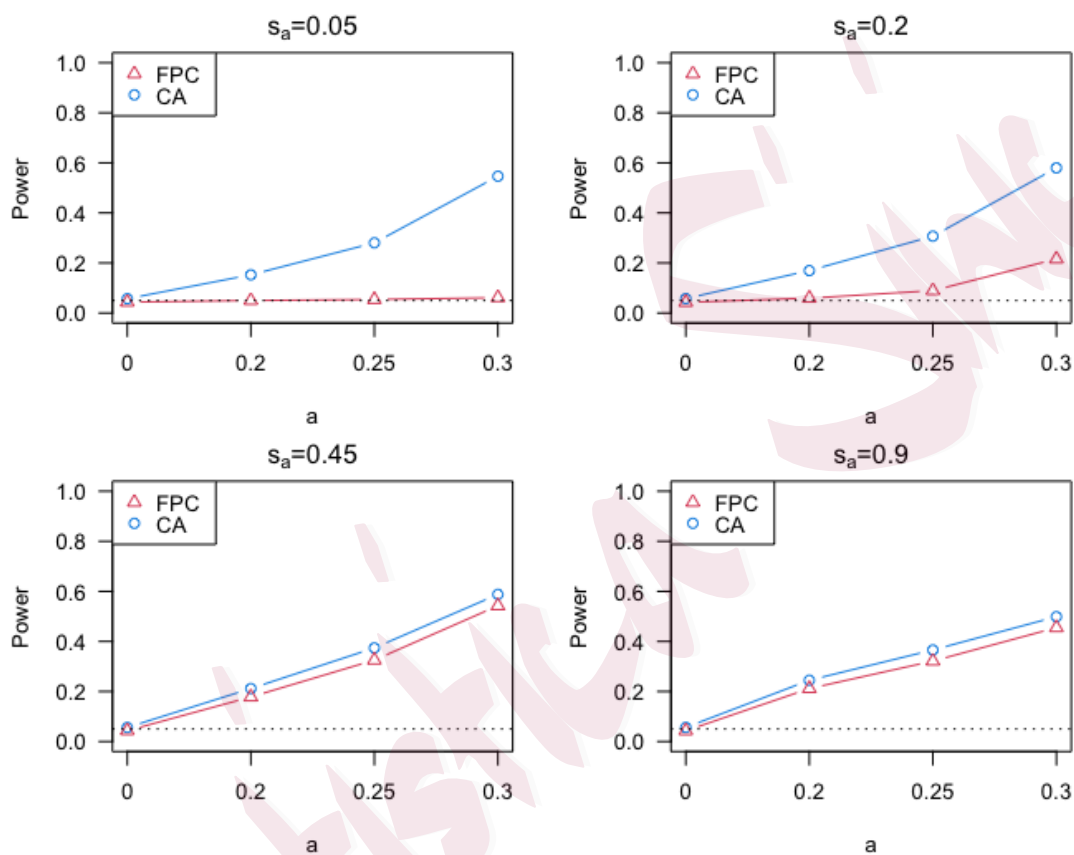Figure 1: Power Comparison. The dotted black line signifies the significance level.

4.5   Detection results when $\theta^* = 1/4$

In this section, we set $\theta^* = 1/4$ and simulate 120 or 200 independent functions following the model

$$X_n(t_j) = \begin{cases} \sum_{d=1}^{B} \xi_{nd} F_{d+5}(t_j) + \epsilon_n(t_j), & 1 \leq n \leq \lfloor N/4 \rfloor, \ j = 1, \ldots, T, \\ \sum_{d=1}^{B} \xi_{nd} F_{d+5}(t_j) + \delta(t_j) + \epsilon_n(t_j), & \lfloor N/4 \rfloor + 1 \leq n \leq N, \ j = 1, \ldots, T, \end{cases}$$

where $\delta(t) = aF_2(t)$, and $a = 0, 0.3, 0.35, 0.4, 0.45$, and $s = 0.5, 1.5, 2.5$. Here we consider the standardized detector $T_{w,N}(\theta) = T_N(\theta)/\{\theta(1-\theta)\}$, since it is more powerful to detect change point locating away from the middle. Suppose that there exists some $\epsilon > 0$, so that $\theta^* \in [\epsilon, 1-\epsilon]$, then under H_0,

$$T_{w,N}(\hat{\theta}_N^*) \xrightarrow{\mathcal{D}} \sup_{\theta \in [\epsilon, 1-\epsilon]} \frac{\boldsymbol{B}'(\theta)\Sigma_D\boldsymbol{B}(\theta)}{\theta(1-\theta)}, \qquad N \to \infty,$$

where $\hat{\theta}_N^* = \inf\{\theta : T_{\omega,N}(\theta) = \sup_{\epsilon \leq \theta' \leq 1-\epsilon} T_{\omega,N}(\theta')\}$. This result follows from Theorem A.1 in Aue et $al.$ (2009).

Remark 4. Compared with the non-standardized detector, the standardized detector gives more accurate detection when the true change point is near the boundary of the sequence, but when the true change point is near the middle, the detected change point of the standardized detector typically has higher variation.

Table 3 displays the power and type-I error of the three standardized detectors, and the enhanced detector is superior to the other competitors with respect to power

while controlling the type-I error well. The box-plots of the detected change points are shown in Figure 2. It shows that the variation of change points detected by the fPC-based method is substantially higher comparing to other methods when s is small. This is because, when s is small, the selected fPCs are not capable of capturing the jump function and thus the resulting fPC-based detector is not reliable.

Table 3: Empirical sizes and powers under different values of a and s.

s	N	method	a 0.00	0.30	0.35	0.40	0.45
0.5	120	CA	0.073	0.118	0.153	0.202	0.292
		FF	0.066	0.102	0.125	0.167	0.247
		fPC	0.056	0.060	0.065	0.081	0.155
	200	CA	0.072	0.174	0.273	0.474	0.843
		FF	0.063	0.150	0.234	0.429	0.803
		fPC	0.051	0.060	0.069	0.163	0.577
1.5	120	CA	0.069	0.114	0.142	0.197	0.286
		FF	0.064	0.103	0.124	0.173	0.249
		fPC	0.052	0.088	0.112	0.159	0.245
	200	CA	0.067	0.164	0.258	0.443	0.756
		FF	0.063	0.159	0.251	0.434	0.745
		fPC	0.054	0.138	0.240	0.433	0.748
2.5	120	CA	0.070	0.104	0.128	0.181	0.265
		FF	0.056	0.092	0.113	0.149	0.215
		fPC	0.047	0.075	0.093	0.129	0.192
	200	CA	0.068	0.166	0.261	0.417	0.647
		FF	0.062	0.136	0.222	0.371	0.602
		fPC	0.056	0.125	0.208	0.369	0.601

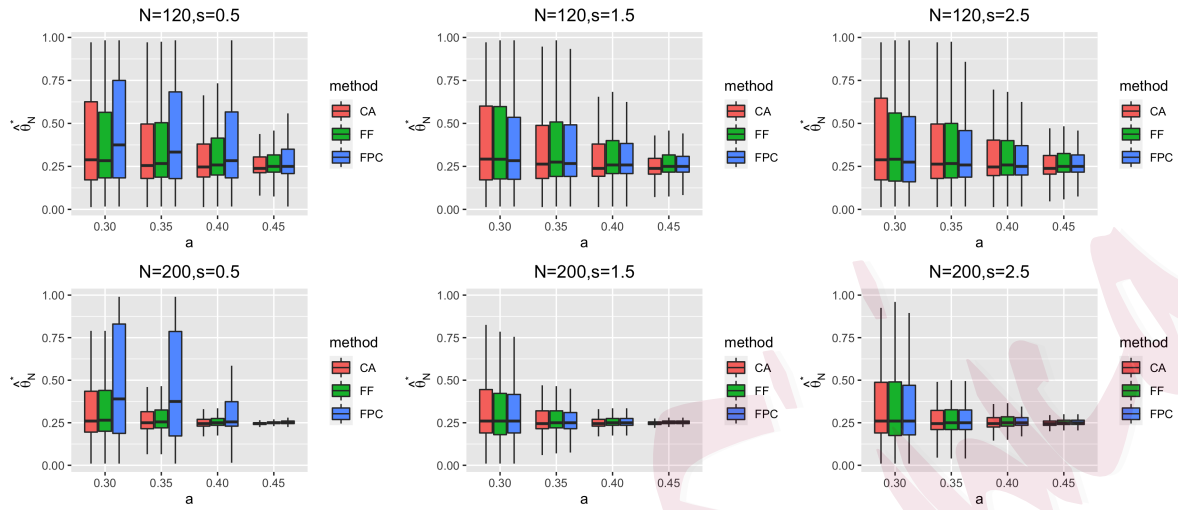Figure 2: Boxplots of $\hat{\theta}_N^*$.

4.6   Summary of Simulations

The comparisons are summarized as follows.

1. The type-I error of the proposed detector is well controlled around the nominal level under both $i.i.d.$ and dependent case.

2. The performance of the proposed change-aligned detector is *robust* to the selection of ρ and κ, and thus is not highly influenced by tuning parameters.

3. The power of the proposed detector is obviously higher than that of the fully functional detector and the fPC-based detector, especially when the noise variation is high. The fPC-based detector typically gives the worst performance especially when the leading ordinary fPCs cannot explain the change. The fully

functional approach, though performs better than the fPC-based approach, still gives suboptimal performance compared to the new approach especially when the noises become substantial. One explanation is that the fully functional approach incorporates the random noises into the detection procedure, and the nuisance effect of the random noises reduce the detection power. It numerically demonstrates the necessities of careful selection of basis functions.

5. Application to Annual Humidity Trajectories

In this section, the proposed approach is applied to daily humidity trajectories obtained in Basel-City, Switzerland in 2021. The raw data consist of $N = 365$ daily measurements of humidity recordings (one observation per hour, 24 observations for each day) that are converted into functional objects by using 24 Fourier basis functions. The data can be downloaded at www.meteoblue.com. Figure 3 displays the trajectories. For comparison, the proposed detector and the other two competitors (the fPC-based detector and the fully functional detector) are applied to date the time of the structural breaks.

5.1 Dynamic Segmentation

To attenuate the violation of at most one change-point assumption (AMOC), we first segment the entire sequence into multiple disjoint blocks. The segmentation approach
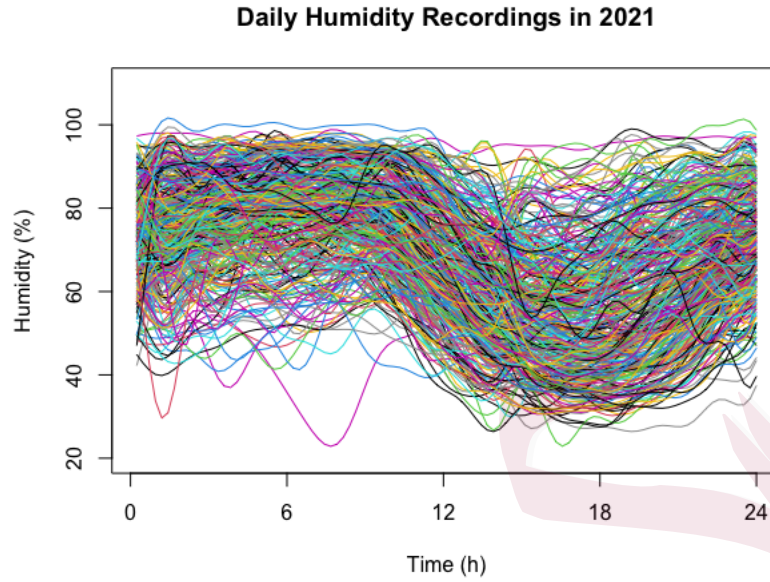
Daily Humidity Recordings in 2021

Figure 3: Daily humidity curves in Basel-City, Switzerland.

employed here is motivated by the dynamic segmentation approach (see Chiou, Chen

and Hsing (2019)) and is adapted for our own purpose, which is described below.

First we segment the whole functional sequences into 10 equal-length blocks

$$\{[\theta_r^{(0)}, \theta_{r+1}^{(0)}) \colon r = 1, \ldots, 10\},$$

where $\theta_1^{(0)} = 1$ and $\theta_{11}^{(0)} = 365$. Then recursively update the segment points as follows.

Given a subinterval $[\theta_r, \theta_{r+1})$ of $[1, 365]$ and any θ in the subinterval, the sample

covariance is calculated as follows

$$S_{[\theta_r, \theta_{r+1})}^{(\theta)}(t, s) = \frac{1}{\lfloor N\theta_{r+1} \rfloor - \lfloor N\theta_r \rfloor} \sum_{n=\lfloor N\theta_r \rfloor}^{\lfloor N\theta_{r+1} \rfloor} \{X_n(t) - \bar{X}_n^{(\theta)}(t)\}\{X_n(s) - \bar{X}_n^{(\theta)}(s)\},$$

where

$$
\bar{X}_n^{(\theta)}(t) =
\begin{cases}
\dfrac{1}{\lfloor N\theta \rfloor - \lfloor N\theta_r \rfloor} \displaystyle\sum_{n=\lfloor N\theta_r \rfloor}^{\lfloor N\theta \rfloor} X_n, \ n \in [\lfloor N\theta_r \rfloor, \lfloor N\theta \rfloor], \\[4mm]
\dfrac{1}{\lfloor N\theta_{r+1} \rfloor - \lfloor N\theta \rfloor} \displaystyle\sum_{n=\lfloor N\theta \rfloor+1}^{\lfloor N\theta_{r+1} \rfloor} X_n, \ n \in (\lfloor N\theta \rfloor, \lfloor N\theta_{r+1} \rfloor].
\end{cases}
$$

Suppose in the i-th interation, the segmentation points are $\{[\theta_r^{(i)}, \theta_{r+1}^{(i)}]: r = 1, \ldots, 10\}$, where $\theta_1^{(i)} = 1$ and $\theta_{11}^{(i)} = 365$ for all iteration i. For each $r > 1$, find the $\theta \in [\theta_{r-1}^{(i+1)}, \theta_{r+1}^{(i)})$ that minimizes $\|S_{[\theta_{r-1}^{(i+1)}, \theta_{r+1}^{(i)}]}^{(\theta)}\|_{\mathcal{S}}$, which is set as $\theta_r^{(i+1)}$. The iteration stops when $\max_{1 \leq r \leq 10} |\theta_r^{(i+1)} - \theta_r^{(i)}| < 1/N$. The final segmentation points are denoted by $\{\tilde{\theta}_r, r = 2, \ldots, 10\}$, and $\tilde{\theta}_1 = 1$ and $\tilde{\theta}_{11} = 365$.

Our proposal is that the whole sequence $[1, N]$ is segmented by $\{(\tilde{\theta}_r + \tilde{\theta}_{r+1})/2: r \geq 2\}$. Note that, in Chiou, Chen and Hsing (2019), $\{\tilde{\theta}_r, r = 2, \ldots, 10\}$ are considered as change-point candidates. Each candidate will be tested under the AMOC assumption, and the statistically nonsignificant ones are removed. Here we divide the sequence $[1, N]$ disjointly so that each segment contains one such candidate. The initial segmentation of $[1, 365]$ is displayed in Figure 4.

5.2   Backward Elimination

For each segment, we apply the three detectors considered in the simulation to detect and date the change-point under the AMOC assumption. If there is no change point detected in the subinterval $[\theta_r, \theta_{r+1}]$, then remove θ_{r+1} and test the change-point in

Annual Humidity in 2021

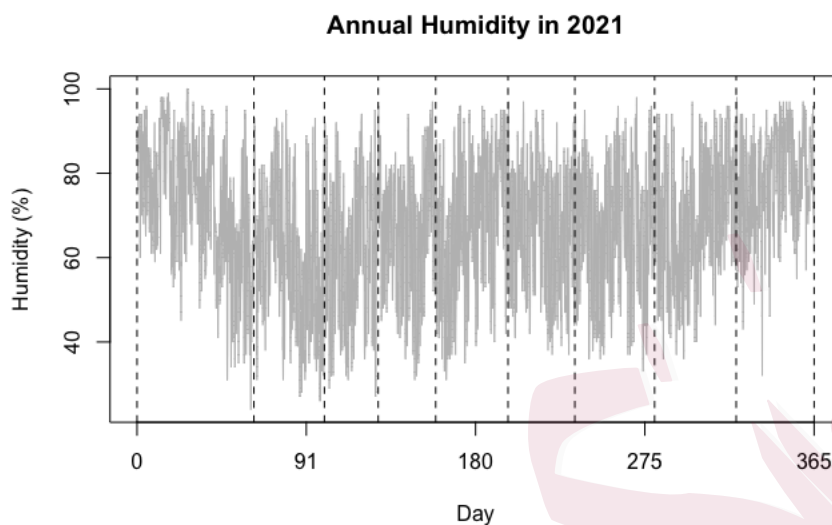Figure 4: Initial segmentation.

the longer subinterval $[\theta_r, \theta_{r+2}]$. The elimination procedure stops till no segmentation point is removed.

Here, $\ell = 3$, and $\rho = N^{0.4}$. Both our approach and the fully functional approach detect 4 change-points at significance level 0.05, which are displayed in Figure 5, while the fPC-based approach detect two change-points only, say, the 43th and 304th day of the year. The mean functions of the 5 segments are displayed in Figure 6.

In this application, although the proposed detector and the fully functional detector work similarly, there are cases when our approach is superior to the fully functional approach. There is evidence to believe that the developed procedure offers a more reliable method to detect change points in functional means.
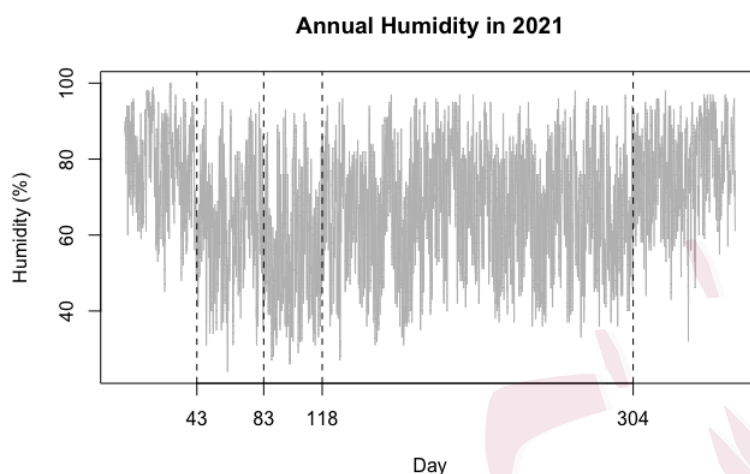
Figure 5: Detected days of change-points.

6. Conclusions

In this paper, a new change-aligned detector is introduced to detect and date the structural breaks in mean function of weakly dependent functional data. This detector has several advantages compared to the existing representative approaches including the fPC-based detector and the fully functional detector. Specifically, the fPC-based approach does not work while the employed fPCs fail to explain the structural breaks, and the fully functional approach essentially selects all basis functions that span the functional space, and thus suffers more from the nuisance effect of the irrelevant basis functions than the developed change-aligned procedure. The proposed detector relies on the carefully selected basis functions that are informative
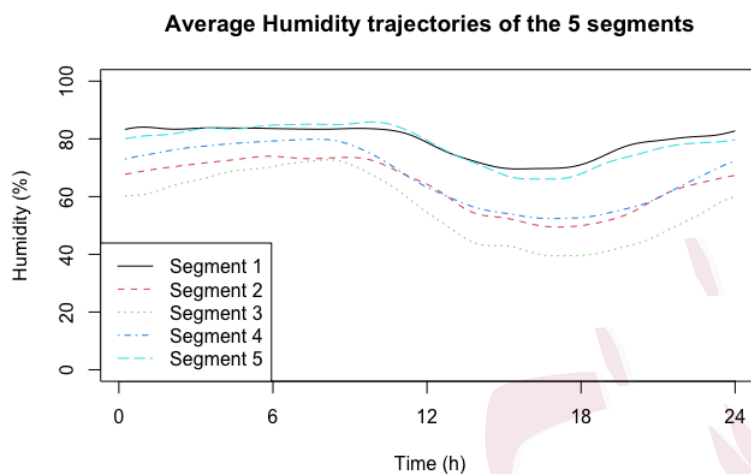
Average Humidity trajectories of the 5 segments

Figure 6: Average Humidity Trajectories of each segment.

to the change in mean, making it more reliable to detect the change while control-ling the type-I error close to the nominal level. In the simulation study, it is shown that the proposed detector performs better than the fully functional and fPC-based detectors, especially when the functions are contaminated by random errors or the leading fPCs cannot explain the change of mean.

Supplementary Materials

The supplementary materials contain the technical proofs, box-plots, and additional tables.

References

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

ASTON, J. A. & KIRCH, C. (2012b). Evaluating stationarity via change point alternatives with applications to fMRI data. *The Annals of Applied Statistics* **6**, 1906–1948.

AUE, A., GABRYS, R., HORVÁTH, L. & KOKOSZKA, P. (2009). Estimation of a change point in the mean function of functional data. *Journal of Multivariate Analysis* **100**, 1043–1073.

AUE, A., HÖRMANN, S., HORVÁTH, L. & HUŠKOVÁ, M. (2014). Dependent functional linear models with applications to monitoring structural change. *Statistica Sinica* **100**, 2254–2269.

AUE, A., HÖRMANN, S., HORVÁTH, L. & REIMHERR, M. (2009). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics* **37**, 4046–4087.

AUE, A., RICE, G. & 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**, 509–529.

AUE, A., RICE, G. & SÖNMEZ, O. (2020). Structural break analysis for spectrum and trace of covariance operators. *Environmetrics* **31**, e2617.

AUE, A. & VAN DELFT, A. (2020). Testing for stationarity of functional time series in the frequency domain. *The Annals of Statistics* **48**, 2505–2547.

BERKES, I., GABRYS, R., HORVÁTH, L. & KOKOSZKA, P. (2009). Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**, 927–946.

BUCCHIA, B. & WENDLER, M. (2017). Change-point detection and bootstrap for Hilbert space valued random fields. *Journal of Multivariate Analysis* **155**, 344–368.

CAI, T.-T. & HALL, P. (2006). Prediction in function linear regression. *The Annals of Statistics* **34**, 2159–2179.

CHIOU, J.-M., CHEN, Y.-T., & HSING, T. (2019). Identifying multiple changes for a functional data sequence with application to freeway traffic segmentation. *The Annals of Applied Statistics* **13**, 1430–1463.

CHEN, Y.-T., CHIOU, J.-M., & HUANG, T.-M. (2021). Greedy Segmentation for a Functional Data Sequence. *Journal of the American Statistical Association (online).* DOI: 10.1080/01621459.2021.1963261.

REFERENCES

FREMDT, S., HORVÁTH, L., KOKOSZKA, P. & STEINEBACH, J. (2014). Functional
data analysis with an increasing number of projections. *Journal of Multivariate
Analysis* **124**, 313–332.

GROMENKO, O., KOKOSZKA, P. & REIMHERR, M. (2017). Detection of change in
the spatiotemporal mean function. *Journal of the Royal Statistical Society: Series
B (Statistical Methodology)* **79**, 29–50.

HORVÁTH, L., KOKOSZKA, P. & RICE, G. (2014). Testing stationarity of functional
time series. *Journal of Econometrics* **179**, 66–82.

HÖRMANN, S., KIDZIŃSKI, L. & Hallin, M. (2015). Dynamic functional principal
components. *Journal of the Royal Statistical Society: Series B (Statistical Method-
ology)* **77**, 319-348.

HÖRMANN, S. & KOKOSZKA, P. (2010). Weakly dependent functional data. *The
Annals of Statistics* **38**, 1845–1884.

JARUŠKOVÁ, D. (2013). Testing for a change in covariance operator. *Journal of
Statistical Planning and Inference* **143**, 1500–1511.

JIAO, S., FROSTIG, R. D. & OMBAO, H. (2022). Breaking point detection
in functional covariance. *Scandinavia Journal of Statistics (online)*. DOI:
10.1111/sjos.12589.

RICE, G. & SHANG, H. L. (2017). A plug-in bandwidth selection procedure for long-run covariance estimation with stationary functional time series. *Journal of Time Series Analysis* **38**, 591–609.

SHARIPOV, O., TEWES, J. & WENDLER, M. (2016). Sequential block bootstrap in a Hilbert space with application to change point analysis. *Canadian Journal of Statistics* **44**, 300-322.

STOEHR, C., ASTON, J. A. & KIRCH, C. (2021). Detecting changes in the covariance structure of functional time series with application to fMRI data. *Econometrics and Statistics* **18**, 44-62.

TORGOVITSKI, L. (2015). Detecting changes in Hilbert space data based on "repeated" and change-aligned principal components. *arXiv preprint arXiv:1509.07409*

ZHANG, X., HAYHOE, K., WUEBBLES, D. & SHAO, X. (2010). Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electronic Journal of Statistics* **5**, 1765–1796.