Ann Inst Stat Math (2018) 70:691–715
https://doi.org/10.1007/s10463-017-0606-0

A more powerful test identifying the change in mean of functional data

Buddhananda Banerjee[1] · Satyaki Mazumder[2]

Received: 26 June 2016 / Revised: 25 January 2017 / Published online: 9 March 2017

Abstract An existence of change point in a sequence of temporally ordered functional data demands more attention in its statistical analysis to make a better use of it. Introducing a dynamic estimator of covariance kernel, we propose a new methodology for testing an existence of change in the mean of temporally ordered functional data. Though a similar estimator is used for the covariance in finite dimension, we introduce it for the independent and weakly dependent functional data in this context for the first time. From this viewpoint, the proposed estimator of covariance kernel is more natural one when the sequence of functional data may possess a change point. We prove that the proposed test statistics are asymptotically pivotal under the null hypothesis and consistent under the alternative. It is shown that our testing procedures outperform the existing ones in terms of power and provide satisfactory results when applied to real data.

Keywords Change point detection · Functional data analysis · Covariance kernel

1 Introduction

Functional data analysis (FDA) is becoming increasingly popular because of its wide applicability in various fields of statistics. It is noticed that in some cases the func-

✉ Buddhananda Banerjee
bbanerjee@maths.iitkgp.ernet.in

Satyaki Mazumder
satyaki@iiserkol.ac.in

1 Department of Mathematics, Indian Institute of Technology Kharagpur, Kharagpur 721302, India

2 Department of Mathematics and Statistics, Indian Institute of Science Education and Research Kolkata, Mohanpur 741246, India

tional representation of real-life data is more appropriate and appealing over its finite dimensional representation. In such case, often functional principal component analysis (FPCA) leads to more accurate inference than the usual principal component analysis (see Berkes et al. 2009). Ramsay and Silverman (2005) have enriched the literature with a detailed discussions on several techniques and usefulness of FPCA. Nonparametric inference of functional data is elaborately discussed by Ferraty and Vieu (2006). Some recent developments in many more aspects of FDA can be found in Ferraty (2011) and Bongiorno et al. (2014). Specifically the ANOVA technique in FDA is described by Zhang (2013). A rigorous mathematical treatment of functional data analysis is available in Hsing and Eubank (2015). The review articles by Cuevas (2014) and Goia and Vieu (2016) with the references therein are worth viewing for accessible references in FDA.

However, inference and especially prediction may alter if there exists an inherent change in the stochastic structure of the temporally observed functional data. The change may occur at an unknown point of time within the chronological sequence of data, but it is always challenging to test whether the change has occurred or not. In the cases of scalar and vector valued data, a considerable amount of contributions can be found in the works by Cobb (1978), Inclán and Tiao (1994), Davis et al. (1995), Antoch et al. (1997), Horváth et al. (1999), Kokoszka and Leipus (2000), Kirch et al. (2015) and references therein, among many others. In the context of functional data a change may occur in the mean function or in the covariance kernel of the data or both. Recently, Berkes et al. (2009) and Aue et al. (2009) have proposed a method for detecting changes in the mean functions of an observed set of functional data. Berkes et al. (2009), in their pioneering work in this context, have provided an elegant test procedure to decide the existence of a significant change in the mean function, whereas Aue et al. (2009) following the method of Berkes et al. (2009) have dealt with the detection of the position of the change in the mean function. One may look at the works by Hörmann and Kokoszka (2010), Aston and Kirch (2012) in the context of weakly dependent data for the similar problem, whereas the test of independence for functional data is given by Horváth et al. (2013). In practice, both are important to judge the existence of significant change in the mean function of the data and to identify the location of change point in the sequence. For example, while analysing the temperature of a certain region over a long period of time, it is very important to environmentalist to identify the time point after which a significant change in the mean temperature is observed as a possible effect of global warming.

In this paper, we come up with different methodologies to analyse the functional data subject to a possible change point and propose new statistical tests which are more powerful than the existing ones, for detecting the presence of a change in the mean function of the independent as well as the weakly dependent data. We introduce a new estimator of the covariance kernel of the functional data in the context of the change point problem. A similar approach to estimate the underlying variance is available for scalar data (see Fotopoulos and Jandhyala 2010; Bhattacharya 1987; Gombay and Horváth 1994; Hinkley 1970; Mei 2006; Shao and Zhang 2010). Unlike the finite dimensional data, functional principal component analysis is used to identify the eigenfunctions and the corresponding eigenvalues. All convergences are studied under $L^2(\cdot,\cdot)$ norm in a separable Hilbert space which makes the analysis technically more challenging than the finite dimensional cases. Under the null hypothesis, i. e. with no change in the mean function of the data, the uniform convergence of the newly proposed estimator of the covariance operator is established for the independent and the weakly dependent cases. Here we also show that under the null hypothesis the proposed test statistics converge in distribution to a functional of the Brownian bridges, as shown in Berkes et al. (2009) for independent data and as given in Hörmann and Kokoszka (2010) for weakly dependent data. Moreover, we prove here that the test is consistent under the alternative hypothesis when the number of the observations becomes large enough. Besides the consistency of the estimator of the covariance kernel under the null hypothesis, it also enjoys less asymptotic bias compared to that of the estimator provided by Berkes et al. (2009), Aue et al. (2009) or Hörmann and Kokoszka (2010) under the alternative hypothesis. We find the reduction in the asymptotic bias while estimating the covariance kernels leads to tests for an existence of a change point in the mean function with better power than the existing methods by Berkes et al. (2009) and Hörmann and Kokoszka (2010) for independent and weakly dependent data, respectively. The outcomes of an extensive simulation study reflect the same. It is also noted that our methods outperform the existing methods in a wide margin for small samples. Therefore, it is more advantageous to use the proposed methods in practice for deciding with the presence of significant change in the mean of the functional data for independent and weakly dependent data as well, especially when the data size is not big enough.

The organization of the paper is as follows. In Sect. 2, we introduce the required notation and definitions for introducing the subject. The details of the model and the null and alternative hypothesises, discussed in the paper, are mentioned in this section. In Sect. 3, the proposed estimator of the covariance kernel is introduced which is used for both independent and weakly dependent functional data. In Sect. 4, the uniform convergence of the estimator of the covariance operator is established under the null hypothesis for independent observations. In this section we provide the main theorem about the reduction of bias in estimation of the covariance kernel under the alternative hypothesis. The testing methodology and the asymptotic properties of test statistic for independent observations are developed in Sect. 5. Inference with weakly dependent data in the same context is dealt in Sect. 6. In Sect. 7, simulation results are provided in great detail for the independent and weakly dependent data separately. Here we show that our methods substantially improve over the existing methods of Berkes et al. (2009) and that of Hörmann and Kokoszka (2010) with independent and weakly dependent data, respectively. In Sect. 8, we show the performance of our test in real data. Here we analyse the data set used by Berkes et al. (2009) and compared with their results assuming independence. Moreover, we analyse average global temperature anomaly data assuming weak dependency. Remarks and conclusion of the work are given in the Sect. 9. Finally, we provide the required proofs of the results of Sect. 4 in "Appendix" section.

2 Preliminaries and assumptions

Let $X_i(t)$ for $i = 1, \ldots, N$, be random functions defined over a compact set $\tau = [0, 1]$ belonging to a separable Hilbert space with the inner product $\langle x, y\rangle = \int_0^1 x(t)y(t)\mathrm{d}t$. We assume that $X_i s$ are independent. We are interested in testing the equality of the mean functions of X_i for all $i = 1, 2, \ldots, N$. More precisely, the null hypothesis to test will be

$$H_0 : E(X_1(t)) = E(X_2(t)) = \cdots = E(X_N(t))$$

against the alternative

$$H_1 : E(X_1(t)) = \cdots = E(X_{k^*}(t)) \neq E(X_{k^*+1}(t)) = \cdots = E(X_N(t)),$$

for some $1 \leq k^* < N$. It is important to note that nothing is presumed about any property of the common mean under the null hypothesis. We deal with the situation when the data contain at most one change point; however, in case of applications we elaborate how to implement this method with multiple change points case. In Sect. 8, we specifically deal with the situation with more than one change points. There the data can be subdivided into several consecutive parts and within each part the mean function remains constant but it deviates between different contiguous parts. The details of the model with single change point are discussed in the Sect. 2.1. Under the null hypothesis, we express X_i, $i = 1, \ldots, N$, in the following manner.

$$\begin{aligned} X_i(t) &= \mu(t) + Y_i(t) \\ E(Y_i(t)) &= 0. \end{aligned} \tag{1}$$

The covariance kernel is defined as

$$c(t, s) = E(Y_i(t)Y_i(s)) \qquad t, s \in \tau, \tag{2}$$

where $c(t, s) \in L^2(\tau \times \tau)$. If $E\left(\int Y_i^2(t)\mathrm{d}t\right)^2$ is finite then the covariance operator of Y, which is a positive definite symmetric Hilbert–Schmidt (H-S) operator mapping from $L^2(\tau)$ to itself, will be of the form

$$C(x) = E[\langle Y, x\rangle Y] \text{ with } ||C||^2 = \int\int c^2(t, s)\mathrm{d}t\mathrm{d}s. \tag{3}$$

The evaluation of $C(x)$ at t, i.e. $C(x)(t)$, is given by

$$C(x)(t) = \int c(t, s)x(s)\,\mathrm{d}s \qquad \forall t \in \tau.$$

Moreover, Mercer's theorem in (Indritz 1963, Chap. 4) implies that $c(t,s)$ has the following spectral decomposition:

$$c(t,s)=\sum_{l=1}^{\infty}\lambda_l \upsilon_l(t)\upsilon_l(s) \qquad t,s\in\tau, \tag{4}$$

where each real scalar λ_l and function υ_l (in $L^2(\tau)$) are defined, for $t\in\tau$, as

$$C(\upsilon_l)(t)=\lambda_l\upsilon_l(t),\ l=1,2\ldots,$$
$$\text{i.e.}\ \int c(t,s)\upsilon_l(s)\mathrm{d}s=\lambda_l\upsilon_l(t),\ l=1,2,\ldots. \tag{5}$$

In other words, $\lambda_l s$ and $\upsilon_l s$ are the eigenvalues and the corresponding eigenfunctions, respectively, of the operator $C(\cdot)$. Since the eigenfunctions of the positive definite symmetric operator, $C(\cdot)$, form an orthonormal basis of $L^2(\tau)$ and the eigenvalues are positive, Karhunen–Loéve representation of Y_i holds good in $L^2(\tau)$ and is given by

$$Y_i(t)=\sum_{l=1}^{\infty}\sqrt{\lambda_l}\delta_{i,l}\upsilon_l(t), \tag{6}$$

where $\sqrt{\lambda_l}\delta_{i,l}=\langle Y_i,\upsilon_l\rangle=\int Y_i(s)\upsilon_l(s)$ is known as lth functional principal component score. By construction, the elements of the sequence $\{\delta_{i,l}\}_l$ are uncorrelated random variables with zero mean and unit variance and $\{\delta_{i,l}\}_l$ and $\{\delta_{j,l}\}_l$ are independent for $i\neq j$. Keeping this in consideration, first we estimate the covariance kernel in the following section and then develop a new methodology to test H_0. Now we specify the assumptions about the mean function μ and random element Y_i, based on which the asymptotic behaviour of the test statistic can be determined. From here onwards all integrations are computed over the compact set τ, unless otherwise mentioned.

2.1 Assumptions with independent observations

A1. $Y_i s$, are independent and identically distributed random elements with zero mean. Moreover, $Y_i s$ *and* $\mu\in L^2(\tau)$. $Y_i s$ also satisfy

$$E||Y_i||^4=E\left(\int Y_i^2(t)\mathrm{d}t\right)^2<\infty. \tag{7}$$

A2. First $d+1$ largest eigen values are distinct and positive for some for some natural number d.

A3. Under the alternative, with an existence of single change point the observations, $X_i, i=1,\ldots,N$ can be represented as follows

$$X_i(t)=\begin{cases}\mu_1(t)+Y_i(t), & 1\le i\le k^*\\ \mu_2(t)+Y_i(t), & k^*<i<N,\end{cases} \tag{8}$$

where $Y_i, i = 1, \ldots, N$ satisfy the assumption A1, $\mu_j(t), j = 1, 2$ are in $L^2(\tau)$ and $k^* = [N\theta]$, with $\theta \in (0, 1)$. Therefore, we assume that under the alternative hypothesis a single change in the mean may occur.

3 Estimation of covariance kernel

To estimate the covariance kernel, let us define the piecewise sample means for two segments

$$\widehat{\mu}_k(t) = \frac{1}{k}\sum_{i=1}^{k} X_i(t), \tag{9}$$

$$\widetilde{\mu}_k(t) = \frac{1}{N-k}\sum_{i=k+1}^{N} X_i(t), \tag{10}$$

where $k = [Nu]$ with $u \in (0, 1)$, implying $1 \leq k < N$. For $u = 1$ we define $\widehat{\mu}_N(t) = \frac{1}{N}\sum_{i=1}^{N} X_i(t)$. With the help of Eqs. (9) and (10), the newly proposed estimator of the covariance kernel for $k = [Nu]$ and $u \in (0, 1)$ is

$$\widehat{c}_u(t,s) = \frac{1}{N}\left[\sum_{i=1}^{k}(X_i(t)-\widehat{\mu}_k(t))(X_i(s)-\widehat{\mu}_k(s)) + \sum_{i=k+1}^{N}(X_i(t)-\widetilde{\mu}_k(t))(X_i(s)-\widetilde{\mu}_k(s))\right]. \tag{11}$$

For $u = 1$, we define $\widehat{c}_1(t,s) = \frac{1}{N}\left[\sum_{i=1}^{N}(X_i(t)-\widehat{\mu}_N(t))(X_i(s)-\widehat{\mu}_N(s))\right]$, which is commonly used as an estimator of the covariance kernel (see Berkes et al. 2009; Aue et al. 2009; Hörmann and Kokoszka 2010) for independent and weakly dependent data as well. However, we suggest to use the newly proposed estimator [Eq. (11)] of the covariance kernel in the context of change point analysis in mean function for both independent and weakly dependent functional data. The advantages of the proposed estimator [Eq. (11)] over the existing one are discussed in the subsequent sections.

4 Properties of $\hat{c}_u(t,s)$ with independent data

With the newly proposed estimator of the covariance kernel, we obtain the most important finding of this paper which is stated in the following theorem.

Theorem 1 *Defining* $c_u(t,s) := c(t,s)+\theta(1-\theta)\Delta(t)\Delta(s)\, f_\theta(u)$, *under the assumption A3,*

$$\int\int [\widehat{c}_u(t,s) - c_u(t,s)]^2\, \mathrm{d}t\mathrm{d}s \xrightarrow{P} 0, \textit{ as } N \uparrow \infty,$$

where

$$f_\theta(u) = \frac{\max\{u, \theta\} - \min\{u, \theta\}}{\max\{u, \theta\}(1 - \min\{u, \theta\})} \in [0, 1]$$

with $\theta \in (0, 1)$, $u \in (0, 1]$ *and* $\Delta(t) = \mu_1(t) - \mu_2(t)$.

Proof The proof of the theorem is provided in "Appendix" section. □

Theorem 2 *If the null hypothesis is true then* $\sup_{u\in(0,1]} ||\hat{C}_u - C|| \xrightarrow{P} 0$, *where the operators* $\hat{C}_u$ *and* C *have the kernels* $\widehat{c}_u(t, s)$ *and* $c(t, s)$*, respectively.*

Proof The proof is provided in "Appendix" section. □

Some more interesting observations, which show the greater applicability of the Theorem 1, are as follows.

Remark 1 It can be easily checked that $c_u(t, s)$ is a positive definite, symmetric satisfying

$$\int\int c_u^2(t, s)\, \mathrm{d}t\mathrm{d}s < \infty,$$

and hence is a covariance kernel.

Remark 2 If $u = 1$, that is, if commonly used estimator of $c(t, s)$ is used, then it is readily observable that, under the alternative, $\widehat{c}_1(t, s) \xrightarrow{P} c(t, s) + \theta(1-\theta)\Delta(t)\Delta(s) = \widetilde{c}(t, s)$, say, which is also proved by Berkes et al. (2009). We note here that whenever H_0 is false, $\widehat{c}_1(t, s)$ has a constant bias $\theta(1 - \theta)\Delta(t)\Delta(s)$. Therefore, for any $u \in (0, 1)$, the asymptotic bias of the estimator $\widehat{c}_u(t, s)$ is less than that of $\widehat{c}_1(t, s)$ under the alternative hypothesis.

Remark 3 If $u = \theta$, that is, when the data are partitioned in true position, then $\widehat{c}_\theta(t, s) \xrightarrow{P} c(t, s)$ and in that case asymptotic bias of $\widehat{c}_\theta(t, s)$ is zero, whereas asymptotic bias of $\widehat{c}_1(t, s)$ remains $\theta(1 - \theta)\Delta(t)\Delta(s)$.

5 Testing with independent observations

A few more notation and definitions are needed to be introduced here to state the further results.

Definition 1 The orthonormal functions $\omega_l^{(u)}(t)$ in $L^2(\tau)$ corresponding to the real scalars $\gamma_l^{(u)}$ are defined as orthonormal eigenfunctions associated with the eigenvalues $\gamma_l^{(u)}$ of the covariance operator $C_u(\cdot)$ from $L^2(\tau)$ to $L^2(\tau)$, defined as $C_u(x)(t) = \int c_u(t, s)x(s)\, \mathrm{d}s$, satisfying the relation

$$\int c_u(t, s)\omega_l^{(u)}(s)\, \mathrm{d}s = \gamma_l^{(u)}\omega_l^{(u)}(t). \tag{12}$$

Definition 2 The estimates of the eigenvalues $\gamma_l^{(u)}$ and eigenfunction $\omega_l^{(u)}(\cdot)$ are denoted by $\widehat{\lambda}_l^{(u)}$ and $\widehat{\upsilon}_l^{(u)}(\cdot)$, satisfying the relation

$$\int \widehat{c}_u(t,s)\widehat{\upsilon}_l^{(u)}(s)\,\mathrm{d}s = \widehat{\lambda}_l^{(u)}\widehat{\upsilon}_l^{(u)}(t). \tag{13}$$

With the above two definitions, the following important observations can be noted. It is immediate from Hörmann and Kokoszka (2010) that under the Assumption A3, for every $1 \le l \le d$ and $u \in (0,1]$, we have

$$\widehat{\lambda}_l^{(u)} \xrightarrow{P} \gamma_l^{(u)} \text{ and} \tag{14}$$

$$\int [\widehat{\upsilon}_l^{(u)}(t) - \widehat{h}_l^{(u)}\,\omega_l^{(u)}(t)]^2\,\mathrm{d}t \xrightarrow{P} 0, \tag{15}$$

where $\widehat{h}_l^{(u)} = \mathrm{sgn}\langle\omega_l^{(u)}(\cdot), \widehat{\upsilon}_l^{(u)}(\cdot)\rangle$. The result follows from the Theorem 1 and Lemmas 4.2 and 4.3 of Bosq (2000).

Remark 4 Under the null hypothesis, for all $1 \le l \le d$ and $u \in (0,1]$, $\widehat{\lambda}_l^{(u)} \xrightarrow{P} \lambda_l$ and $\widehat{\upsilon}_l^{(u)}(\cdot)$ converges to υ_l in probability, in $L^2(\tau)$. Moreover, under the alternative hypothesis, if $u=\theta$ then for all $1 \le l \le d$, $\widehat{\lambda}_l^{(\theta)} \xrightarrow{P} \lambda_l$ and $\widehat{\upsilon}_l^{(\theta)}(.)$ converges to υ_l in probability, in $L^2(\tau)$. It, in fact, can be easily seen that

$$\sup_{0<u\le 1} \int [\widehat{\upsilon}_l^{(u)}(t) - \widehat{h}_l^{(u)}\,\omega_l^{(u)}(t)]^2\,\mathrm{d}t \xrightarrow{P} 0.$$

In the direction of the eigenfunctions $\widehat{\upsilon}_l^{(u)}(\cdot)$ corresponding to the largest d eigenvalues $\widehat{\lambda}_l^{(u)}$, the non-central scores can be obtained as

$$\widehat{\eta}_{i,l}(u) = \int X_i(t)\widehat{\upsilon}_l^{(u)}(t)\,\mathrm{d}t,\ i=1,\ldots,N,\ l=1,\ldots d. \tag{16}$$

Utilizing the score functions, as defined above, we provide a statistic and its distributional convergence in the following theorem which will be important to know to construct the test statistic and perform the asymptotic test. First we define the statistic based on the self-normalized partial sums in d dimensions

$$R_N(u) = \frac{1}{N}\sum_{l=1}^{d}\frac{1}{\widehat{\lambda}_l^{(u)}}\left(\sum_{i=1}^{[Nu]}\hat{\eta}_{i,l}(u) - u\sum_{i=1}^{N}\hat{\eta}_{i,l}(u)\right)^2. \tag{17}$$

Further denoting $B_1(\cdot),\ldots,B_d(\cdot)$ the standard independent Brownian bridges, the following theorem is provided.

Theorem 3 *Let the assumptions A1–A3 hold. Then with the proper embedding of Skorohod topology in* $D[0, 1]$*, under the null hypothesis and as* $N \uparrow \infty$*,*

$$R_N(u) \xrightarrow{d} \sum_{l=1}^{d} B_l^2(u), \qquad 0 \le u \le 1. \tag{18}$$

Proof The proof follows from the Theorem 1, Eqs. 14 and 15 and the proof of Theorem 6.1 of Horváth and Kokoszka (2012). □

5.1 Test statistic

Finally we define the test statistic $H_{N,d} := \int_0^1 R_N(u)\mathrm{d}u$ which can be computed as:

$$\frac{1}{N^2} \sum_{l=1}^{d} \sum_{[Nu]=1}^{N} \frac{1}{\widehat{\lambda}_l^{(u)}} \left(\sum_{i=1}^{[Nu]} \hat{\eta}_{i,l}(u) - u \sum_{i=1}^{N} \hat{\eta}_{i,l}(u) \right)^2. \tag{19}$$

Remark 5 Using Theorem 3, it is immediate to see that $H_{N,d} \xrightarrow{d} \int_0^1 \sum_{l=1}^{d} B_l^2(u)\mathrm{d}u$ under the null hypothesis, because integral is a continuous functional and $U(R_N(\cdot)) \xrightarrow{d} U\left(\sum_{l=1}^{d} B_l^2(\cdot)\right)$ for any continuous functional $U : D[0, 1] \to \mathbb{R}$ (see Berkes et al. 2009 for further details).

Remark 6 The test statistic, $S_{N:d}$, proposed by Berkes et al. (2009) can be expressed in our notation as

$$S_{N,d} = \frac{1}{N^2} \sum_{l=1}^{d} \frac{1}{\widehat{\lambda}_l^{(1)}} \sum_{[Nu]=1}^{N} \left(\sum_{i=1}^{[Nu]} \hat{\eta}_{i,l}(1) - u \sum_{i=1}^{N} \hat{\eta}_{i,l}(1) \right)^2. \tag{20}$$

In Sect. 7.1.2, we compare the performance of $H_{N,d}$ with $S_{N:d}$.

Remark 7 The distribution of $\int_0^1 \sum_{l=1}^{d} B_l^2(u)\mathrm{d}u$ can be found in Kiefer (1959), and its $(1-\alpha)$th quantile are given in the Table 1 of Berkes et al. (2009). We use these asymptotic critical values for performing the tests, and H_0 is rejected at $100(1-\alpha)\%$ confidence level if the observed value of $H_{N,d}$ is bigger than the tabulated $(1-\alpha)$th quantile $K_d(\alpha)$ in Berkes et al. (2009).

Now we show that the proposed test is consistent under the alternative hypothesis. Basically we show here that $H_{N,d} \xrightarrow{P} \infty$ under the alternative hypothesis with single change point. The following theorem assures the claim.

Theorem 4 *Under the assumption A3 and $N \uparrow \infty$,*

$$\frac{1}{N} H_{N,d} \xrightarrow{P} \sum_{l=1}^{d} \int_0^1 \frac{g_l^2(u)}{\gamma_l^{(u)}} \mathrm{d}u,$$

where $g_l(u) = \min\{\theta, u\}\,(1 - \max\{\theta, u\}) \int \Delta(t)\omega_l^{(u)}(t)\mathrm{d}t$.

Proof The proof follows from the Theorem 1 and the following lemma. □

Lemma 1 *Under the assumption A3,* $\sup_{0 \le u \le 1} \left| N^{-1} R_N(u) - \sum_{l=1}^{d} \frac{g_l^2(u)}{\gamma_l^{(u)}} \right| = o_P(1).$

Proof Proof of the lemma follows from the proof of the Theorem 2 of Berkes et al. (2009). □

Clearly from the Theorem 4 if $\int_0^1 \frac{g_l^2(u)}{\gamma_l^{(u)}}\mathrm{d}u > 0$ for some $1 \le l \le d$, then $H_{N,d} \xrightarrow{P} \infty$.

Similar to Berkes et al. (2009), the change point θ is estimated by finding the value of u which maximizes the function $R_N(u)$. For uniqueness we define the estimator formally as

$$\widehat{\theta}_N = \inf\{u' : R_N(u') = \sup_{0 \le u \le 1} R_N(u)\}. \tag{21}$$

It can be easily shown that (using Lemma 1), under the assumption A3, $\widehat{\theta}_N \xrightarrow{P} \theta$ provided $\langle \Delta, \omega_l^{(u)}(\cdot) \rangle \ne 0$ for all $u \in (0, 1]$ (see, for example, the proposition 1 and its proof of Berkes et al. (2009)).

6 Inference with weakly dependent observations

To construct a testing procedure for detecting change in mean function for weakly dependent data, we need to introduce the following definition and assumptions. Let $L_H^p = L_H^p(\Omega, \mathscr{A}, P)$ be the space of Hilbert valued random functions Y with common probability space $(\Omega, \mathscr{A}, P)$ such that $(E||Y||^p)^{1/p} < \infty$ for any $p > 1$. Now we have the definition from Hörmann and Kokoszka (2010).

Definition 3 A sequence $\{Y_n\} \in L_H^p$ is called $L^p - m-$approximable if every Y_n has the representation,

$$Y_n = f(\epsilon_n, \epsilon_{n-1}, \ldots),$$

where the ϵ_i are i.i.d. elements taking values in a measurable space S, and f is a measurable function $f : S^\infty \mapsto L_H^2$. Moreover, we assume that if ϵ_i' is an independent copy of ϵ_i defined on the same probability space, then letting

$$Y_n^{(m)} = f(\epsilon_n, \epsilon_{n-1}, \ldots \epsilon_{n-m+1}, \epsilon_{n-m}', \epsilon_{n-m-1}', \ldots)$$

we have $\sum_{m=1}^{\infty}(E||Y_n - Y_n^{(m)}||^p)^{1/p} < \infty$. For our applications, choosing $p = 4$ will be sufficient to establish the convergence of the estimated covariance kernel and operator.

If we assume the followings

A4. Each $X_n = \mu + Y_n$ is measurable $(\mathscr{B}_{[0,1]} \times \mathscr{A})/\mathscr{B}_{\mathbb{R}}$ and
A5. $\sup\limits_{t\in[0,1]} E|X(t)|^2 < \infty$

then we can define an integral covariance operator W with the kernel $w(t, s) = E(Y_i(t), Y_i(s))$. Using Eqs. (9) and (10) for all $u \in (0, 1]$, we get an estimator of covariance kernel $w(t, s)$ as

$$\widehat{c}_u(t,s) = \frac{1}{N}\left[\sum_{i=1}^{k}(X_i(t) - \widehat{\mu}_k(t))(X_i(s) - \widehat{\mu}_k(s)) + \sum_{i=k+1}^{N}(X_i(t) - \widetilde{\mu}_k(t))(X_i(s) - \widetilde{\mu}_k(s))\right], \tag{22}$$

for each $[Nu] = k$ similar to the Eq. (11).

6.1 Properties of $\hat{c}_u(t, s)$

In this subsection, we study the properties of $\hat{c}_u(t, s)$ for weakly dependent data. We have the following theorem similar to the Theorem 1.

Theorem 5 *When $\{Y_i\}$ is a centred, stationary, ergodic and $L_H^4 - m-$approximable sequence, defining $w_u(t, s) := w(t, s) + \theta(1-\theta)\Delta(t)\Delta(s)\ f_\theta(u)$, under the assumption A4,*

$$\int\int [\widehat{c}_u(t,s) - w_u(t,s)]^2\, \mathrm{d}t\mathrm{d}s \xrightarrow{P} 0, \textit{ as } N \uparrow \infty,$$

where $f_\theta(u)$ is as defined in Theorem 1 with $\theta \in (0, 1), u \in (0, 1]$ and $\Delta(t) = \mu_1(t) - \mu_2(t)$.

Proof The proof is similar to that of the Theorem 1 following the arguments in Aston and Kirch (2012) and Rao (1962). □

The consistencies of the eigenelements obtained naturally from the above estimate of the covariance kernel are ensured by the following theorem.

Theorem 6 *Let $\{X_i\} \in L_H^4$ be $L^4 - m-$approximable sequence. If the null hypothesis is true then $\sup\limits_{u\in(0,1]} ||\hat{C}_u - W|| \xrightarrow{P} 0$, where the operators $\hat{C}_u$ and W have the kernels $\widehat{c}_u(t, s)$ and $w(t, s)$, respectively.*

Proof The proof similar to Theorem 2. □

6.2 Construction of test statistic

Definitions 1 and 2 hold good for the weakly dependent case also, and so the consistency of corresponding estimated eigenelements are ensured. Under the assumption A2, we denote the estimated non-central score vector for the ith observation by column vector

$$\hat{\boldsymbol{\xi}}_i(u) = (\hat{\xi}_{i,1}(u), \hat{\xi}_{i,2}(u), \ldots, \hat{\xi}_{i,d}(u))^T, \text{ with } u \in (0,1], \tag{23}$$

where $\hat{\xi}_{i,l}(u)$ is the score of random function X_i along the direction of the eigenfunction corresponding to the lth largest estimated eigenvalue of $\hat{C}_u$. Let us assume that $\boldsymbol{\Sigma_{LR}} = \sum_{r=-\infty}^{\infty} \Gamma_r$ stands for long-run covariance matrix of a d-dimensional multivariate time series with Γ_r denoting the covariance matrix of lag $r \in \mathbb{Z}$. Representing the centred scores by

$$\tilde{\boldsymbol{\xi}}_{a:b}(u) = \left[\hat{\boldsymbol{\xi}}_a(u) - \mu_{a:b}(\hat{\boldsymbol{\xi}}(u)), \hat{\boldsymbol{\xi}}_{a+1}(u) - \mu_{a:b}(\hat{\boldsymbol{\xi}}(u)), \ldots, \hat{\boldsymbol{\xi}}_b(u) - \mu_{a:b}(\hat{\boldsymbol{\xi}}(u))\right],$$

where $\mu_{a:b}(\hat{\boldsymbol{\xi}}(u)) = \sum_{i=a}^{b} \hat{\boldsymbol{\xi}}_i(u)/(b-a+1)$ we define

$$\boldsymbol{S_q^{[m:n]}(u)} = \left(\tilde{\boldsymbol{\xi}}_{m:n}(u)\tilde{\boldsymbol{\xi}}_{m:n}^T(u) + \sum_{j=1}^{q} \varpi_q(j)\left[\tilde{\boldsymbol{\xi}}_{m:(n-j)}(u)\tilde{\boldsymbol{\xi}}_{(m+j):n}^T(u) + \tilde{\boldsymbol{\xi}}_{(m+j):n}(u)\tilde{\boldsymbol{\xi}}_{m:(n-j)}^T(u)\right]\right), \tag{24}$$

for $1 \leq m < n$, with some suitable weight function $\varpi_q(j)$, where q is so chosen that $q^4/N \to 0$ for large N (see Hörmann and Kokoszka 2010). So we get a consistent kernel estimate of $\boldsymbol{\Sigma_{LR}}(\hat{\boldsymbol{\xi}}(u))$ by

$$\widehat{\boldsymbol{\Sigma}}_{LR}(\hat{\boldsymbol{\xi}}(u)) = \frac{1}{N}\left[\boldsymbol{S_q^{[1:[Nu]]}(u)} + \boldsymbol{S_q^{[1+[Nu]:N]}(u)}\right], \tag{25}$$

for each $[Nu] = 1, 2, \ldots, N$. Finally we define the test statistic as

$$H_{N,d}^* = \frac{1}{N}\int_0^1 \left[\sum_{i=1}^{[Nu]} (\hat{\boldsymbol{\xi}}_i(u) - \mu_{1:n}(\hat{\boldsymbol{\xi}}(u)))\right]^T \widehat{\boldsymbol{\Sigma}}_{LR}^{-1}(\hat{\boldsymbol{\xi}}(u)) \left[\sum_{i=1}^{[Nu]} (\hat{\boldsymbol{\xi}}_i(u) - \mu_{1:n}(\hat{\boldsymbol{\xi}}(u)))\right] \mathrm{d}u. \tag{26}$$

Following in Hörmann and Kokoszka (2010) and (Chap. 14, Horváth and Kokoszka 2012) we get that under the null hypothesis $H_{N,d}^* \xrightarrow{d} \int_0^1 \sum_{l=1}^{d} B_l^2(u)\mathrm{d}u$, where $B_l(u)s$ are independent stranded Brownian bridges on $[0, 1]$ for $l = 1 \ldots d$. The consistency of the proposed test statistic $H_{N:d}^*$ can be ensured from Theorem 3.2 by Aston and Kirch (2012) and following the Lemma 1 of this article. In Sect. 7.2, we compare the performance of $H_{N,d}^*$ and $T_{N,d}$, the proposed test statistic by Hörmann and Kokoszka (2010). $T_{N,d}$ can be expressed with our notation as

$$T_{N,d} = \frac{1}{N}\int_0^1 \left[\sum_{i=1}^{[Nu]}(\hat{\boldsymbol{\xi}}_i(1) - \mu_{1:n}(\hat{\boldsymbol{\xi}}(1)))\right]^T \widehat{\boldsymbol{\Sigma}}_{LR}^{-1}(\hat{\boldsymbol{\xi}}(1))$$
$$\left[\sum_{i=1}^{[Nu]}(\hat{\boldsymbol{\xi}}_i(1) - \mu_{1:n}(\hat{\boldsymbol{\xi}}(1)))\right] \mathrm{d}u. \tag{27}$$

7 Simulation studies

7.1 Simulation in independent set-up

In this section, we report a summary of the extensive simulation studies that we have conducted for moderate and large sample sizes. As proposed in Sect. 3, we reject the null hypothesis when the observed value of $H_{N,d}$ exceeds the corresponding critical value $K_d(\alpha)$. The critical values are available in Table 1 of Berkes et al. (2009). Without loss of generality, initial mean function is considered to be zero. For the first set of simulation studies, the samples are generated from the standard Brownian motion (BM) over the interval [0, 1] and a drift of amount t and $\sin(t)$ are considered after the presumed locations of change point. The same is done for the standard Brownian bridge over [0, 1] and the mean shift after the change point is considered to be a quadratic function $0.8t(1-t)$. To generate a sample from each of such Gaussian processes 1000 equidistant grid points are used. 750 B-spline basis functions are used to convert the grid data to functional data. As noted by Berkes et al. (2009) and Aue et al. (2009) we also observe that the power of the test decreases as we increase the number of principal components d.

First 3(=d) eigenfunctions, which are sufficient to explain at least 85% of variation for each $u \in (0, 1]$, are used to execute the testing procedures. We choose in particular 85% of the total variation so that the simulation findings can be compared with Berkes et al. (2009).

For a pre-decided sample size and a specific change point, the entire process is replicated 10000 times to assess the power of the test. The considered sample sizes (N) are 50, 100, 150, 200, 300, 500. For any particular sample size, different possible locations of change points (k^*) are chosen, to cover a wide range, which are summarized in Table 1 and Figs. 1, 2. For all practical purposes, we use the complete data together for computing the estimated covariance kernel when $k = 1$ or $k = N - 1$, otherwise as proposed in Eq. (11).

7.1.1 Small sample bias correction

For small sample size (less than or equals to 100, say) we observe some fluctuations in the empirical size of the proposed test based on $H_{N,d}$. To overcome this instability, we propose a bias correction which helps us to get empirical size reasonably close to 0.05. Under the null, it is easy to observe that

$$E[\widehat{c}_u(t,s)] = \left(1 - \frac{2}{N}\right) c(t,s). \tag{28}$$

Table 1 Power comparison of two tests with test statistics $S_{N,d}$ and $H_{N,d}$ for different k^*

$N = 100, d = 3$	BM, BM $+t$		BM, BM $+\sin(t)$		BB,BB $+0.8(1-t)t$	
k^*	$S_{N,d}$	$H_{N,d}$	$S_{N,d}$	$H_{N,d}$	$S_{N,d}$	$H_{N,d}$
0	4.6†	5.5	4.6†	5.5	4.6†	5.0
15	36.3	39.9	30.9	35.4	11.2	12.9
20	57.3	62.0	44.6	49.5	15.3	16.5
25	72.0	75.6	61.2	64.7	19.5	21.3
35	92.9	94.2	80.1	83.4	28.0	31.8
50	94.9†	95.8	88.0†	90.1	34.7	37.4
65	91.0	92.9	81.5	83.7	31.1	33.9
75	74.3	78.1	59.0	64.4	21.9	23.8
80	58.8	64.3	46.1	50.2	13.7	16.1
85	36.4	40.1	27.8	33.0	12.9	14.1

† The values are reported from the tables provided by Berkes et al. (2009, Table 4)

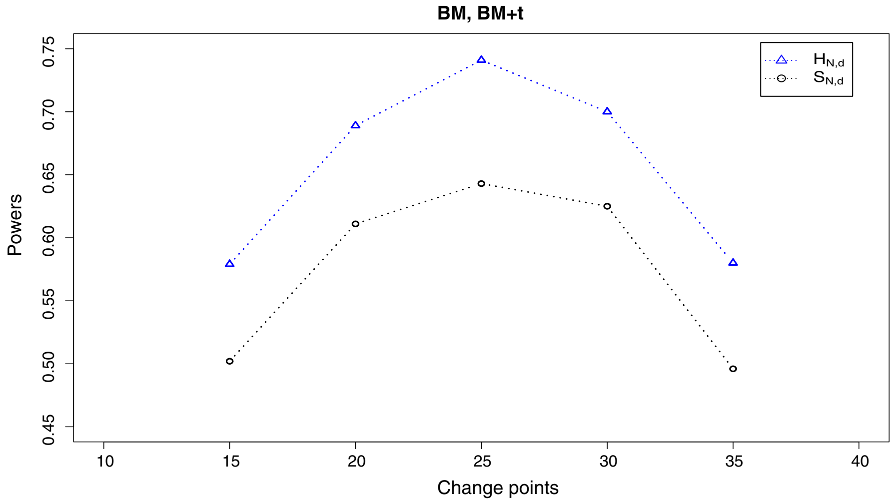

Fig. 1 Power comparison of $H_{N,d}$ and $S_{N,d}$ for $N = 50$ and $d = 3$ with $\Delta(t) = t$

So we suggest to multiply the correction factor with $(1 - 2/N)^{-1}$ with $\widehat{c}_u(t, s)$ to obtain the satisfactory results. Indeed for the large sample the effect of the correction factor vanishes automatically and it hardy matters whether we use it or not.

7.1.2 Simulation findings

In all of the cases, we find that the power curves for the proposed test based on $H_{N,d}$ strictly dominates that of the $S_{N,d}$ proposed by Berkes et al. (2009). For large sample (200 and above, say), the two power curves get very close to each other. But for small sample, we observe a remarkable gap between these two. In particular, we provide the

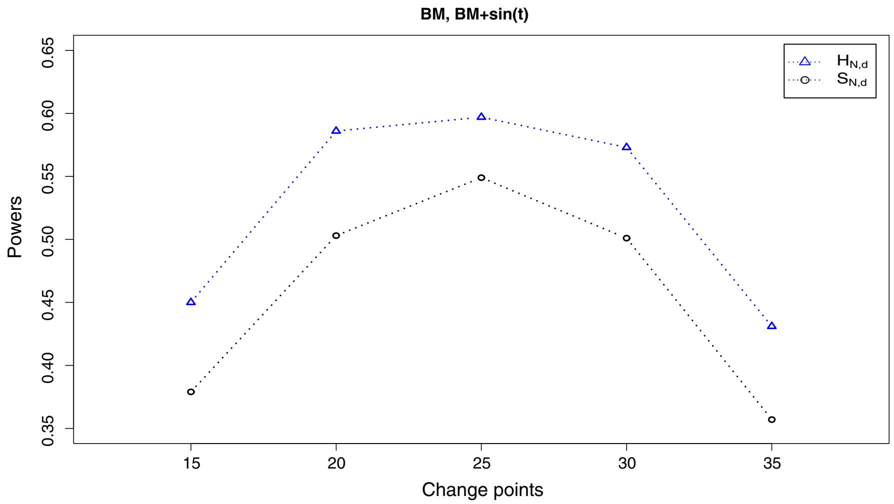

Fig. 2 Power comparison of $H_{N,d}$ and $S_{N,d}$ for $N = 50$ and $d = 3$ with $\Delta(t) = \sin(t)$

details of power for $N = 100$ and $d = 3$ at different point of change points starting from 15 to 85 in the Table 1. We add two different functions, namely t, $\sin t$ with the mean of the Brownian motion on [0, 1] and add $0.8t(1-t)$ with the mean of the standard Brownian bridge. The Fig. 1 and the Fig. 2 show the powers of two methods for sample size $50(=N)$ at different points of changes, with the mean shift t and $\sin t$ to the standard Brownian motion. It can be clearly observed that even for the small sample size our method is outperforming the method of Berkes et al. (2009) with a much larger difference. We also have done simulations for the shift functions, e. g. t^2, $\sqrt{t}$, $\exp(t)$, $\cos(t)$, being added to the mean of Brownian motion and Brownian bridge, and in all cases similar results are obtained. This finding is quite intuitive because both tests are asymptotic tests (converging to the same asymptotic distribution) and the bias in the newly proposed estimator of the covariance kernel under the alternative is smaller than that in the usual estimator of the covariance kernel used elsewhere. This satisfies the desirable quality of a better asymptotic test. We also observe quite good performance of the test statistic when the location of change point is $\leq N/4$ and $\geq 3N/4$.

7.2 Simulation in dependent set-up

We consider functional autoregressive process $\{X_i\}$ of order one according to the assumption A4, where $Y_i = \Psi(Y_{i-1}) + \epsilon_i$. Here ϵ_is are independent Brownian bridges on [0, 1]. Under the alternative hypothesis, we assume

$$X_i(t) = \begin{cases} \mu_1(t) + Y_i(t), & 1 \leq i \leq k^* \\ \mu_2(t) + Y_i(t), & k^* < i < N, \end{cases} \tag{29}$$

Table 2 Comparing $H_N^*(d)$ and $T_N(d)$ for AR(1) process

$N = 200$, BB, BB $+\ 0.8t(1-t)$, $q = 4$, $d = 3$					
k^*	0	25	50	75	100
$T_N(d)$	3.7[†]	11.2	49.5	76.9	83.1
$H_N^*(d)$	6.2	21.8	67.2	89.1	92.1

[†] The value is reported from the tables provided by Hörmann and Kokoszka (2010)

Table 3 Comparing $H_N^*(d)$ and $T_N(d)$ for AR(1) process with $\Delta(t) = t$

	$N = 100$, BM, BM $+\ t$, $q = 3$, $d = 3$								
k^*	0	15	20	25	35	50	65	75	80	85
$T_N(d)$	2.1[†]	25.6	50.6	75.1	97.4	98.7	96.5	81.1	52.5	27.4
$H_N^*(d)$	5.6	66.1	88.4	97.7	99.9	99.9	99.7	97.3	88.7	67.5

[†] The value is reported from the tables provided by (Horváth and Kokoszka 2012 Chap. 16)

where $\Delta(t) = \mu_2(t) - \mu_1(t) = 0.8t(1-t)$, whereas kernel of Ψ is $\psi(t,s) = \varphi \cdot (2 - (2t-1)^2 - (2s-1)^2)$ and the constant φ is so chosen that $||\Psi|| = 0.6$. Uniform 1000 grids in [0, 1] are used to generate the data, and 750 B-spline basis functions are used to convert the data into functional observations. We use the test statistic $H_N^*(d)$ as the proposed test statistic with $q = 4$ for 200(=N) sample size and

$$\varpi_q(j) = \begin{cases} 1 - |j|/(1+q) & \text{if } |j| \le q \\ 0 & \text{otherwise.} \end{cases} \tag{30}$$

Here we adopt the formula $q \approx 1.1447(aN)^{1/3}$ where $a = \frac{4||\Psi||^2}{(1+||\Psi||)^4}$ from Hörmann and Kokoszka (2010). With 10000 iteration we have empirical level 6.2% considering the cut-off value 1.0031 form Berkes et al. (2009) for $d = 3$ explaining 85% of the total variation. With repeated simulations, we found that this empirical level is stable and very close to the nominal level 5%. We compare the performance of proposed test statistic $H_{N,d}^*$ and $T_{N,d}$, the suggested one by Hörmann and Kokoszka (2010) and the powers are reported in Table 2 for different values of k^*.

We also conduct another set of simulation with sample size $N = 100$. As defined above we continue with AR(1) model but $\epsilon_i s$ being independent Brownian motions on [0, 1] and with Gaussian kernel $\psi(t,s) = \phi \exp\{(t^2 + s^2)/2\}$. for the operator Ψ. The constant ϕ is so chosen that $||\Psi|| = 0.5$. We perform two simulation studies one with $\Delta(t) = t$ and another with $\Delta(t) = \sin(t)$. The corresponding power comparisons are provided in Tables 3 and 4, respectively. In both of the cases, 1000 uniform grids are considered on [0, 1] and 750 B-spline basis functions are used to get smooth functions. In all of the above simulations, we find that the proposed $H_{N,d}^*$ out performs $T_{N,d}$ in terms of the power. The constants φ and ϕ are chosen to match the specification of Horváth and Kokoszka (2012) in Chap. 16, for the comparison purpose.

Table 4 Comparing $H_N^*(d)$ and $T_N(d)$ for AR(1) process with $\Delta(t) = \sin(t)$

	$N = 100, BM, BM + \sin(t), q = 3, d = 3$								
k^*	0	15	20	25	35	50	65	75	80	85
$T_N(d)$	2.1[†]	18.6	40.7	62.1	88.9	97.4	91.6	60.9	42.4	22.2
$H_N^*(d)$	5.6	52.2	79.1	91.8	98.8	99.9	99.2	91.6	80.0	53.3

[†] The value is reported from the tables provided by (Horváth and Kokoszka 2012, Chap. 16)

8 Real data analysis

8.1 Independent set-up

The findings of real data analysis to show the performance of proposed test are demonstrated in this section. Two temperature data have been analysed. One data consist of average daily temperatures of central England for 228 years, from 1780 to 2007. The data have been taken from the website of the British Atmospheric Data Centre. The other data, taken from Carbon Dioxide Information Analysis Center, consist of monthly global average anomaly of the temperatures from 1850 to 2012. Thus, these two data sets can be viewed as 228 curves with 365 measurements on each curve and 163 curves with 12 measurements on each curve, respectively. These two data sets are converted to functional data using 12 B-spline basis functions and 8 B-spline basis functions, respectively. Now we discuss the performance of the test statistics on these two temperature data sets individually.

To use the proposed test statistic, developed in the Sect. 5, for temperature data of the central England we assume independence of the data and use first $8(=d)$ eigenfunctions explaining about 85% of the total variability. Given the test indicates a change, the change point is estimated by calculating $\widehat{\theta}_N$ as described in the Lemma 1. Thereafter dividing the data set into two parts, the procedure is repeated for each part until the test fails to reject the null hypothesis. The outcome of our method when used on these data is provided in Table 5. It can be seen that the change points detected by our method and by the method of Berkes et al. (2009) are very much adjacent. Both of the methods have detected 1850 and 1926 as possible change points. In case of other years of change point, it is observed that the timings are very close, for example our method has detected a change in 1810, whereas Berkes et al. (2009) has detected a change in 1808 and in the recent years our method has detected a change in 1989 and Berkes et al. (2009) has detected 1993 as a possible change point. Overall, it is important to note that both of these methods have detected four change points in the given data. We have plotted the mean function for each of the different segments in the Fig. 3, which clearly shows an upward trend in the mean temperature over the said periods. The mean curves of different time segments are very similar to that of Berkes et al. (2009) which make sure the little observed difference in change points among two methods in this particular real data are not major. Table 5 also shows the p values corresponding to the observed value of the statistic for both of the methods. From the

Table 5 Comparisons of the performance of $S_{N,d}$ and $H_{N,d}$ for UK temperature data

Performance of $S^{\dagger}_{N,d}$				Performance of $H_{N,d}$			
Year Segment	Observed $S_{N,d}$	Obtained p value	Estimated Change point	Year Segment	Observed $H_{N,d}$	Obtained p value	Estimated Change point
1780–2007	8.020593	0.00000	1926	1780–2007	9.820036	0.00000	1926
1780–1925	3.252796	0.00088	1808	1780–1926	3.764348	0.00011	1850
1808–1925	2.351132	0.02322	1850	1780–1850	2.403308	0.01900	1810
1926–2007	2.311151	0.02643	1993	1927–2007	2.649414	0.00797	1989

† The values are reported from the tables provided by Berkes et al. (2009, Table 4)

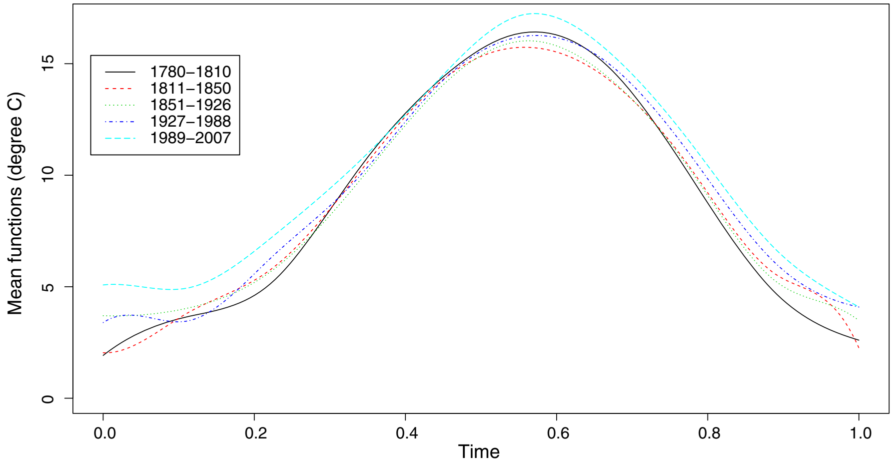

Fig. 3 Segment wise mean functions of central England temperature data

Table 6 Change points for average anomaly global temperature data

Performance of $H_{N,d}$			
Year segment	Observed $H^{*}_{N,d}$	Obtained p value	Estimated change point
1850–2012	5.964259	0.00000	1928
1939–2012	4.913237	0.00000	1979
1980–2012	4.184527	0.00000	1996

p values, it is noted that the p values of proposed test are much smaller than the p values of the existing method showing the greater power of our test.

8.2 Weakly dependent set-up

In this section, we apply the methodology developed in Sect. 6 for weakly dependent data to the monthly average anomaly of the global temperature data of 163 years. Here

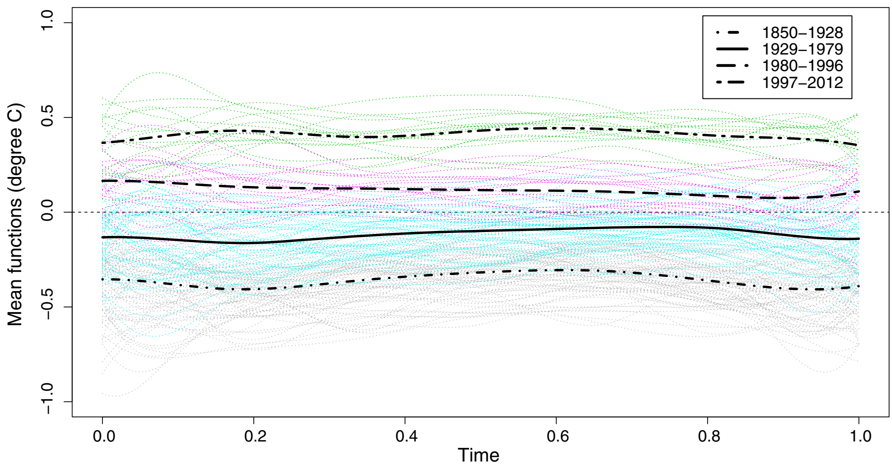

Fig. 4 Segment wise mean functions of average anomaly of global temperature data

first 2 ($=d$) eigenfunctions are used which explains at least about 85% variability of the total variation for each $u \in (0, 1]$. For estimating long-run covariance, see Eq. (25), we choose $q = 3$ following the method suggested by Hörmann and Kokoszka (2010). For detecting segment wise changes, we apply the same procedure as done in previous data set. Table 6 shows the outcomes of the test. The functional data representation of the complete data and segment wise mean functions are shown in Fig. 4 which reflects the prominent changes around the mentioned period of years. From the analysis of this data set, we clearly observe that the global temperature is changing (more specifically increasing) significantly over the period of time.

9 Discussions and conclusions

In this paper, we have proposed a new test for testing the existence of a change point in a given sequence of independent as well as weakly dependent functional data. It is shown that the proposed test enjoys many desirable properties of a test, namely it is proven that the null distribution of test is asymptomatically pivotal (a functional of the sum of squares of Brownian bridges), and proposed test is consistent as the sample sizes increases to infinity. While developing the test statistic, we have proposed an alternative estimator of the covariance kernel, which is a consistent estimator of the true covariance kernel under the null hypothesis for both the independent and the dependent data. Moreover, it has a lesser bias than the existing usual estimate of the covariance kernel under the alternative hypothesis. In fact, it is successfully shown that even under the alternative hypothesis, if the data are divided at the true point of change, then our estimator has zero asymptotic bias, whereas the existing estimate of the covariance kernel, mostly used in the literature of change point analysis in functional data, has a constant asymptotic bias. We have demonstrated that the smaller bias in the proposed covariance estimator leads to improved power of the tests detecting the presence of change point in a given sequence of functional data through the extensive simulation studies. Especially when the data size is not very big, then our method outperforms the existing one with a great margin. We applied our method in two real data for illustration purpose. First data set is the central England temperature which is also used in Berkes et al. (2009), and the other one is the global temperature data. For the first data, we have assumed independent set-up so that the outcomes could be compared with that of Berkes et al. (2009). It is observed that the findings of both the methods are very close. Importantly both the methods suggest four changes in the mean of the central England temperature. The second data, global monthly temperature anomaly data from 1850 to 2012, are analysed assuming weakly dependent set-up. It is found that there exist three change points around 1928, 1979 and 1996, which clearly shows that in the last three decades the temperature has changed significantly over the past. We strongly believe that the analysis of global temperature in terms of finding change points will help the scientists working on the global temperature. In nutshell, we evince that the proposed method has asymptotic null pivotal distribution with greater power than the existing method for testing the presence of change in a sequence of independent and weakly dependent functional data. So the proposed methodologies can be used in practice with more confidence. We conclude our paper with remarks on multiple change points in next paragraph.

Remarks on multiple change points: There are possibilities of multiple change points in practice, and we have provided an working idea for detecting multiple change points segment wise in real data following Berkes et al. (2009). However, the proof of convergence of the estimator of the covariance kernel will be technically very challenging and that will obfuscate the main idea of the paper (see also Berkes et al. 2009). For example, it is assume that there are two change points in the sequence and then we can split the data in three parts as $X_1, \ldots, X_{k_1}$; $X_{k_1+1}, \ldots, X_{k_2}$ and $X_{k_2+1}, \ldots, X_N$, choosing $0 < u < v \leq 1$, where $[Nu] = k_1$ and $[Nv] = k_2$. Then we can estimate the covariance kernel in a similar fashion. After computing the scores along the direction of the estimated principle eigen functions, we can get the test statistic using CUSUM process as done in this paper. A similar approach can be found for the finite dimensional data in the paper by Shao and Zhang (2010). Another alternative approach can be suggested. Assuming sample size $N = 2K$ is even, we can define $Z_i = Y_{2i} - Y_{2i-1}$ for $i = 1, 2, \ldots K$. Now we can estimate estimate the twice of the covariance kernel, i.e. $2c(t, s)$ using $\{Z_1, Z_2, \ldots Z_K\}$ as a whole following the approach of Berkes et al. (2009). Then the estimation of the covariance kernel and its eigenelements is trivial. But the advantage of this method is for single or multiple change points, depending on its odd or even locations, the estimator of the covariance kernel will be unbiased or asymptomatically unbiased. On the other hand, reduction of sample size will reduce the rate of convergence of the corresponding estimators. Rest of the testing process can be similar to that of Berkes et al. (2009). Our experience supports that the suggested method works well for univariate independent Gaussian sequence.

Acknowledgements The authors are thankful to the British Atmospheric Data Centre and Carbon Dioxide Information Analysis Center for real data. The Daily Central England Temperature data have been taken from the Hadley Centre for Climate Prediction and Research (2007), and monthly global average anomaly of temperatures is taken from Jones et al. (2013). The authors also thank the anonymous referee for his insightful comments and suggestions which helped to make a significant improvement of the manuscript.

Appendix

Proof of the Theorem 1

Define $\widehat{\mu}_k(t) = \frac{1}{k}\sum_{i=1}^{k} X_i(t)$ and $\widetilde{\mu}_k(t) = \frac{1}{N-k}\sum_{i=k+1}^{N} X_i(t)$ for some $k = [Nu]$ and $k^* = [N\theta]$ to express the estimated covariance kernel as

$$\widehat{c}_u(t,s) = \frac{1}{N}\left[\sum_{i=1}^{k}\{X_i(t)-\widehat{\mu}_k(t)\}\{X_i(s)-\widehat{\mu}_k(s)\} + \sum_{i=k+1}^{N}\{X_i(t)-\widetilde{\mu}_k(t)\}\{X_i(s)-\widetilde{\mu}_k(s)\}\right].$$

It immediately gives

$$\begin{aligned}\widehat{c}_u(t,s) = &\frac{1}{N}\sum_{i=1}^{N}\{X_i(t)-\widehat{\mu}_N(t)\}\{X_i(s)-\widehat{\mu}_N(s)\} - \frac{k}{N}\{\widehat{\mu}_k(t)-\widehat{\mu}_N(t)\}\{\widehat{\mu}_k(s)-\widehat{\mu}_N(s)\}\\ &-\frac{k}{N}\{\widetilde{\mu}_k(t)-\widehat{\mu}_N(t)\}\{\widetilde{\mu}_k(s)-\widehat{\mu}_N(s)\}.\end{aligned}$$

For $k \le k^*$, note that

$$\widehat{\mu}_k(t) = \widehat{\overline{Y}}_k(t) + \mu_1(t),\ \text{where } \widehat{\overline{Y}}_k(t) = \frac{1}{k}\sum_{i=1}^{k} Y_i(t),$$

$$\widetilde{\mu}_k(t) = \widetilde{\overline{Y}}_k(t) + \mu_2(t) + \left(\frac{k^*-k}{N-k}\right)\Delta(t),\ \text{where } \widetilde{\overline{Y}}_k(t) = \frac{1}{k}\sum_{i=k+1}^{N} Y_i(t),$$

and

$$\widehat{\mu}_N(t) = \widehat{\overline{Y}}_N(t) + \left(\frac{k^*}{N}\right)\mu_1(t) + \left(\frac{N-k^*}{N}\right)\mu_2(t).$$

Now observe that

$$\widehat{\mu}_k(t) - \widehat{\mu}_N(t) = \widehat{\overline{Y}}_k(t) - \widehat{\overline{Y}}_N(t) + \left(1-\frac{k^*}{N}\right)\Delta(t)$$

and

$$\widetilde{\mu}_k(t) - \widehat{\mu}_N(t) = \widetilde{\overline{Y}}_k(t) - \widehat{\overline{Y}}_N(t) - \frac{k(N-k^*)}{(N-k)N}\Delta(t)$$

to get the following deductions,

$$
\begin{aligned}
\widehat{c}_u(t,s) &= \frac{1}{N}\sum_{i=1}^{N}\{X_i(t)-\widehat{\mu}_N(t)\}\{X_i(s)-\widehat{\mu}_N(s)\} - \Delta(t)\Delta(s)\left(\frac{N-k^*}{N}\right)^2\left(\frac{k}{N-k}\right)\\
&\quad -\frac{k}{N}\left\{\widehat{\overline{Y}}_k(t)-\widehat{\overline{Y}}_N(t)\right\}\left\{\widehat{\overline{Y}}_k(s)-\widehat{\overline{Y}}_N(s)\right\}-\left(1-\frac{k}{N}\right)\left\{\widetilde{\overline{Y}}_k(t)-\widehat{\overline{Y}}_N(t)\right\}\left\{\widetilde{\overline{Y}}_k(s)-\widehat{\overline{Y}}_N(s)\right\}\\
&\quad -\frac{k}{N}\left(1-\frac{k^*}{N}\right)\left[\left\{\widehat{\overline{Y}}_k(t)-\widetilde{\overline{Y}}_k(t)\right\}\Delta(s)+\left\{\widehat{\overline{Y}}_k(s)-\widetilde{\overline{Y}}_k(s)\right\}\Delta(t)\right].
\end{aligned}
$$

Again

$$
\begin{aligned}
\widehat{c}_1(t,s) &= \frac{1}{N}\sum_{i=1}^{N}\{X_i(t)-\widehat{\mu}_N(t)\}\{X_i(s)-\widehat{\mu}_N(s)\}\\
&= \frac{1}{N}\sum_{i=1}^{N}\left\{Y_i(t)-\widehat{\overline{Y}}_N(t)\right\}\left\{Y_i(s)-\widehat{\overline{Y}}_N(s)\right\}+\frac{k^*}{N}\left(1-\frac{k^*}{N}\right)\Delta(t)\Delta(s)\\
&\quad +\frac{k^*}{N}\left(1-\frac{k^*}{N}\right)\left[\left\{\widehat{\overline{Y}}_k(t)-\widetilde{\overline{Y}}_k(t)\right\}\Delta(s)+\left\{\widehat{\overline{Y}}_k(s)-\widetilde{\overline{Y}}_k(s)\right\}\Delta(t)\right]
\end{aligned}
$$

gives

$$
\begin{aligned}
\widehat{c}_u(t,s) &= \frac{1}{N}\sum_{i=1}^{N}\left\{Y_i(t)-\widehat{\overline{Y}}_N(t)\right\}\left\{Y_i(s)-\widehat{\overline{Y}}_N(s)\right\}+\frac{k^*}{N}\left(1-\frac{k^*}{N}\right)\left[1-\frac{(N-k^*)k}{(N-k)k^*}\right]\Delta(t)\Delta(s)\\
&\quad +\left(1-\frac{k^*}{N}\right)\Delta(s)\left[\frac{k^*}{N}\left\{\widehat{\overline{Y}}_{k^*}(t)-\widetilde{\overline{Y}}_{k^*}(t)\right\}-\frac{k}{N}\left\{\widehat{\overline{Y}}_k(t)-\widetilde{\overline{Y}}_k(t)\right\}\right]\\
&\quad +\left(1-\frac{k^*}{N}\right)\Delta(t)\left[\frac{k^*}{N}\left\{\widehat{\overline{Y}}_{k^*}(s)-\widetilde{\overline{Y}}_{k^*}(s)\right\}-\frac{k}{N}\left\{\widehat{\overline{Y}}_k(s)-\widetilde{\overline{Y}}_k(s)\right\}\right]\\
&\quad -\frac{k}{N}\left(1-\frac{k}{N}\right)\left\{\widehat{\overline{Y}}_k(t)-\widetilde{\overline{Y}}_k(t)\right\}\left\{\widehat{\overline{Y}}_k(s)-\widetilde{\overline{Y}}_k(s)\right\}\\
&\equiv \frac{1}{N}\sum_{i=1}^{N}\left\{Y_i(t)-\widehat{\overline{Y}}_N(t)\right\}\left\{Y_i(s)-\widehat{\overline{Y}}_N(s)\right\}+\theta(1-\theta)\Delta(t)\Delta(s)\ f_\theta(u)\\
&\quad +r_1(t,s)+r_2(t,s)+r_3(t,s),\quad \text{say.}
\end{aligned}
\tag{31}
$$

Using the law of large numbers for independent and identically distributed Hilbert-space-valued random variables (see for example Theorem 2.4 of Bosq 2000), we obtain

$$
\int\int r_1^2(t,s)\mathrm{d}t\mathrm{d}s \xrightarrow{P} 0 \quad \text{and} \quad \int\int r_2^2(t,s)\mathrm{d}t\mathrm{d}s \xrightarrow{P} 0 \quad \text{as} \quad N\to\infty.
$$

At the same time using Theorem 5.1 of Horváth and Kokoszka (2012) we get

$$
N^2\int\int r_3^2(t,s)\mathrm{d}t\mathrm{d}s \xrightarrow{d} \left(\int \Gamma^2(t)\mathrm{d}t\right)^2,
$$

where $\{\Gamma(t) : t \in \tau\}$ is a Gaussian process with $E(\Gamma(t)) = 0$ and $E(\Gamma(t)\Gamma(s)) = c(t, s)$, which in turn implies that

$$\int\int r_3^2(t, s)\mathrm{d}t\mathrm{d}s \xrightarrow{P} 0 \text{ as } N \to \infty.$$

These help to conclude that

$$\int\int [\widehat{c}_u(t, s) - c_u(t, s)]^2\mathrm{d}t\mathrm{d}s \xrightarrow{P} 0 \text{ as } N \to \infty.$$

The similar proof holds when $k > k^*$. It is easy to see that under the null hypothesis

$$\widehat{c}_u(t, s) \xrightarrow{P} c(t, s) \quad \forall u \in (0, 1] \text{ as } N \to \infty.$$

□

Proof of Theorem 2

Under the H_0 for $k = [Nu]$

$$\widehat{c}_u(t, s) = \frac{1}{N}\sum_{i=1}^{N}\left\{Y_i(t) - \widehat{\overline{Y}}_N(t)\right\}\left\{Y_i(s) - \widehat{\overline{Y}}_N(s)\right\} - \frac{k}{N}\left(1 - \frac{k}{N}\right)\left\{\widehat{\overline{Y}}_k(t) - \widetilde{\overline{Y}}_k(t)\right\}\left\{\widehat{\overline{Y}}_k(s) - \widetilde{\overline{Y}}_k(s)\right\}. \tag{32}$$

It is enough to show that $NE \sup_{u\in(0,1]} ||\hat{C}_u - C||^2$ is bounded with a quantity which is does not involve N. Note that

$$NE \sup_{u\in(0,1]} ||\hat{C}_u - C|| \leq 2NE\int\int \times \left[\frac{1}{N}\sum_{i=1}^{N}\left\{Y_i(t) - \widehat{\overline{Y}}_N(t)\right\}\left\{Y_i(s) - \widehat{\overline{Y}}_N(s)\right\} - c(t, s)\right]^2 \mathrm{d}t\mathrm{d}s + 2NE \max_{1\leq k\leq N}\int\int\left[\left(\frac{k}{N}\right)\left(1-\frac{k}{N}\right)\left\{\widehat{\overline{Y}}_k(t)-\widetilde{\overline{Y}}_k(t)\right\}\left\{\widehat{\overline{Y}}_k(s)-\widetilde{\overline{Y}}_k(s)\right\}\right]^2 \mathrm{d}t\mathrm{d}s. \tag{33}$$

The first term is free from k and is ensured to be bounded by Hörmann and Kokoszka (2010) and the proof will be complete if we can show the the second term is bounded too. Now consider,

$$NE \max_{1\leq k\leq N}\int\int\left[\left(\frac{k}{N}\right)\left(1 - \frac{k}{N}\right)\left\{\widehat{\overline{Y}}_k(t) - \widetilde{\overline{Y}}_k(t)\right\}\left\{\widehat{\overline{Y}}_k(s) - \widetilde{\overline{Y}}_k(s)\right\}\right]^2 \mathrm{d}t\mathrm{d}s$$

$$= \frac{1}{N} E \max_{1 \le k \le N} \left(\int \frac{k(N-k)}{N} \left(\widehat{\overline{Y}}_k(t) - \widetilde{\overline{Y}}_k(t) \right)^2 \mathrm{d}t \right)^2$$

$$= \frac{1}{N} E \max_{1 \le k \le N} \left(\int \frac{k(N-k)}{N} \left(\sum_{i=1}^{N} Y_i(t) \zeta_i \right)^2 \mathrm{d}t \right)^2,$$

$$\text{where } \zeta_i = \left(\frac{1}{k} \right) \mathbf{1}_{\{i \le k\}} + \left(\frac{-1}{N-k} \right) \mathbf{1}_{\{i > k\}}$$

$$= \frac{1}{N} E \max_{1 \le k \le N} \left(\int \frac{k(N-k)}{N} \left(\sum_{i=1}^{N} \sum_{j=1}^{N} Y_i(t) Y_j(t) \zeta_i \zeta_j \right) \mathrm{d}t \right)^2$$

$$\le \frac{1}{N} E \max_{1 \le k \le N} \left(\frac{k(N-k)}{N} \left(\sum_{i=1}^{N} \sum_{j=1}^{N} ||Y_i|| ||Y_j|| \zeta_i \zeta_j \right) \right)^2$$

$$= \frac{1}{N} E \max_{1 \le k \le N} \left(\frac{\boldsymbol{\zeta}' \Sigma_{\mathbf{Y}} \boldsymbol{\zeta}}{\boldsymbol{\zeta}' \boldsymbol{\zeta}} \right)^2, \text{ where } \boldsymbol{\zeta}' = (\zeta_1, \zeta_2, \ldots \zeta_N) \text{ with } \boldsymbol{\zeta}' \boldsymbol{\zeta} = \frac{N}{k(N-k)}$$

$$\le \frac{1}{N} E \left(\sup_{\boldsymbol{\beta} \ne 0} \frac{\boldsymbol{\beta}' \Sigma_{\mathbf{Y}} \boldsymbol{\beta}}{\boldsymbol{\beta}' \boldsymbol{\beta}} \right)^2, \text{ where } \boldsymbol{\beta} \in \mathbb{R}^N$$

$$\le \frac{1}{N} E \left(\mathrm{trace}(\Sigma_{\mathbf{Y}}) \right)^2,$$

where $\Sigma_{\mathbf{Y}}$ is a non-negative definite matrix with $((\Sigma_{\mathbf{Y}}))_{ij} = ||Y_i|| ||Y_j||$

$$\le \limsup_{N \uparrow \infty} \frac{1}{N} E \left(\sum_{i=1}^{N} ||Y_i||^2 \right)^2 < \infty,$$

the last inequality follows from Hörmann and Kokoszka (2010).

This completes the proof. □

References

Antoch, J., Husková, M., Prásková, Z. (1997). Effect of dependence on statistics for determination of change. *Journal of Statistical Planning and Inference*, *60*, 291–310.

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

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*(10), 2254–2269.

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*(5), 927–946.

Bhattacharya, P. K. (1987). Maximum likelihood estimation of a change-point in the distribution of independent random variables: General multi-parameter case. *Journal of Multivariate Analysis*, *23*, 183–208.

Bongiorno, E. G., Salinelli, E., Goia, A., Vieu, P. (2014). *Contributions in infinite-dimensional statistics and related topics*. Chicago: Societa Editrice Esculapio.

Bosq, D. (2000). *Linear processes in function spaces* (vol. 149). Lecture notes in statistics theory and applications. New York: Springer.

Cobb, G. (1978). The problem of the nile: Conditional solution to a change-point problem. *Biometrika*, *65*, 243–251.

Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference*, *147*, 1–23.

Davis, R., Huang, D., Yao, Y.-C. (1995). Testing for a change in the parameter values and order of an autoregressive model. *The Annals of Statistics*, *23*, 282–304.

Ferraty, F. (2011). *Recent advances in functional data analysis and related topics*. Berlin: Springer.

Ferraty, F., Vieu, P. (2006). *Nonparametric functional data analysis*: *Theory and practice*. Chicago: Springer.

Fotopoulos, S., Jandhyala, V. (2010). Exact asymptotic distribution of change point mle for change in the mean of gaussian sequences. *Annals of Applied Statistics*, *4*, 1081–1104.

Goia, A., Vieu, P. (2016). An introduction to recent advances in high/infinite dimensional statistics. *Journal of Multivariate Analysis*, *146*, 1–6. Special Issue on Statistical models and methods for high or infinite dimensional spaces.

Gombay, E., Horváth, L. (1994). An application of the maximum likelihood test to the change-point problem. *Stochastic Processes and their Applications*, *50*, 1–17.

Hadley Centre for Climate Prediction and Research (2007). *Daily Mean, Minimum and Maximum Central England Temperature series*. National Centre for Atmospheric Science (NCAS) British Atmospheric Data Centre. http://catalogue.ceda.ac.uk/uuid/b621ef77b07d3c8e116b5b31fd5eb92b. Accessed 12 Feb 2015.

Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika*, *57*, 1–17.

Hörmann, S., Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics*, *38*(3), 1845–1884.

Horváth, L., Kokoszka, P. (2012). *Inference for functional data with applications*. Chicago: Springer.

Horváth, L., Kokoszka, P., Steinebach, J. (1999). Testing for changes in multivariate dependent observations with applications to temperature changes. *Journal of Multivariate Analysis*, *68*, 96–119.

Horváth, L., Hušková, M., Rice, G. (2013). Test of independence for functional data. *Journal of Multivariate Analysis*, *117*, 100–119.

Hsing, T., Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators* (1st ed.). Harvard: Wiley. Series in Probability and statistics.

Inclán, C., Tiao, G. C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association*, *89*(427), 913–923.

Indritz, J. (1963). *Methods in analysis*. New York: The Macmillan Co.

Jones, P., Parker, D., Osborn, T., Briffa, K. (2013). Global and hemispheric temperature anomalies land and marine instrumental records. In *Trends*: *A compendium of data on global change*. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge,Tennessee.

Kiefer, J. (1959). K-sample analogues of the kolmogorov-smirnov and cramér-v. mises tests. *The Annals of Mathematical Statistics*, *30*, 420–447.

Kirch, C., Muhsal, B., Ombao, H. (2015). Detection of changes in multivariate time series with application to eeg data. *Journal of the American Statistical Association*, *110*(511), 1197–1216.

Kokoszka, P., Leipus, R. (2000). Change-point estimation in arch models. *Bernoulli*, *6*, 513–539.

Mei, Y. (2006). Sequential change-point detection when unknown parameters are present in the pre-change distribution. *The Annals of Statistics*, *34*, 92–122.

Ramsay, J. O., Silverman, B. (2005). *Functional data analysis*. New York: Springer.

Rao, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, *33*, 659–680.

Shao, X., Zhang, X. (2010). Testing for change point in time series. *Journal of American Statistical Association*, *105*, 1228–1240.

Zhang, J.-T. (2013). *Analysis of variance for functional data*. Harvard: CRC Press.