

Dette, Holger; Quanz, Pascal

**Article — Published Version**

## Detecting relevant changes in the spatiotemporal mean function

Journal of Time Series Analysis

**Provided in Cooperation with:**

John Wiley & Sons

*Suggested Citation:* Dette, Holger; Quanz, Pascal (2023) : Detecting relevant changes in the spatiotemporal mean function, Journal of Time Series Analysis, ISSN 1467-9892, John Wiley & Sons, Ltd, Oxford, UK, Vol. 44, Iss. 5-6, pp. 505-532, <https://doi.org/10.1111/jtsa.12674>

This Version is available at:

<https://hdl.handle.net/10419/288140>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by-nc-nd/4.0/>

## DETECTING RELEVANT CHANGES IN THE SPATIOTEMPORAL MEAN FUNCTION

HOLGER DETTE  AND PASCAL QUANZ

*Fakultät für Mathematik, Ruhr-Universität Bochum, Bochum, Germany*

For a spatiotemporal process  $\{X_j(s, t) \mid s \in S, t \in T\}_{j=1, \dots, n}$ , where  $S$  denotes the set of spatial locations and  $T$  the time domain, we consider the problem of testing for a change in the sequence of mean functions  $\{\mu_j(s, t) \mid s \in S, t \in T\}_{j=1, \dots, n}$ . In contrast to most of the literature, we are not interested in arbitrarily small changes but only in changes with a norm exceeding a given threshold. Asymptotically distribution free tests are proposed, which do not require the estimation of the long-run spatiotemporal covariance structure. In particular, we consider a fully functional approach and a test based on the cumulative sum paradigm, investigate the large sample properties of the corresponding test statistics and study their finite sample properties by means of simulation study.

*Received 16 March 2022; Accepted 21 November 2022*

Keywords: Spatiotemporal process; functional data analysis; change point analysis; self-normalization; relevant hypotheses.

**JEL.** 62M10; 62R10.

### 1. INTRODUCTION

In many applications such as in the analysis of weather- or pollution-related data, measurements are obtained at different spatial locations over a certain time period at a high temporal frequency. Often there exists a natural segmentation of the time series such that it is reasonable to model at each spatial coordinate, say  $s$ , and on each segment, say  $j$ , the resulting data as a function, say  $t \rightarrow X_j(s, t)$  of the time (on the corresponding segment). Typical examples are measurements at different geographical locations. For example, within the United States Climate Reference Network (USCRN) high resolution infrared surface temperature measurements at 126 stations in the US are publicly available on the website of the NOAA U.S. government agency. Here at each location  $s$  and each day  $j$  one observes the daily temperature curve  $t \rightarrow X_j(s, t)$  (Diamond *et al.*, 2013). Other examples include yearly curves at different locations over different years such as the daily mean temperature records from 1916 to 2018 in 40 representative Canadian cities, which are publicly available from the government of Canada website. In these applications data is typically modeled in the form

$$X_j(s, t), \quad s \in S, \quad t \in T, \quad j = 1, \dots, n, \quad (1.1)$$

where  $S$  is an arbitrary finite set and  $T$  is a dense set (we will later consider an interval).

A typical question in this context is if the sequence of mean functions  $\{\mu_j(s, t) \mid s \in S, t \in T\}_{j=1, \dots, n}$ , of a spatiotemporal process  $\{X_j(s, t) \mid s \in S, t \in T\}_{j=1, \dots, n}$  has changed over a specific time period. For a fixed location this corresponds to the meanwhile classical change point problem in functional data analysis (see, e.g. Berkes *et al.*, 2009; Zhang *et al.*, 2011; Aston and Kirch, 2012; Horváth and Kokoszka, 2012; Aue *et al.*, 2018; Dette *et al.*, 2020b, among many others). On the other hand, in the spatiotemporal context as considered in model (1.1)

\* Correspondence to: Holger Dette, Fakultät für Mathematik, Ruhr-Universität Bochum, Universitätsstraße 150, 44780 Bochum, Germany.  
E-mail: holger.dette@rub.de

the change point problem is not well studied. Recently, Gromenko *et al.* (2017) proposed a test for the hypothesis of the existence of a change point in the mean function, say  $\mu_j(s, t) = \mathbb{E}[X_j(s, t)]$ , in a sequence of independent observations. They formulated the null hypothesis and alternative in the form

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_n$$

and

$$H_1 : \mu := \mu_1 = \cdots = \mu_{\lfloor n\vartheta_0 \rfloor} \neq \mu_{\lfloor n\vartheta_0 \rfloor + 1} = \cdots = \mu_n =: \mu + \delta$$

for some  $\vartheta_0 \in (0, 1)$  and combined the CUSUM principle with classical principal component analysis to construct a test for these hypotheses, which generalizes the approach of Berkes *et al.* (2009) to the spatiotemporal model (1.1). We also refer to the recent article of Zhao *et al.* (2021) who proposed change point analysis based on a composite likelihood criterion for a different spatiotemporal model.

In contrast to this literature (and also to most of the literature on change point analysis for functional data), this article takes a different look at the change point problem. Our work is motivated by the observation that in many applications one might not be interested in arbitrarily ‘small’ changes in the mean function (in fact, one often does not believe that this function is completely constant over the whole time period for all locations). As an alternative, we therefore propose to test the hypothesis of the existence of a time point  $\lfloor n\vartheta_0 \rfloor$  such that the difference, say  $\delta$ , between the mean functions before and after this point in time is *relevant*. For this purpose, we define two measures of relevance. The first one corresponds to the fully functional approach as advocated in Aue *et al.* (2018) and is based on a norm of the difference function  $\delta$ . The second one is related to the FPCA approach as considered in Berkes *et al.* (2009) and Gromenko *et al.* (2017) and uses the norm of the projection of the difference on the leading principal components. The null hypothesis is then stated in the form that the squared norm is less or equal than a given threshold  $\Delta > 0$ , that is  $H_0 : \|\delta\|^2 \leq \Delta$  (see Section 2 for details). We derive pivotal tests for both testing problems with neither requiring estimation of the long-run variance of the process  $\{X_j\}_{j=1, \dots, n}$  nor the estimation of the covariance structure of the random field  $\{X_j(s, t) \mid s \in S, t \in T\}$ .

In Section 2 we introduce the basic terminology and carefully define the two types of hypotheses considered in this article. Section 3 is devoted to the fully functional approach, while the problem of testing relevant hypotheses by projections on the functional principal components is investigated in Section 4. Finally, in Section 5 we illustrate our approach by means of a small simulation study and by the analysis of a data example.

## 2. RELEVANT CHANGES IN THE SPATIOTEMPORAL MEAN

For an arbitrary finite set  $S$  let  $L^2(S \times T)$  denote the set of all square integrable functions of the form  $f : S \times T \rightarrow \mathbb{R}$  with the common inner product

$$\langle f, g \rangle = \sum_{s \in S} \int_T f(s, t)g(s, t)dt$$

and the corresponding norm  $\|f\| = \langle f, f \rangle^{1/2}$ . From now on we put  $T = [0, 1]$ . Let  $\{X_j\}_{j \in \mathbb{Z}}$  be a sequence of square integrable random functions on  $S \times [0, 1]$ , where

$$X_j = \mu_j + \eta_j, \quad j \in \mathbb{Z}, \quad (2.1)$$

$\{\eta_j\}_{j \in \mathbb{Z}}$  is a centered error process and  $\{\mu_j\}_{j \in \mathbb{Z}}$  is a sequence of mean functions in  $L^2(S \times [0, 1])$ . We assume that the mean functions are of the form

$$\mu = \mu_1 = \cdots = \mu_{\lfloor n\vartheta_0 \rfloor}, \quad \mu_{\lfloor n\vartheta_0 \rfloor + 1} = \cdots = \mu_n = \mu + \delta,$$

where  $\mu, \delta$  denote arbitrary deterministic but unknown elements in  $L^2(S \times [0, 1])$  and  $\vartheta_0 \in (0, 1)$  is a potential (unknown) change point. The case  $\delta \equiv 0$  corresponds to the situation of no change point. As explained in Section 1, we are not interested in ‘small’ deviations before and after a potential change point and therefore consider the problem of monitoring the sequence for a relevant change in the mean function by testing the *relevant* hypotheses

$$H_0 : \|\delta\|^2 \leq \Delta \text{ versus } H_1 : \|\delta\|^2 > \Delta. \quad (2.2)$$

Here  $\Delta > 0$  is a predefined threshold, which defines the difference before and after the time point  $[n\vartheta_0]$  as relevant. Note that the case  $\Delta = 0$  corresponds to the *classical* hypotheses (see Gromenko *et al.*, 2017), but this case is not considered here. Our interest in hypotheses of the form (2.2) with  $\Delta > 0$  stems from the fact that in applications it is often questionable to look for arbitrary small deviations. Instead, it is more reasonable to focus on (scientifically) relevant deviations, which are here defined by the threshold  $\Delta$  in (2.2). The choice of this threshold depends sensitively on the specific application (see Remark 3.4 for some discussion and Dette and Wied, 2014, for an example in the context of portfolio analysis based on multivariate data). We also note that for hypotheses of the form (2.2) the choice of the norm matters, as objects might be identified as close with respect to one norm (such as the  $L^2$ -norm), while they might be considered as different with respect to another norm (such as the sup-norm). Moreover, we also mention that the null hypothesis and alternative in (2.2) can easily be changed, that is

$$H_0 : \|\delta\|^2 > \Delta \text{ versus } H_1 : \|\delta\|^2 \leq \Delta. \quad (2.3)$$

This formulation is attractive because it allows to decide for a non-relevant change (such that one can continue working under the assumption of a nearly constant mean function) at a controlled type I error. For real valued data, hypotheses of the form (2.2) and (2.3) have found considerable attention in the literature (see, e.g. the monographs of Chow and Liu, 1992; Wellek, 2010). This concept has also been used by (Liu *et al.*, 2009; Gsteiger *et al.*, 2011) and Dette *et al.* (2018) to establish the similarity of different parametric regression curves which are estimated from real valued data. In the context of functional data analysis, relevant hypotheses have been considered by (Fogarty and Small, 2014; Dette *et al.*, 2020b) and Dette *et al.* (2020c) among others. A pivotal test for the hypotheses (2.2) (and as a consequence also for the hypotheses (2.3)) will be developed in Section 3.

Recently, Gromenko *et al.* (2017) considered a different quantity to measure deviations of the difference  $\delta$  from the function  $\delta \equiv 0$ , which is closely related to functional principal component analysis. More precisely, assume that  $\mathcal{B} := \{b_1, b_2, \dots\}$  is a basis in  $L^2(S \times [0, 1])$  such that the linear span of  $\mathcal{B}$  is dense in  $L^2(S \times [0, 1])$ . Then these authors proposed to test for a fixed order  $d \in \mathbb{N}$  whether the sum of the squared scores  $\sum_{k=1}^d \langle \delta, b_k \rangle^2$  vanishes. In the context of testing relevant hypotheses, we are therefore interested in testing hypotheses of the form

$$H_0 : \sum_{k=1}^d \langle \delta, b_k \rangle^2 \leq \Delta \text{ versus } H_1 : \sum_{k=1}^d \langle \delta, b_k \rangle^2 > \Delta.$$

A pivotal test for these hypotheses, where the basis functions are given by the eigenfunctions of a convex combination of the covariance kernels before and after the change point, will be developed in Section 4.

We conclude this section by presenting several assumptions, which are required to prove the results in the next and the following sections.

### Assumption 2.1.

(A1) The process  $\{X_j\}_{j \in \mathbb{Z}}$  in model (2.1) satisfies

$$X_j = \begin{cases} \mu + \eta_j^{(1)} & j \leq n\vartheta_0, \\ \mu + \delta + \eta_j^{(2)}, & j > n\vartheta_0, \end{cases}$$

where  $\{\eta_j^{(1)}\}_{j \in \mathbb{Z}}$  and  $\{\eta_j^{(2)}\}_{j \in \mathbb{Z}}$  are stationary processes in  $L^2(S \times [0, 1])$ .

(A2)  $\{\eta_j^{(1)}\}_{j \in \mathbb{Z}}$  and  $\{\eta_j^{(2)}\}_{j \in \mathbb{Z}}$  form sequences of Bernoulli-shifts, that is, there exist a measurable space  $S$ , measurable functions  $f_1, f_2 : S^\infty \rightarrow L^2(S \times [0, 1])$  and a sequence of i.i.d.,  $S$ -valued and jointly (in  $(s, t, \omega)$ ) measurable random functions  $\{\varepsilon_j\}_{j \in \mathbb{Z}} = \{\varepsilon_j(s, t, \omega)\}_{j \in \mathbb{Z}}$  such that

$$\eta_j^{(\ell)} = f_\ell(\varepsilon_j, \varepsilon_{j-1}, \dots), \quad \ell = 1, 2$$

for all  $j \in \mathbb{Z}$ .

(A3) There exists a constant  $\psi \in (0, 1)$  such that  $E\|\eta_j^{(\ell)}\|^{2+\psi} < \infty$ , for  $\ell = 1, 2$ .

(A4) The sequences  $\{\eta_j^{(1)}\}_{j \in \mathbb{Z}}$  and  $\{\eta_j^{(2)}\}_{j \in \mathbb{Z}}$  can be approximated by  $m$ -dependent sequences  $\{\eta_{j,m}^{(1)}\}_{j \in \mathbb{Z}}$  and  $\{\eta_{j,m}^{(2)}\}_{j \in \mathbb{Z}}$ , respectively, in the sense that for some  $\kappa > 2 + \psi$

$$\sum_{m=1}^{\infty} \left( E\|\eta_0^{(\ell)} - \eta_{0,m}^{(\ell)}\|^{2+\psi} \right)^{1/\kappa} < \infty, \quad \ell = 1, 2,$$

where  $\eta_{j,m}^{(\ell)}$  is defined by

$$\eta_{j,m}^{(\ell)} = f_\ell(\varepsilon_j, \dots, \varepsilon_{j-m+1}, \varepsilon_{j,m}^*), \quad \ell = 1, 2 \tag{2.4}$$

with  $\varepsilon_{j,m}^* = (\varepsilon_{j,m,j-m}^*, \varepsilon_{j,m,j-m-1}^*, \dots)$  and  $\varepsilon_{j,m,k}^*$  are i.i.d. copies of  $\varepsilon_0$  and independent of  $\{\varepsilon_j\}_{j \in \mathbb{Z}}$ .

Assumption (A1) defines two segments before and after the change point. On each segment the corresponding spatiotemporal time series is assumed to be stationary, but there might be differences (in the mean and the error process) before and after the change point. The stochastic processes  $\{\eta_j^{(1)}\}_{j \in \mathbb{Z}}$  and  $\{\eta_j^{(2)}\}_{j \in \mathbb{Z}}$  in Assumption (A1) are not independent as they are defined by the same sequence of innovations  $\{\varepsilon_j\}_{j \in \mathbb{Z}}$  (see Assumption (A2)). The dependence concept specified by condition (A4) has been used by (Aue *et al.*, 2009; Hörmann and Kokoszka, 2010; Berkes *et al.*, 2013; Horváth *et al.*, 2014) among others. We also emphasize that our assumptions are different from those in Gromenko *et al.* (2017), who considered an independent and identically distributed error process  $\{\eta_j\}_{j \in \mathbb{Z}}$ . In particular, we allow for different long-run variances before and after the change point  $[n\vartheta_0]$ . Moreover, these authors postulate separability in the spatiotemporal variance structure (in our case a long-run variance), which means that it factors into a purely spatial and a purely temporal component. This assumption simplifies the definition and the asymptotic analysis of their test statistics substantially. We will demonstrate below that, by using the concept of self-normalization, we can construct (asymptotically) pivotal test statistics for relevant hypotheses without any of these assumptions.

### 3. FULLY FUNCTIONAL DETECTION OF RELEVANT CHANGE POINTS

We first consider a fully functional approach for testing the relevant hypotheses in (2.2). As in the case of the classical hypothesis  $H_0 : \|\delta\| = 0$ , it is based on the CUSUM statistic, but it turns out that for relevant hypotheses it will be more difficult to obtain asymptotic quantiles of a corresponding test statistic. To be precise, we consider the common estimator for the unknown change point  $\vartheta_0$  (see Hariz *et al.*, 2007; Jandhyala *et al.*, 2013, among many others) defined by

$$\hat{\vartheta}_n := \frac{1}{n} \arg \max_{\lfloor n\varepsilon \rfloor + 1 \leq k \leq n - \lfloor n\varepsilon \rfloor} \frac{k(n-k)}{n^2} \left\| \frac{1}{k} \sum_{j=1}^k X_j - \frac{1}{n-k} \sum_{j=k+1}^n X_j \right\|^2, \tag{3.1}$$

where  $\varepsilon \in (0, 1/2)$  is a predefined constant. It can be shown by similar arguments as in Proposition 3.1 of Dette *et al.* (2020c) that, under Assumption 2.1, the estimator  $\hat{\vartheta}_n$  is consistent whenever  $\|\delta\|^2 > 0$  and  $\vartheta_0 \in (\varepsilon, 1 - \varepsilon)$ , that is

$$\hat{\vartheta}_n = \vartheta_0 + o_p(n^{-1/2}) \tag{3.2}$$

as  $n \rightarrow \infty$ . Next, for  $\lambda \in [0, 1]$ ,  $\vartheta \in (\varepsilon, 1 - \varepsilon)$  we define the quantity

$$D_n(\lambda, \vartheta) := \frac{1}{[n\vartheta]} \sum_{j=1}^{[\lambda[n\vartheta]]} X_j - \frac{1}{n - [n\vartheta]} \sum_{j=[n\vartheta]+1}^{[n\vartheta]+[\lambda(n-[n\vartheta])]} X_j \in L^2(S \times [0, 1]), \tag{3.3}$$

where we also use the notation  $D_n(s, t, \lambda, \vartheta)$  simultaneously to make the dependence on the spatial and temporal coordinate explicit. Note that  $D_n$  is well defined if  $\varepsilon \geq \frac{1}{n}$  and we will assume throughout this article that  $n$  is sufficiently large such that this condition is satisfied. If  $\vartheta = k/n$  and  $\lambda = 1$  the quantity  $D_n(1, k/n)$  coincides with the expression in the squared norm in (3.1). Therefore,  $D_n(1, \hat{\vartheta}_n)$  is a natural estimator of the function  $\delta$ , which defines the difference before and after the change point. Consequently, it is reasonable to reject the null hypothesis in (2.2) for large values of the statistic

$$\|D_n(1, \hat{\vartheta}_n)\|.$$

It will be shown later that  $\sqrt{n} (\|D_n(1, \hat{\vartheta}_n)\|^2 - \|\delta\|^2)$  converges weakly to a normal distribution with a complicated variance depending on a linear combination of the long-run variances of the processes  $\{\eta_j^{(1)}\}_{j \in \mathbb{Z}}$  and  $\{\eta_j^{(2)}\}_{j \in \mathbb{Z}}$ . To avoid its estimation, we will construct a pivotal statistic. Our main tool for this construction is the following result, which provides the weak convergence of the process  $\left\{ \sqrt{n} (\|D_n(\lambda, \hat{\vartheta}_n)\|^2 - \lambda^2 \|\delta\|^2) \right\}_{\lambda \in [0, 1]}$ . For its statement we denote by

$$K_\ell((s_1, t_1), (s_2, t_2)) := \sum_{h \in \mathbb{Z}} \text{Cov}(\eta_0^{(\ell)}(s_1, t_1), \eta_h^{(\ell)}(s_2, t_2))$$

the long-run covariance kernel of the process  $\{\eta_j^{(\ell)}\}_{j \in \mathbb{Z}}$  for  $\ell = 1, 2$ , which exists under Assumption 2.1 and by

$$\bar{K}((s_1, t_1), (s_2, t_2)) := \frac{1}{\vartheta_0} K_1((s_1, t_1), (s_2, t_2)) + \frac{1}{1 - \vartheta_0} K_2((s_1, t_1), (s_2, t_2)) \tag{3.4}$$

a scaled convex combination of these kernels. The following results can be proved by similar arguments as given in Dette *et al.* (2020b) and the details are omitted for the sake of brevity.

**Theorem 3.1.** If Assumption 2.1 is satisfied and  $\|\delta\| > 0$ , then

$$\sqrt{n} \left\{ \|D_n(\lambda, \vartheta_0)\|^2 - \lambda^2 \|\delta\|^2 \right\}_{\lambda \in [0, 1]} \rightsquigarrow \tau_{\delta, \vartheta_0} \{ \lambda \mathbb{B}(\lambda) \}_{\lambda \in [0, 1]}, \tag{3.5}$$

$$\sqrt{n} \left\{ \|D_n(\lambda, \hat{\vartheta}_n)\|^2 - \lambda^2 \|\delta\|^2 \right\}_{\lambda \in [0, 1]} \rightsquigarrow \tau_{\delta, \vartheta_0} \{ \lambda \mathbb{B}(\lambda) \}_{\lambda \in [0, 1]} \tag{3.6}$$

as  $n \rightarrow \infty$ , where  $\rightsquigarrow$  denotes weak convergence in  $\mathcal{C}^\infty([0, 1])$ ,  $\{\mathbb{B}(\lambda)\}_{\lambda \in [0, 1]}$  denotes a standard Brownian motion and

$$\tau_{\delta, \vartheta_0}^2 := 4 \sum_{s_1, s_2 \in S} \iint \delta(s_1, t_1) \delta(s_2, t_2) \bar{K}((s_1, t_1), (s_2, t_2)) dt_1 dt_2 \tag{3.7}$$

with  $\bar{K}$  defined in (3.4).

This result leads to a very simple and pivotal test for the relevant hypotheses in (2.2). To be precise, we define

$$\begin{aligned} \hat{\mathbb{D}}_n &:= \|D_n(1, \hat{\vartheta}_n)\|^2, \\ \hat{\mathbb{V}}_n &:= \left( \int_0^1 (\|D_n(\lambda, \hat{\vartheta}_n)\|^2 - \lambda^2 \|D_n(1, \hat{\vartheta}_n)\|^2)^2 d\nu(\lambda) \right)^{1/2}, \end{aligned} \tag{3.8}$$

where  $\hat{\vartheta}_n$  is defined as in (3.1) and  $\nu$  is a probability measure on the interval (0, 1). We propose to reject the null hypothesis in (2.2) whenever

$$\hat{\mathbb{D}}_n > \Delta + q_{1-\alpha}(\mathbb{W}) \hat{\mathbb{V}}_n, \tag{3.9}$$

where  $q_{1-\alpha}(\mathbb{W})$  is the  $(1 - \alpha)$ -quantile of distribution of the random variable

$$\mathbb{W} := \frac{\mathbb{B}(1)}{\left( \int \lambda^2 (\mathbb{B}(\lambda) - \lambda \mathbb{B}(1))^2 d\nu(\lambda) \right)^{1/2}}. \tag{3.10}$$

The following result shows that the decision rule (3.9) defines a consistent and asymptotic level  $\alpha$  test for the hypotheses (2.3).

**Theorem 3.2.** If Assumption 2.1 is satisfied,  $\Delta > 0$  and  $\vartheta_0 \in (\epsilon, 1 - \epsilon)$ , we have

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \hat{\mathbb{D}}_n > \Delta + q_{1-\alpha}(\mathbb{W}) \hat{\mathbb{V}}_n \right) = \begin{cases} 0, & \text{if } \|\delta\|^2 < \Delta, \\ \alpha, & \text{if } \|\delta\|^2 = \Delta \text{ and } \tau_{\delta, \vartheta_0}^2 > 0, \\ 1, & \text{if } \|\delta\|^2 > \Delta. \end{cases}$$

**Remark 3.3.** The distribution of the random variable  $\mathbb{W}$  in (3.10) is symmetric. To see this, note that the numerator and denominator of  $\mathbb{W}$  are independent. This follows by using the  $L^2$  representation  $\mathbb{B}(t) = tZ_0 + \sum_{k \geq 1} Z_k \frac{\sin(\pi kt)}{\pi k}$  of the Brownian motion for  $t \in [0, 1]$  and comparing  $\mathbb{B}(1)$  and  $\mathbb{B}(t) - t\mathbb{B}(1)$ . Lastly, since  $\mathbb{B}(1)$  is symmetric, the claim follows because

$$\left( -\mathbb{B}(1), \left( \int \lambda^2 (\mathbb{B}(\lambda) - \lambda \mathbb{B}(1))^2 d\nu(\lambda) \right)^{-1/2} \right) \stackrel{D}{=} \left( \mathbb{B}(1), \left( \int \lambda^2 (\mathbb{B}(\lambda) - \lambda \mathbb{B}(1))^2 d\nu(\lambda) \right)^{-1/2} \right).$$

**Remark 3.4.**

- (1) Note that the test (3.9) depends on the specification of the measure  $\nu$  on the interval (0, 1), which has to be chosen in advance by the data analyst. However, in numerical experiments it turned out that this dependence does not have a significant influence on the rejection probabilities if the support of the measures has some distance to the boundaries 0 and 1 of this interval (see Section 5 for some practical results). A heuristic explanation for this observation lies in the fact that the measure  $\nu$  appears in the definition of the statistic  $\hat{\mathbb{V}}_n$  in (3.8) and in the quantiles of the random variable  $\mathbb{W}$  in (3.10). Thus, intuitively there is a cancelation effect in the decision rule (3.9).
- (2) An important problem in applications is the choice of the threshold  $\Delta$  which is problem-specific. For this choice a careful discussion with experts from the field of application is recommended to understand in which difference they are really interested. Moreover, there are also several alternatives if this choice is difficult after these discussions. In particular, it follows from the proof of Theorem 3.1 that

$$\frac{\hat{\mathbb{D}}_n - \|\delta\|^2}{\hat{\mathbb{V}}_n} \xrightarrow{D} \mathbb{W}.$$

Consequently, an (asymptotic)  $(1 - \alpha)$  confidence interval for the squared norm  $\|\delta\|^2 \geq 0$  of the difference of the mean functions before and after the change point is given by

$$\left[0, \hat{\mathbb{D}}_n + q_{1-\alpha}(\mathbb{W})\hat{\mathbb{V}}_n\right]. \quad (3.11)$$

Similarly, if it can be ruled out that the squared norm vanishes, a two sided interval for  $\|\delta\|^2 > 0$  is given by

$$\left(\max\{0, \hat{\mathbb{D}}_n - q_{1-\alpha/2}(\mathbb{W})\hat{\mathbb{V}}_n\}, \hat{\mathbb{D}}_n + q_{1-\alpha/2}(\mathbb{W})\hat{\mathbb{V}}_n\right]. \quad (3.12)$$

- (3) It is also possible to test the relevant hypotheses in (2.2) for a finite number of thresholds  $\Delta^{(1)} < \dots < \Delta^{(L)}$  simultaneously. In particular, acceptance of the null hypothesis with the threshold  $\Delta^{(L_0)}$  also implies acceptance for all larger thresholds  $\Delta^{(L_0+1)}, \dots, \Delta^{(L)}$ . Correspondingly, rejection for a  $\Delta^{(L_0)}$  means rejection for all smaller thresholds. In this sense, evaluating the test for several thresholds is logically consistent for the user and it is possible to determine for fixed level  $\alpha$  the largest threshold such that the null hypothesis is rejected.

**Remark 3.5.** We expect that under additional assumptions the results of this article remain correct if the data is observed at discrete but sufficiently dense time points in the interval  $[0, 1]$  and the methodology provided in this article is applied to continuous (with respect to  $t$ ) curves obtained after smoothing the original data. To be a little more precise, assume, for example, that for each  $j = 1, \dots, n$  and each location  $s \in S$  we observe

$$\{X_j(s, t_v) \mid v = 1, \dots, N\}, \quad (3.13)$$

where  $t_0 := 0 < t_1 < \dots < t_{N-1} < t_N = 1$  defines a partition of the interval  $[0, 1]$  such that  $\max_{j=1}^N |t_j - t_{j-1}| = O(1/N)$  and  $N$  converges to infinity as  $n \rightarrow \infty$ . For fixed  $j$  let  $\hat{X}_j(s, t)$  denote the local linear estimator of the function  $(s, t) \rightarrow X_j(s, t)$  from the data in (3.13). Then under additional smoothness assumptions on the error processes  $t \rightarrow \eta^{(\ell)}(s, t)$  and the mean functions  $t \rightarrow \mu(s, t)$  ( $s \in S, \ell = 1, 2$ ) and for a sufficiently dense partition of the interval  $[0, 1]$  the estimation error  $\max_{s \in S} \max_{t \in [0, 1]} |X_j(s, t) - \hat{X}_j(s, t)|$  is of order  $o_p(n^{-1/2})$  (uniformly with respect to  $j$ ). Moreover, the procedures defined in this and the following section remain valid if they are applied to the smoothed data  $\{\hat{X}_j(s, t) \mid s \in S, t \in [0, 1]\}_{j=1, \dots, n}$ .

**Remark 3.6.** As pointed out by a referee, there are numerous applications where there might exist more than one change point. For Euclidean data the literature in this context investigates multiple change points by considering classical hypotheses (see, e.g. Dette *et al.*, 2020a, and the references therein), that is: one is interested in detecting arbitrarily small changes. On the other hand, the problem of testing for multiple relevant change points is challenging and has not been addressed rigorously so far. In the situation where  $0 < \vartheta_1 < \vartheta_2 < \dots < \vartheta_K < 1$  are unknown change points and

$$X_j = \mu + \delta_{\ell-1} + \eta_j^{(\ell)} \quad \text{if } \lfloor n\vartheta_{\ell-1} \rfloor + 1 \leq j \leq \lfloor n\vartheta_{\ell} \rfloor; \quad 1 \leq \ell \leq K + 1 \quad (3.14)$$

(with  $\delta_0 = 0; \vartheta_0 = 0, \vartheta_{K+1} = 1$ ) Dette *et al.* (2020c) described a test for the hypotheses

$$H_0^{L^2} : \frac{1}{K} \sum_{\ell=1}^K \|\delta_{\ell} - \delta_{\ell-1}\|^2 \leq \Delta \quad \text{versus} \quad H_1^{L^2} : \frac{1}{K} \sum_{\ell=1}^K \|\delta_{\ell} - \delta_{\ell-1}\|^2 > \Delta$$

(the cumulative sum of changes is relevant) for  $L^2([0, 1])$ -valued data. This approach can be directly extended to the situation of spatiotemporal data as considered in this article. However, for other hypotheses

such as

$$H_0^{\max} : \max_{\ell=1}^K \|\delta_\ell - \delta_{\ell-1}\|^2 \leq \Delta \quad \text{versus} \quad H_1^{\max} : \max_{\ell=1}^K \|\delta_\ell - \delta_{\ell-1}\|^2 > \Delta$$

(there exists at least one relevant change point) the situation is more complicated and will be a topic of future research.

#### 4. TESTING RELEVANT HYPOTHESES BY CUMULATIVE SUMS OF SQUARED SCORES

We address the problem of detecting relevant changes in the mean of a stationary functional time series by estimating scores. For functional data this approach has been successfully used by several authors in the context of testing *classical* hypotheses in the one-sample, two-sample and change point problem (see Benko *et al.*, 2009; Berkes *et al.*, 2009; Zhang and Shao, 2015, among others) and it has been generalized to spatiotemporal data by Gromenko *et al.* (2017). To the best authors' knowledge tests for relevant hypotheses have not been constructed using this approach.

To be precise, consider model (2.1) and note that in this scenario, it is possible that the covariance function also changes at the point  $[n\vartheta_0]$ . Therefore, we denote by  $c^{(1)}$  and  $c^{(2)}$  the covariance kernels corresponding to the samples  $X_1, \dots, X_{[n\vartheta_0]}$  and  $X_{[n\vartheta_0]+1}, \dots, X_n$  before and after the change point, respectively, and define

$$c_{\vartheta_0} := \vartheta_0 c^{(1)} + (1 - \vartheta_0) c^{(2)} \quad (4.1)$$

as a convex combination of these two kernels. We denote by  $\tau_1 \geq \tau_2 \geq \dots$  the ordered eigenvalues of the operator having covariance kernel  $c_{\vartheta_0}$  with the corresponding orthonormal eigenfunctions  $w_1, w_2, \dots$  in  $L^2(S \times [0, 1])$ . For a fixed integer  $d \in \mathbb{N}$  we are interested in testing the (relevant) hypotheses

$$H_0 : \sum_{k=1}^d \langle \delta, w_k \rangle^2 \leq \Delta \quad \text{versus} \quad H_1 : \sum_{k=1}^d \langle \delta, w_k \rangle^2 > \Delta, \quad (4.2)$$

where  $\Delta > 0$  is a predefined threshold. We will briefly comment on the choice of  $d$  in Section 5. Note that by Parseval's identity  $\|\delta\|^2 = \sum_{k=1}^{\infty} \langle \delta, w_k \rangle^2$ , and therefore - similar as for testing classical hypotheses - a test for the hypotheses (4.2) can also be used for the hypotheses (2.2). We refer to Remark 4.4 for a more detailed discussion of this approach in the context of testing relevant hypotheses. For the statements in this section we require the following assumptions.

**Assumption 4.1.** The process  $\{X_j\}_{j \in \mathbb{Z}}$  in model (2.1) satisfies conditions (A1) and (A2) of Assumption 2.1. Furthermore,  $\{X_j\}_{j \in \mathbb{Z}}$  satisfies

(A3') There exists a constant  $\psi \in (0, 1)$  such that  $\mathbb{E} \left\| \eta_j^{(\ell)} \right\|^{4+\psi} < \infty$ ,  $\ell = 1, 2$ .

(A4') The sequences  $\{\eta_j^{(1)}\}_{j \in \mathbb{Z}}$  and  $\{\eta_j^{(2)}\}_{j \in \mathbb{Z}}$  can be approximated by  $m$ -dependent sequences  $\{\eta_{j,m}^{(1)}\}_{j \in \mathbb{Z}}$  and  $\{\eta_{j,m}^{(2)}\}_{j \in \mathbb{Z}}$ , respectively, in the sense that for some  $\kappa > 4 + \psi$

$$\sum_{m=1}^{\infty} \left( \mathbb{E} \left\| \eta_0^{(\ell)} - \eta_{0,m}^{(\ell)} \right\|^{4+\psi} \right)^{1/\kappa} < \infty, \quad \ell = 1, 2,$$

where  $\eta_{j,m}^{(\ell)}$  is defined (2.4).

**Assumption 4.2.** The (ordered) eigenvalues of the covariance operator  $c_{\vartheta_0}$  in (4.1) satisfy  $\tau_1 > \dots > \tau_d > \tau_{d+1} > 0$ , where  $d \in \mathbb{N}$  is the number of scores considered in (4.2).

Recall the definition of the estimator for the change point in (3.1) and for  $f, g \in L^2(S \times [0, 1])$  we define the function  $f \otimes g \in L^2((S \times [0, 1])^2)$  as

$$(f \otimes g)((s_1, t_1), (s_2, t_2)) := f(s_1, t_1)g(s_2, t_2), \quad (s_1, t_1), (s_2, t_2) \in S \times [0, 1].$$

We consider

$$\hat{c}_{\hat{\vartheta}_n, \lambda} := \hat{\vartheta}_n \hat{c}_\lambda^{(1)} + (1 - \hat{\vartheta}_n) \hat{c}_\lambda^{(2)}, \tag{4.3}$$

as a sequential estimator of the convex combination (4.1), where

$$\begin{aligned} \hat{c}_\lambda^{(1)} &:= \frac{1}{[\lambda \lfloor n \hat{\vartheta}_n \rfloor]} \sum_{i=1}^{\lfloor \lambda \lfloor n \hat{\vartheta}_n \rfloor \rfloor} \left( X_i - \frac{1}{[\lambda \lfloor n \hat{\vartheta}_n \rfloor]} \sum_{j=1}^{\lfloor \lambda \lfloor n \hat{\vartheta}_n \rfloor \rfloor} X_j \right) \otimes \left( X_i - \frac{1}{[\lambda \lfloor n \hat{\vartheta}_n \rfloor]} \sum_{j=1}^{\lfloor \lambda \lfloor n \hat{\vartheta}_n \rfloor \rfloor} X_j \right), \\ \hat{c}_\lambda^{(2)} &:= \frac{1}{[\lambda(n - \lfloor n \hat{\vartheta}_n \rfloor)]} \sum_{i=\lfloor n \hat{\vartheta}_n \rfloor + 1}^{\lfloor n \hat{\vartheta}_n \rfloor + \lfloor \lambda(n - \lfloor n \hat{\vartheta}_n \rfloor) \rfloor} \left( X_i - \frac{1}{[\lambda(n - \lfloor n \hat{\vartheta}_n \rfloor)]} \sum_{j=\lfloor n \hat{\vartheta}_n \rfloor + 1}^{\lfloor n \hat{\vartheta}_n \rfloor + \lfloor \lambda(n - \lfloor n \hat{\vartheta}_n \rfloor) \rfloor} X_j \right) \\ &\quad \otimes \left( X_i - \frac{1}{[\lambda(n - \lfloor n \hat{\vartheta}_n \rfloor)]} \sum_{j=\lfloor n \hat{\vartheta}_n \rfloor + 1}^{\lfloor n \hat{\vartheta}_n \rfloor + \lfloor \lambda(n - \lfloor n \hat{\vartheta}_n \rfloor) \rfloor} X_j \right) \end{aligned}$$

are estimates of the covariance functions  $c^{(1)}$  and  $c^{(2)}$  before and after the change point respectively. For  $\ell = 1, 2$  we put  $\hat{c}_\lambda^{(\ell)} := 0$  if  $[\lambda \lfloor n \hat{\vartheta}_n \rfloor] = 0$ . The first result of this section shows that the statistic (4.3) is a uniformly consistent estimator for the convex combination in (4.1).

**Theorem 4.1.** If Assumption 4.1 is satisfied and  $\vartheta_0 \in (\epsilon, 1 - \epsilon)$ , we have

$$\sup_{0 \leq \lambda \leq 1} \sqrt{\lambda} \|\hat{c}_{\hat{\vartheta}_n, \lambda} - c_{\vartheta_0}\|_2 = O_P \left( \frac{\log^{2/\kappa}(n)}{\sqrt{n}} \right),$$

where  $\|\cdot\|_2$  denotes the norm induced by the inner product

$$\langle f, g \rangle_2 := \sum_{s_1, s_2 \in S} \int \int f((s_1, t_1), (s_2, t_2))g((s_1, t_1), (s_2, t_2))dt_1 dt_2 \tag{4.4}$$

on  $L^2((S \times [0, 1])^2)$ .

In the following discussion by  $\hat{w}_{1, \lambda}, \hat{w}_{2, \lambda}, \dots$  we will denote the eigenfunctions of the operator corresponding to the estimated covariance function  $\hat{c}_{\hat{\vartheta}_n, \lambda}$ . Moreover, we use the notation  $\hat{w}_{k, 1} := \hat{w}_k$ . Recall the definition of the process  $D_n(\lambda, \vartheta)$  in (3.3) and note that  $D_n(1, \hat{\vartheta}_n)$  is an estimator for the difference between the mean functions before and after the change point. Consequently, a natural estimator for the quantity  $\sum_{k=1}^d \langle \delta, w_k \rangle^2$  in (4.2) is given by the statistic

$$\sum_{k=1}^d \langle D_n(1, \hat{\vartheta}_n), \hat{w}_k \rangle^2 \tag{4.5}$$

and the null hypothesis in (4.2) will be rejected for large values of this statistic. The following result shows that the process  $\sqrt{n}\{\sum_{k=1}^d (\langle D_n(\lambda, \hat{\vartheta}_n), \hat{w}_k \rangle^2 - \lambda^2 \langle \delta, w_k \rangle^2)\}_{\lambda \in [0,1]}$  converges weakly and - as a by-product - establishes asymptotic normality of (4.5) (after appropriate normalization).

**Theorem 4.2.** If Assumption 4.1 is satisfied,  $\|\delta\|^2 > 0$  and  $\vartheta_0 \in (\varepsilon, 1 - \varepsilon)$ , then the following statements hold in  $\ell^\infty([0, 1])$  as  $n \rightarrow \infty$ .

$$\sqrt{n} \left\{ \sum_{k=1}^d (\langle D_n(\lambda, \vartheta_0), \hat{w}_{k,\lambda} \rangle^2 - \lambda^2 \langle \delta, w_k \rangle^2) \right\}_{\lambda \in [0,1]} \rightsquigarrow \sigma_{\delta, \vartheta_0} \{ \lambda \mathbb{B}(\lambda) \}_{\lambda \in [0,1]}, \tag{4.6}$$

$$\sqrt{n} \left\{ \sum_{k=1}^d (\langle D_n(\lambda, \hat{\vartheta}_n), \hat{w}_{k,\lambda} \rangle^2 - \lambda^2 \langle \delta, w_k \rangle^2) \right\}_{\lambda \in [0,1]} \rightsquigarrow \sigma_{\delta, \vartheta_0} \{ \lambda \mathbb{B}(\lambda) \}_{\lambda \in [0,1]}, \tag{4.7}$$

where

$$\begin{aligned} \sigma_{\delta, \vartheta_0}^2 := & 4 \sum_{h \in \mathbb{Z}} \left( \vartheta_0 \sum_{s_1, \dots, s_4 \in S} \int \text{Cov} \left( \eta_0^{(1)}(s_1, t_1) \eta_0^{(1)}(s_2, t_2), \eta_h^{(1)}(s_3, t_3) \eta_h^{(1)}(s_4, t_4) \right) \right. \\ & \quad \times f((s_1, t_1), (s_2, t_2)) f((s_3, t_3), (s_4, t_4)) d(t_1, t_2, t_3, t_4) \\ & - 2 \sum_{s_1, s_2, s_3 \in S} \int \text{Cov} \left( \eta_0^{(1)}(s_1, t_1) \eta_0^{(1)}(s_2, t_2), \eta_h^{(1)}(s_3, t_3) \right) f((s_1, t_1), (s_2, t_2)) w(s_3, t_3) d(t_1, t_2, t_3) \\ & + (1 - \vartheta_0) \sum_{s_1, \dots, s_4 \in S} \int \text{Cov} \left( \eta_0^{(2)}(s_1, t_1) \eta_0^{(2)}(s_2, t_2), \eta_h^{(2)}(s_3, t_3) \eta_h^{(2)}(s_4, t_4) \right) \\ & \quad \times f((s_1, t_1), (s_2, t_2)) f((s_3, t_3), (s_4, t_4)) d(t_1, t_2, t_3, t_4) \\ & - 2 \sum_{s_1, s_2, s_3 \in S} \int \text{Cov} \left( \eta_0^{(2)}(s_1, t_1) \eta_0^{(2)}(s_2, t_2), \eta_h^{(2)}(s_3, t_3) \right) f((s_1, t_1), (s_2, t_2)) w(s_3, t_3) d(t_1, t_2, t_3) \\ & \left. + 4 \sum_{s_1, s_2 \in S} \int \bar{K}((s_1, t_1), (s_2, t_2)) w(s_1, t_1) w(s_2, t_2) d(t_1, t_2) \right) \end{aligned}$$

with  $\bar{K}$  defined in (3.4) and

$$w(s_1, t_1) := \sum_{k=1}^d \langle \delta, w_k \rangle w_k(s_1, t_1), \tag{4.8}$$

$$f((s_1, t_1), (s_2, t_2)) := \sum_{k=1}^d \langle \delta, w_k \rangle w_k(s_1, t_1) \sum_{i \neq k} \frac{\langle \delta, w_i \rangle}{\tau_k - \tau_i} w_i(s_2, t_2). \tag{4.9}$$

Note that Theorem 4.2 implies that the statistic  $\sqrt{n}\sum_{k=1}^d (\langle D_n(1, \hat{\vartheta}_n), \hat{w}_k \rangle^2 - \langle \delta, w_k \rangle^2)$  converges weakly to a normal distribution with mean 0 and variance  $\sigma_{\delta, \vartheta_0}^2$ . However, the limiting variance  $\sigma_{\delta, \vartheta_0}^2$  is rather difficult to estimate and therefore we propose to proceed by self-normalization again. For this purpose, define

$$\hat{Y}_n := \left( \int_0^1 \left( \sum_{k=1}^d \langle D_n(\lambda, \hat{\vartheta}_n), \hat{w}_{k,\lambda} \rangle^2 - \lambda^2 \langle D_n(1, \hat{\vartheta}_n), \hat{w}_k \rangle^2 \right) d\nu(\lambda) \right)^{1/2}, \tag{4.10}$$

for some probability measure  $\nu$  on the interval  $(0, 1)$ . Then we propose to reject the null hypothesis in (4.2) whenever

$$\sum_{k=1}^d \langle D_n(1, \hat{\vartheta}_n), \hat{w}_k \rangle^2 > \Delta + q_{1-\alpha}(\mathbb{W}) \hat{Y}_n. \quad (4.11)$$

The next result shows that this decision rule defines a reasonable test for the hypotheses (4.2). The proof is obtained by similar arguments as given in the proof of Theorem 3.2 in Dette *et al.* (2020c) and is therefore omitted.

**Theorem 4.3.** If Assumption 4.1 is satisfied,  $\Delta > 0$  and  $\vartheta_0 \in (\varepsilon, 1 - \varepsilon)$ , we have

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \sum_{k=1}^d \langle D_n(1, \hat{\vartheta}_n), \hat{w}_k \rangle^2 > \Delta + q_{1-\alpha}(\mathbb{W}) \hat{Y}_n \right) = \begin{cases} 0, & \text{if } \sum_{k=1}^d \langle \delta, w_k \rangle^2 < \Delta, \\ \alpha, & \text{if } \sum_{k=1}^d \langle \delta, w_k \rangle^2 = \Delta \text{ and } \sigma_{\delta, \vartheta_0}^2 > 0, \\ 1, & \text{if } \sum_{k=1}^d \langle \delta, w_k \rangle^2 > \Delta. \end{cases}$$

**Remark 4.4.** Parseval's identity implies  $\|\delta\|^2 = \sum_{k=1}^{\infty} \langle \delta, w_k \rangle^2$  and therefore one can also use the decision rule (4.11) for testing the hypotheses (2.2) (the null hypothesis is rejected if (4.11) holds). As a consequence of Theorem 4.2, we obtain for the rejection probabilities of this test - provided that all the requirements stated in Theorem 4.3 are satisfied

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \sum_{k=1}^d \langle D_n(1, \hat{\vartheta}_n), \hat{w}_k \rangle^2 > \Delta + q_{1-\alpha}(\mathbb{W}) \hat{Y}_n \right) = \begin{cases} 0, & \text{if } \|\delta\|^2 < \Delta, \\ 1, & \text{if } \sum_{k=1}^d \langle \delta, w_k \rangle^2 > \Delta \end{cases}$$

and if  $\|\delta\|^2 = \Delta$  and  $\sigma_{\delta, \vartheta_0}^2 > 0$ , we have

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \sum_{k=1}^d \langle D_n(1, \hat{\vartheta}_n), \hat{w}_k \rangle^2 > \Delta + q_{1-\alpha}(\mathbb{W}) \hat{Y}_n \right) \leq \alpha.$$

This means that the decision rule (4.11) also defines a (conservative) asymptotic level  $\alpha$  test for the hypotheses (2.2). Moreover, similar as in the case of testing classical hypotheses the test is consistent whenever  $\sum_{k=1}^d \langle \delta, w_k \rangle^2 > \Delta$ .

## 5. FINITE SAMPLE PROPERTIES

We illustrate the finite sample properties of the proposed tests by means of a small simulation study and by the analysis of a data example. Throughout this section, if not mentioned otherwise, we use a uniform distribution  $\nu_{19} = \frac{1}{19} \sum_{i=1}^{19} \delta_{i/20}$  at the points  $1/19, \dots, 18/19$  for the the probability measure  $\nu$  in the pivotal statistic (here  $\delta_x$  denotes the Dirac measure at the point  $x$ ). We also demonstrate below that the tests (3.9) and (4.11) are not very sensitive with respect to the choice of this measure.

### 5.1. Simulation Study

For the illustration of the methods introduced in Section 3 (fully functional) and in Section 4 (cumulative sum of squared scores) we put  $\mu \equiv 0$  and as the difference function before and after the change point we use

$$\delta(s, t) = \sqrt{\gamma} s \cos\left(\frac{\pi}{2} t\right), \quad (s, t) \in S \times [0, 1]. \quad (5.1)$$

The parameter  $\gamma$  will be used to vary the size of the quantity  $\|\delta\|^2$  and  $\sum_{k=1}^d \langle \delta, v_k \rangle^2$  in the hypotheses (2.2) and (4.2) respectively. In all cases we consider four locations, that is  $S = \{1, 2, 3, 4\}$ , the threshold is  $\Delta = 0.15$  and the sample sizes are given by  $n = 150, 250, 500$ . The position of the change point is chosen as  $\vartheta_0 = 0.6$  and the trimming parameter  $\varepsilon$  in the change point estimator defined in (3.1) is set to 0.05. All the results presented are based on 1000 simulation runs.

For both procedures we consider different error processes  $\{\eta_j\}_{j \in \mathbb{Z}}$  in model (2.1), where we assume that the processes before and after the change point have the same distribution, that is  $\{\eta_i\}_{i \in \mathbb{Z}} \stackrel{D}{=} \{\eta_i^{(1)}\}_{i \in \mathbb{Z}} \stackrel{D}{=} \{\eta_i^{(2)}\}_{i \in \mathbb{Z}}$ . The first one consists of independent (scaled) Brownian motions, that is for  $j = 1, \dots, n$

$$\eta_j(s, t) = \frac{s}{4} \mathbb{B}_j(t) \tag{5.2}$$

(here  $\mathbb{B}_j$  denotes the standard Brownian motion on the interval  $[0, 1]$ ). Note that this process has a separable covariance function. Second, we consider a process  $\{\eta_j\}_{j \in \mathbb{Z}}$  of independent functions with non-separable covariance defined by

$$\eta_j(s, t) = \sum_{\ell=1}^{20} \frac{N_{\ell,j}}{2\pi\ell} (\sin(2\pi\ell t) + s \cdot \cos(2\pi\ell t)), \quad (s, t) \in S \times [0, 1], \tag{5.3}$$

where  $N_{\ell,j}$  denote independent standard normally distributed random variables. The third process is a functional moving average process (fMA) of order 1. More precisely, we define independent processes

$$\varepsilon_j(s, t) = \sum_{k=1}^{\infty} N_{k,j} \sqrt{\lambda_k} v_k(s, t), \quad (s, t) \in S \times [0, 1], \tag{5.4}$$

where  $N_{k,j}$  are independent standard normally distributed random variables,  $\lambda_k = (2\pi k^2)^{-1}$  and  $v_k(s, t) = \frac{\sqrt{2}}{4} \cdot s \cdot \sin(2k\pi t)$  and consider the fMA(1) process

$$\eta_j(s, t) = \varepsilon_j(s, t) + 0.7\varepsilon_{j-1}(s, t), \quad (s, t) \in S \times [0, 1] \tag{5.5}$$

(we use the first 40 terms in the expansion (5.4)). The data is generated and stored in Fourier basis representations using the R-package `fda`.

### 5.1.1. Fully Functional Detection of Relevant Changes

We begin with an investigation of the fully functional approach and display in Figure 1 the rejection probabilities of the test (3.9) for the hypotheses (2.2) (with  $\Delta = 0.15$ ) for the three different error processes introduced in the previous subsection. We observe a qualitatively similar behavior in all three cases as predicted by Theorem 3.2. The rejection probability is strictly increasing with  $\|\delta\|^2$ . It is close to the nominal level 5% if  $\|\delta\|^2 = \Delta$  and smaller (larger) than 5% if  $\|\delta\|^2 < \Delta$  ( $\|\delta\|^2 > \Delta$ ). A comparison of the three error processes shows that the best power of the test is obtained for the error process (5.3) followed by the error process (5.5), while for the error process (5.2) the test is least powerful.

This observation can be explained by the fact that the different error processes have different variability. More precisely, it follows from the proof of Theorem 3.2 that for large sample sizes the probability of rejection can be approximated by

$$P\left(\hat{\mathbb{W}} > \frac{\sqrt{n} \Delta - \|\delta\|^2}{\mathbb{V} \tau_{\delta, \vartheta_0}} + q_{1-\alpha}(\mathbb{W})\right), \tag{5.6}$$

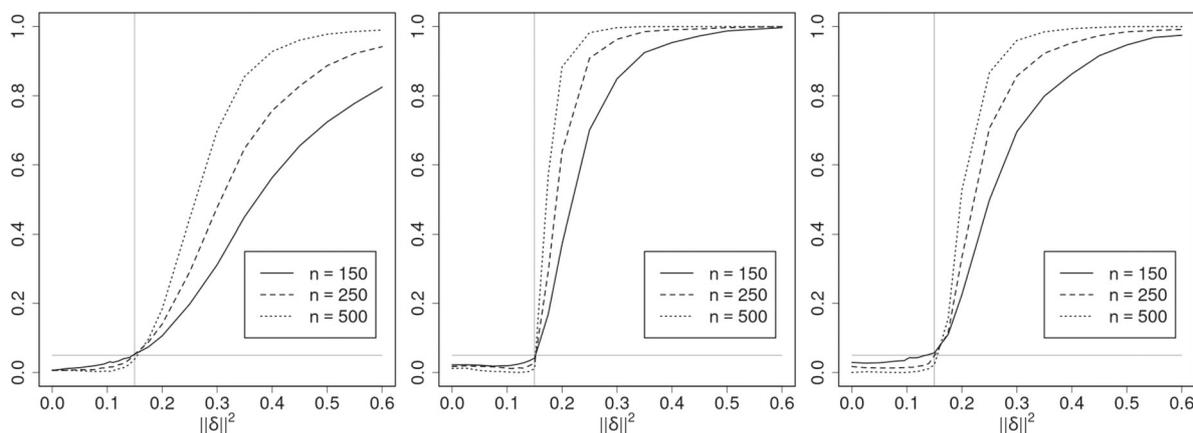


Figure 1. Empirical rejection probabilities of the test (3.9) for the hypotheses (2.2) with  $\Delta = 0.15$ . The difference  $\delta$  between the mean functions is given by (5.1) and different error processes are considered. Left panel: scaled Brownian motion in (5.2); middle panel: non-separable process in (5.3); right panel: fMA(1) process in (5.5)

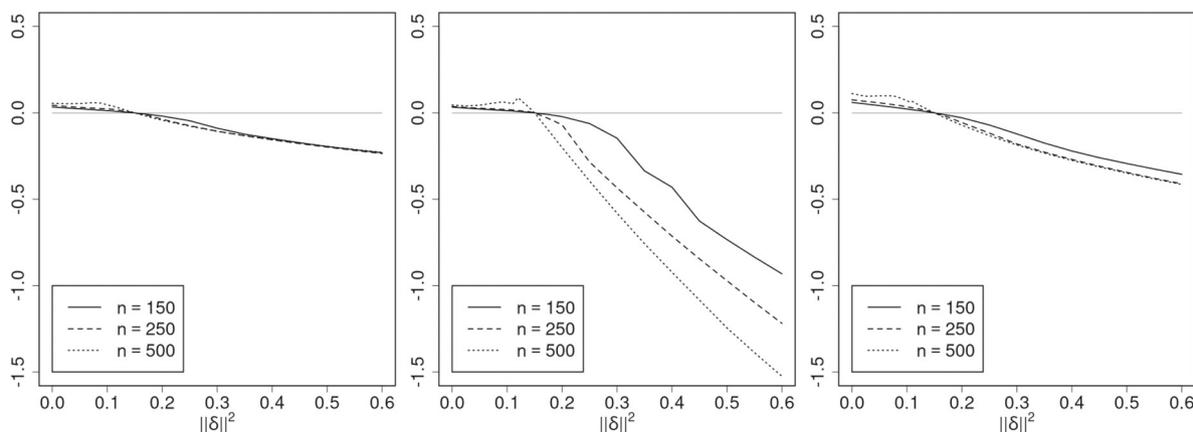


Figure 2. Size of the term  $(\Delta - \|\delta\|^2)/\tau_{\delta,\theta_0}$  as a function  $\|\delta\|^2$  in the approximation of the rejection probability ( $\Delta = 0.15$ ) in (5.6). Left panel: scaled Brownian motion in (5.2); middle panel: non-separable process in (5.3); right panel: fMA(1) process in (5.5)

where  $\tau_{\delta,\theta_0}^2$  is defined in (3.7) and  $\mathbb{V} = \left(\int \lambda^2(\mathbb{B}(\lambda) - \lambda\mathbb{B}(1))^2 d\nu(\lambda)\right)^{1/2}$  denotes a generic random variable (a functional of the Brownian motion). Thus, the power is dominated by the term  $(\Delta - \|\delta\|^2)/\tau_{\delta,\theta_0}$ , which is negative under the alternative. A smaller value of this term results in a larger power and in Figure 2 we display this quantity as a function of  $\|\delta\|^2$ . The results explain the differences in the simulated power for the three processes under consideration.

We also note that the approximation of the nominal level at the boundary of the hypotheses  $\|\delta\|^2 = \Delta$  differs in the scenarios. For small sample sizes it is more accurate for the non-separable process (5.3) compared to the two other cases. A possible explanation for this observation is given by the different accuracies of the change point estimator (3.1) in the three scenarios, which are displayed in Figure 3. We observe that the change point estimator for the error process (5.2) exhibits a larger variability than in the two other cases, while the smallest variability is obtained for the error process (5.3).

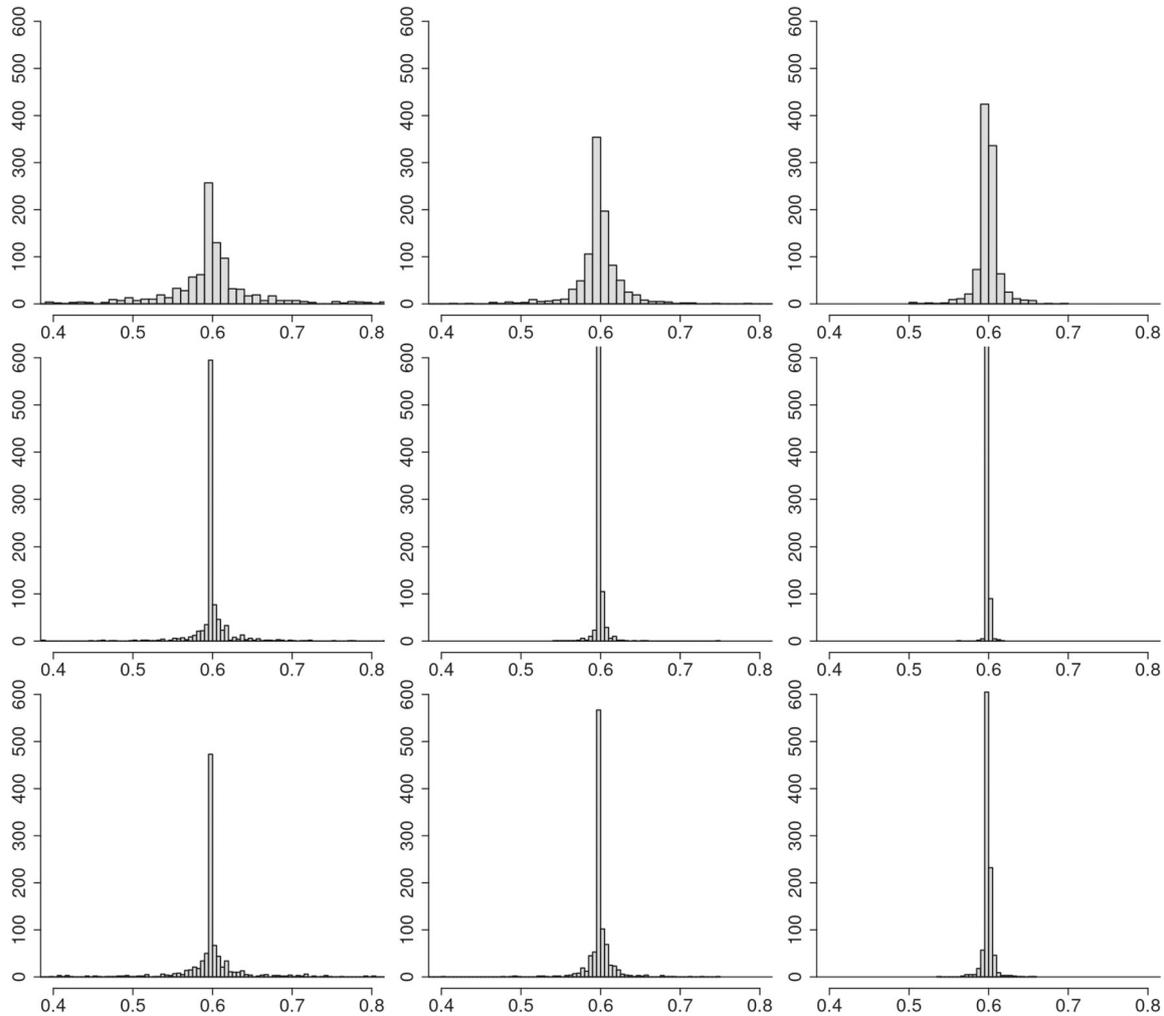


Figure 3. Histograms of the change point estimator defined in (5.2). The difference between the mean functions is given by (5.1),  $\Delta = 0.15$  and the change point is located at  $\vartheta_0 = 0.6$ . Different columns correspond to different sample sizes. Left:  $n = 250$ ; middle:  $n = 500$ ; right:  $n = 1000$ . Different rows correspond to different processes. Top: scaled Brownian motion in (5.2); middle: non-separable process in (5.3); bottom: fMA(1) process in (5.5)

Next we investigate the impact of the measure  $\nu$  in the self-normalizing statistic (3.8) on the properties of the test (3.9). Note that this measure appears in the definition of the statistic  $\hat{V}_n$  in (3.8) and in the random variable  $\mathbb{W}$  in (3.10). Thus, intuitively, there is a cancelation effect in the decision rule (3.9). For the sake of brevity, we restrict ourselves to the case of the fMA(1) error process (5.5) and display in Figure 4 the rejection probabilities of the test (3.9), where we use uniform distributions

$$\nu_k = \frac{1}{k} \sum_{i=1}^k \delta_{i/(k+1)} \tag{5.7}$$

at  $k = 4, 9$  and  $k = 19$  points as measure in the statistic (3.8). We observe a rather similar behavior for all three measures, where  $\nu_{19}$  yields a slightly better approximation of the nominal level at the boundary of the hypotheses (that is  $\|\delta\|^2 = \Delta = 0.15$ ) for the sample size  $n = 150$ .

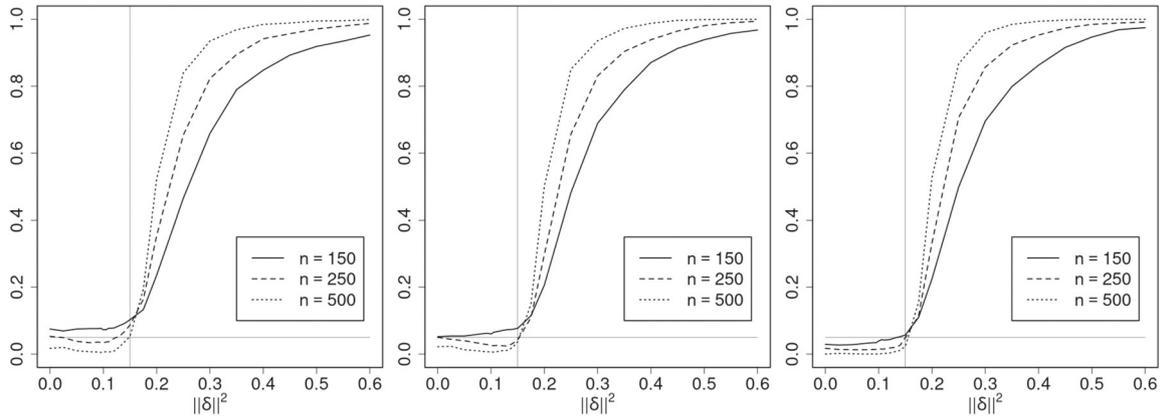


Figure 4. Empirical rejection probabilities of the test (3.9) for the hypotheses (2.2) with  $\Delta = 0.15$ . The error process is given by an fMA(1) model and different measures in the statistic (3.8) are considered. Left panel:  $\nu_4$ ; middle panel:  $\nu_6$ ; right panel:  $\nu_{19}$

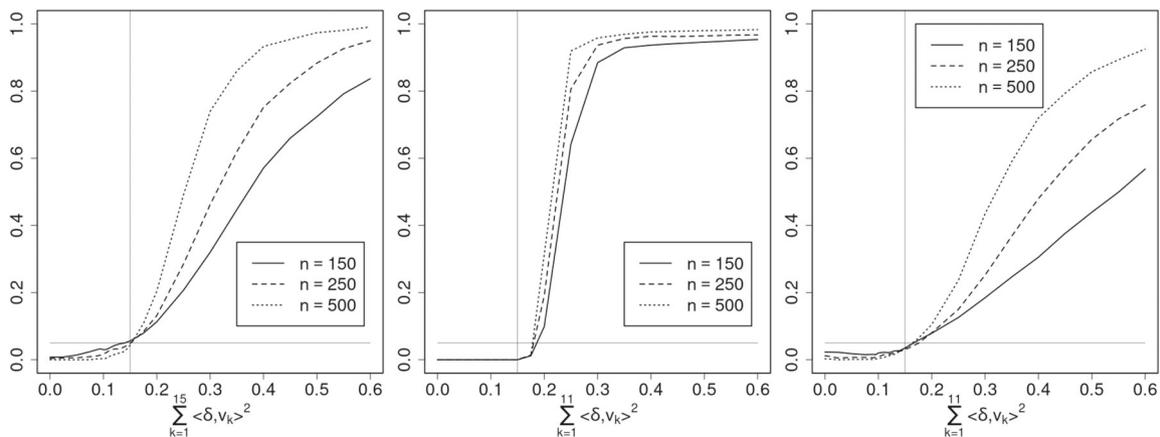


Figure 5. Empirical rejection probabilities of the test (4.11) for the hypotheses (4.2) with  $\Delta = 0.15$ . The difference  $\delta$  between the mean functions is given by (5.1) and different error processes are considered. Left panel: scaled Brownian motion ( $d = 15$ ) in (5.2); middle panel: non-separable process ( $d = 11$ ) in (5.3); right panel: fMA(1) process ( $d = 11$ ) in (5.5)

5.1.2. Relevant Changes by Cumulative Sums of Squared Scores

We briefly illustrate the finite sample properties of the test (4.11) for the hypotheses (4.2), where we use the same scenarios as before. The test requires the choice of the number of principal functional components and we choose the parameter  $d$  such that 95% of the variance in the data will be explained. This results in  $d = 15$ ,  $d = 11$  and  $d = 11$  functional principal components for the models (5.2), (5.3), and (5.5) respectively. The corresponding rejection probabilities are displayed in Figure 5 and we observe the qualitative behavior predicted by Theorem 4.3. We also observe that the test (4.11) is conservative in the case of the non-separable process (5.3), while the nominal level at the boundary of the hypotheses  $\sum_{k=1}^d \langle \delta, b_k \rangle^2 = \Delta$  is very well approximated for the scaled Brownian motion (5.2).

Finally, we investigate the impact of the measure  $\nu$  in the scaling factor (4.10), where we again restrict ourselves to the case of an fMA(1) process and the uniform distributions  $\nu_k$  in (5.7) for  $k = 4, 9$  and  $19$ . The corresponding results are shown in Figure 6 and demonstrate that the test (4.11) is not very sensitive with respect to this choice, either.

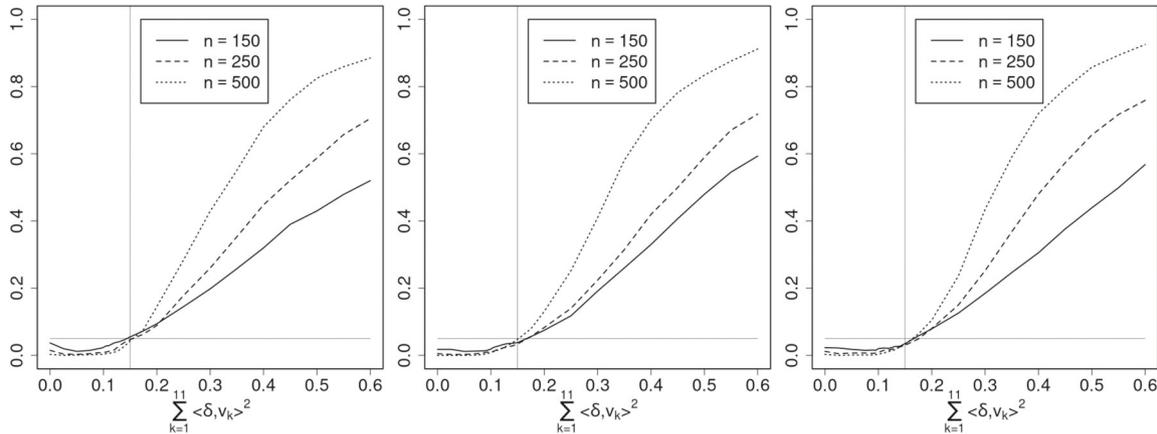


Figure 6. Empirical rejection probabilities of the test (4.11) for the hypotheses (4.2) with  $\Delta = 0.15$ . The error process is given by an fMA(1) model and different measures in the statistic (3.8) are considered. Left panel:  $\nu_4$ ; middle panel:  $\nu_9$ ; right panel:  $\nu_{19}$

Table I. Results of the test (3.9) for the Canadian weather data for different nominal level and different thresholds

$\Delta/\alpha$	10%	5%	1%
0.004	Reject	Reject	Accept
0.005	Reject	Accept	Accept
$\vdots$	$\vdots$	$\vdots$	$\vdots$
2.974	Reject	Accept	Accept
2.975	Accept	Accept	Accept

Note: The value 0.004 represents the maximal threshold in (2.3) such that the null hypothesis of no relevant change is rejected at nominal level 5%, whereas the value 2.974 represents the maximal threshold in (2.3) such that the null hypothesis is rejected at nominal level 10%.

### 5.2. Data Example

We conclude this article with an application of the two test procedures in a real data example. For this purpose, we use Canadian weather data, which consists of daily measurements at 40 representative Canadian cities. Thus, we observe yearly curves over different years (1916 – 2018) at different locations. The data can be downloaded from the government of Canada website: [https://climate.weather.gc.ca/historical\\_data/search\\_historic\\_data\\_e.html](https://climate.weather.gc.ca/historical_data/search_historic_data_e.html). The available data contains several different measurements such as maximum/minimum temperature or precipitation amount. For the sake of brevity, we focus on the average daily temperatures. Due to missing values in the reported temperature data, four stations were chosen such that a large amount of available data overlaps and only little parts had to be interpolated or removed: CALGARY INTERNATIONAL AIRPORT, ALBERTA (ID: 2205), MEDICINE HAT AIRPORT, ALBERTA (ID: 2273), INDIAN HEAD CDA, SASKATCHEWAN (ID: 2925) and OTTAWA CDA, ONTARIO (ID: 4333). In the notation of the previous sections this means  $S = \{1, 2, 3, 4\}$  and the sample size is given by  $n = 113$ , which corresponds to the period from the years 1891 to 2007 without the years 1910, 1911, 1993 and 1995. The change point estimator in (3.1) yields  $\hat{\delta}_n = 0.708$ , which approximately corresponds to the year 1969. We emphasize that this result depends on the definition of the segmentation and we have used the common yearly segmentation starting on January 1st and ending on December 31st (see also Besse *et al.*, 2000; Sharipov *et al.*, 2016; Gromenko *et al.*, 2017, who used this segmentation as well). If one uses a different yearly segmentation the results may change slightly. For example, the yearly segmentation ranging from July 1st to June 30th yields  $\hat{\delta}_n = 0.699$  as estimate for the change point corresponding to the years 1968/69.

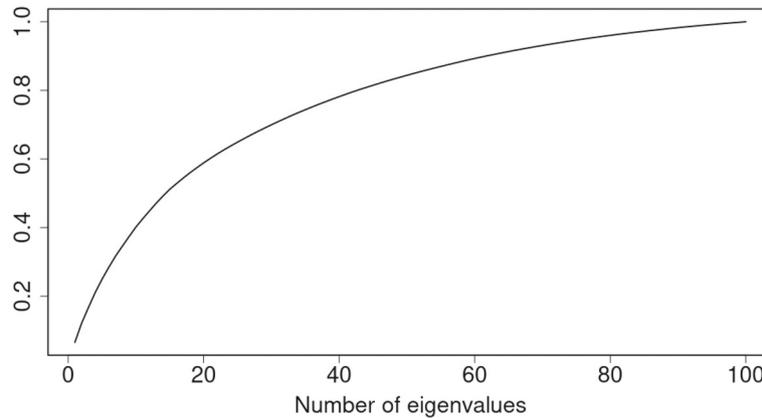


Figure 7. The function  $e$  in (5.8) for  $j = 1, \dots, 100$

Table II. Results of the test (4.11) with  $d = 51$  principal components for the Canadian weather data for different nominal level and different thresholds

$\Delta/\alpha$	10%	5%	1%
0.531	Reject	Reject	Reject
0.532	Reject	Reject	Accept
⋮	⋮	⋮	⋮
3.301	Reject	Reject	Accept
3.302	Reject	Accept	Accept
⋮	⋮	⋮	⋮
4.467	Reject	Accept	Accept
4.468	Accept	Accept	Accept

Note: The value 4.467 represents the maximal threshold in (4.2) such that the null hypothesis of no relevant change is rejected at nominal level 10%, the value 3.301 is the maximal threshold for the nominal level 5% and the value 0.531 is the maximal threshold for the nominal level 1%.

Next we apply the testing procedures introduced in Section 3 and 4 starting with the fully functional approach. The results of the test (3.9) for different thresholds and different nominal levels are given in Table I. We observe that  $\Delta = 2.974$  is the largest threshold such that the test (3.9) rejects the null hypothesis at nominal level  $\alpha = 0.1$ . Because there are four stations, this corresponds to an average effect of  $2.974/4 \approx 0.74$ . Finally, we note that the one-sided confidence interval for  $\|\delta\| \geq 0$  in (3.11) is given by  $[0, 4.397]$ , while the two-sided interval for  $\|\delta\| > 0$  in (3.12) is obtained as  $(0.067, 4.675]$ .

Lastly, we consider the test based on cumulative sums of squared scores developed in Section 4. In this case, a choice for the value of  $d$  must be made and we display in Figure 7 the ratios

$$j \rightarrow e(j) := \frac{\sum_{k=1}^j \hat{\lambda}_k}{\text{trace}(\hat{c}_{\hat{\delta}_n})}, \tag{5.8}$$

where  $\hat{\lambda}_1, \hat{\lambda}_2, \dots$  are the estimated eigenvalues of the covariance operator corresponding to the covariance function in (4.1). We observe that the eigenvalues are slowly decreasing and we choose  $d = 51$ , which results in a value of  $e(51) \approx 0.849$  of explained variance. The results of the test (4.11) for the relevant hypotheses (4.2) are shown in Table II for different values of  $\Delta$  and  $\alpha$ . We observe that the maximal threshold in (4.2) such that the null hypothesis of no relevant change is rejected at nominal level 10%, is given by 4.467. Finally, one and two-sided confidence

intervals for the quantity  $\left(\sum_{k=1}^d (\delta, w_k)^2\right)^{1/2}$  are obtained in same way as described in Remark 3.4 and are given by [0, 3.254] and (1.877, 3.396], respectively ( $d = 51$ ).

#### ACKNOWLEDGEMENTS

The authors would like to thank two unknown referees for their constructive comments on an earlier version of this article. This research has been supported in part by the German Research Foundation (DFG), project number 45723897. Open Access funding enabled and organized by Projekt DEAL.

#### DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available at [https://climate.weather.gc.ca/historical\\_data/search\\_historic\\_data\\_e.html](https://climate.weather.gc.ca/historical_data/search_historic_data_e.html).

#### REFERENCES

- Aston JA, Kirch C. 2012. Detecting and estimating changes in dependent functional data. *Journal of Multivariate Analysis* **109**:204–220.
- 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**(6B):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**(3):509–529.
- Aue A, Dette H, Rice G. 2023. Two-sample tests for relevant differences in the eigenfunctions of covariance operators. *Statistica Sinica* **33**:1–27.
- Benko M, Härdle W, Kneip A. 2009. Common functional principal components. *The Annals of Statistics* **37**(1):1–34.
- 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.
- Berkes I, Horváth L, Rice G. 2013. Weak invariance principles for sums of dependent random functions. *Stochastic Processes and their Applications* **123**(2):385–403.
- Besse PC, Cardot H, Stephenson DB. 2000. Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics* **27**(4):673–687.
- Chow S-C, Liu P-J. 1992. *Design and Analysis of Bioavailability and Bioequivalence Studies* Marcel Dekker, New York.
- Dette H, Wied T. 2014. Detecting relevant changes in time series models. *Journal of the Royal Statistical Society* **78**(2):371–394.
- Dette H, Möllenhoff K, Volgushev S, Bretz F. 2018. Equivalence of regression curves. *Journal of the American Statistical Association* **113**:711–729.
- Dette H, Eckle T, Vetter M. 2020a. Multiscale change point detection for dependent data. *Scandinavian Journal of Statistics* **47**(4):1243–1274.
- Dette H, Kokot K, Aue A. 2020b. Functional data analysis in the Banach space of continuous functions. *Annals of Statistics* **48**(2):1168–1192.
- Dette H, Kokot K, Volgushev S. 2020c. Testing relevant hypotheses in functional time series via self-normalization. *Journal of the Royal Statistical Society (B)* **82**(3):629–660.
- Diamond HJ, Karl TR, Palecki MA, Baker CB, Bell JE, Leeper RD, Easterling DR, Lawrimore JH, Meyers TP, Helfert MR, Goodge G, Thorne PW. 2013. U.S. Climate Reference Network after One Decade of Operations: Status and Assessment. *Bulletin of the American Meteorological Society* **94**(4):485–498.
- Fogarty CB, Small DS. 2014. Equivalence testing for functional data with an application to comparing pulmonary function devices. *Annals of Applied Statistics* **8**(4):2002–2026.
- Gromenko O, Kokoszka P, Reimherr M. 2017. Detection of change in the spatiotemporal mean function. *Journal of the Royal Statistical Society (B)* **79**(1):29–50.
- Gsteiger S, Bretz F, Liu W. 2011. Simultaneous confidence bands for nonlinear regression models with application to population pharmacokinetic analyses. *Journal of Biopharmaceutical Statistics* **21**(4):708–725.
- Hariz SB, Wylie JJ, Zhang Q. 2007. Optimal rate of convergence for nonparametric change-point estimators for nonstationary sequences. *The Annals of Statistics* **35**:1802–1826.
- 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* Springer-Verlag, New York.

- Horváth L, Kokoszka P, Rice G. 2014. Testing stationarity of functional time series. *Journal of Econometrics* **179**(1):66–82.
- Jandhyala V, Fotopoulos S, MacNeill I, Liu P. 2013. Inference for single and multiple change-points in time series. *Journal of Time Series Analysis* **34**(4):423–446.
- Liu W, Bretz F, Hayter AJ, Wynn HP. 2009. Assessing non-superiority, non-inferiority of equivalence when comparing two regression models over a restricted covariate region. *Biometrics* **65**(4):1279–1287.
- 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**(3):300–322.
- Wellek S. 2010. *Testing Statistical Hypotheses of Equivalence and Noninferiority* Chapman and Hall/CRC, New York.
- Zhang X, Shao X. 2015. Two sample inference for the second-order property of temporally dependent functional data. *Bernoulli* **21**(2):909–929.
- Zhang X, Shao X, Hayhoe K, Wuebbles DJ. 2011. Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electronic Journal of Statistics* **5**:1765–1796.
- Zhao Z, Ma TF, Ng WL, Yau CY (2021). A composite likelihood-based approach for change-point detection in spatio-temporal process. arXiv:1904.06340.

## APPENDIX : PROOFS

### A.1. Some Preliminary Results

We start with some preparations and present several results which are used in the proofs. We define for  $s \in S$  and  $t \in [0, 1]$

$$S_n(s, t, \lambda) := \frac{1}{n} \sum_{j=1}^{\lfloor \lambda n \rfloor} (X_j(s, t) - E[X_j(s, t)])$$

and state the following result, which can be obtained by generalizing Theorem 1.1 in Berkes *et al.* (2013) to the space  $L^2(S \times [0, 1])$ .

**Theorem A.1.** *If Assumption 2.1 is satisfied, there exists a sequence of Gaussian processes  $(\{\Gamma_n(s, t, \lambda) \mid s \in S, 0 \leq \lambda, t \leq 1\})_{n \in \mathbb{N}}$  such that*

$$\sup_{0 \leq \lambda \leq 1} \sum_{s \in S} \int (\sqrt{n} S_n(s, t, \lambda) - \Gamma_n(s, t, \lambda))^2 dt = \sup_{0 \leq \lambda \leq 1} \left\| \sqrt{n} S_n(\cdot, \cdot, \lambda) - \Gamma_n(\cdot, \cdot, \lambda) \right\|^2 = o_p(1)$$

and

$$\{\Gamma_n(s, t, \lambda) \mid s \in S, 0 \leq \lambda, t \leq 1\} \stackrel{D}{=} \{\Gamma(s, t, \lambda) \mid s \in S, 0 \leq \lambda, t \leq 1\},$$

where

$$\Gamma(s, t, \lambda) := \sum_{i=1}^{\infty} \sqrt{\lambda_i} \phi_i(s, t) W_i(\lambda),$$

$\lambda_i, \phi_i$  are the eigenvalues and eigenvectors of the covariance operator of  $X_j$  and  $\{W_i\}_{i \in \mathbb{N}}$  are independent standard Brownian motions.

Our next auxiliary result quantifies the difference between the processes  $\{D_n(\lambda, \vartheta_0)\}_{\lambda \in [0, 1]}$  and  $\{D_n(\lambda, \hat{\vartheta}_n)\}_{\lambda \in [0, 1]}$  as  $n \rightarrow \infty$ .

**Lemma A.2.** *If Assumption 2.1 is satisfied, then*

- (1)  $\sup_{0 \leq \lambda \leq 1} \|D_n(\lambda, \vartheta)\| = O_p(1)$  for  $\vartheta \in \{\vartheta_0, \hat{\vartheta}_n\}$ ,
- (2)  $\sup_{0 \leq \lambda \leq 1} \|D_n(\lambda, \hat{\vartheta}_n) - D_n(\lambda, \vartheta_0)\| = o_p(n^{-1/2})$ .

*Proof.* We can assume that  $\frac{1}{n} \leq \lambda$  because, by definition,  $D_n(\lambda, \vartheta_0) \equiv D_n(\lambda, \hat{\vartheta}_n) \equiv 0$  if  $0 \leq \lambda < \frac{1}{n}$ , and both assertions are trivially true.

For a proof of part (1) we note that with  $\kappa$  from (A4) we have

$$\sup_{\frac{1}{n} \leq \lambda \leq 1} \frac{1}{\sqrt{[\lambda n]}} \left\| \sum_{j=1}^{[\lambda n]} \eta_j^{(\ell)} \right\| = O_P(\log^{1/\kappa}(n)), \quad \ell = 1, 2, \tag{A1}$$

which follows from Lemma B.1 in the online supplementary material of Aue *et al.* (2018). Let  $\lambda \in [1/n, 1]$ . It suffices to show the assertion for  $\vartheta = \vartheta_0$  because the second part of the Lemma implies the statement for  $\vartheta = \hat{\vartheta}_n$ . Then

$$\|D_n(\lambda, \vartheta_0)\| \leq \|\tilde{D}_n(\lambda, \vartheta_0)\| + \lambda \|\delta\| + o_P(1) = \|\tilde{D}_n(\lambda, \vartheta_0)\| + O_P(1),$$

where the process  $\{\tilde{D}_n(\lambda, \vartheta)\}_{\lambda \in [0,1]}$  in  $L^2(S \times [0, 1])$  is defined by

$$\tilde{D}_n(\lambda, \vartheta) := \frac{1}{[n\vartheta]} \sum_{j=1}^{[\lambda n\vartheta]} \eta_j^{(1)} - \frac{1}{n - [n\vartheta]} \sum_{j=[n\vartheta]+1}^{[n\vartheta]+[\lambda(n-[n\vartheta])]} \eta_j^{(2)}.$$

Note that  $\tilde{D}_n(\lambda, \vartheta)$  is the centered version of  $D_n(\lambda, \vartheta)$  defined in (3.3) and that

$$D_n(\lambda, \vartheta_0) + \lambda\delta = \tilde{D}_n(\lambda, \vartheta_0) + O_P(n^{-1}) \tag{A2}$$

(note that  $\vartheta_0 \in (\varepsilon, 1 - \varepsilon)$  and that  $\varepsilon \geq \frac{1}{n}$  if  $n$  is sufficiently large). We have

$$\sup_{\frac{1}{n} \leq \lambda \leq 1} \|\tilde{D}_n(\lambda, \vartheta_0)\| \leq \sup_{\frac{1}{n} \leq \lambda \leq 1} \frac{1}{[n\vartheta_0]} \left\| \sum_{j=1}^{[\lambda n\vartheta_0]} \eta_j^{(1)} \right\| + \sup_{\frac{1}{n} \leq \lambda \leq 1} \frac{1}{n - [n\vartheta_0]} \left\| \sum_{j=[n\vartheta_0]+1}^{[n\vartheta_0]+[\lambda(n-[n\vartheta_0])]} \eta_j^{(2)} \right\|,$$

where the first term can be estimated as follows

$$\begin{aligned} \sup_{\frac{1}{n} \leq \lambda \leq 1} \frac{1}{[n\vartheta_0]} \left\| \sum_{j=1}^{[\lambda n\vartheta_0]} \eta_j^{(1)} \right\| &\leq \frac{1}{\sqrt{[n\vartheta_0]}} \sup_{\frac{1}{n} \leq \lambda \leq 1} \frac{1}{\sqrt{[\lambda n\vartheta_0]}} \left\| \sum_{j=1}^{[\lambda n\vartheta_0]} \eta_j^{(1)} \right\| \\ &\leq \frac{1}{\sqrt{[n\vartheta_0]}} \sup_{\frac{1}{n} \leq \lambda \leq 1} \frac{1}{\sqrt{[\lambda n]}} \left\| \sum_{j=1}^{[\lambda n]} \eta_j^{(1)} \right\| = O_P\left(\frac{\log^{1/\kappa}(n)}{\sqrt{n}}\right) = o_P(1). \end{aligned}$$

(here in the second inequality, we expanded the set over which we take the supremum and in the last step, we used the estimate (A1)). A similar argument provides the same rate for the second term, which proves part (1) of Lemma A.2.

For the proof of the second assertion we need the following (slightly more general) statement from the beginning of Section B.1 in Dette *et al.* (2020c), which states that the sequence of processes  $\{\Gamma_n(s, t, \lambda) \mid s \in S, 0 \leq t, \lambda \leq 1\}_{n \in \mathbb{N}}$  in Theorem A.1 satisfies

$$\sup_{\substack{\nu, \lambda \in [0,1]: \\ |\nu - \lambda| \leq \kappa_n}} \|\Gamma_n(\cdot, \cdot, \nu) - \Gamma_n(\cdot, \cdot, \lambda)\|^2 = o_P(1). \tag{A3}$$

for any and positive sequence  $(\kappa_n)_{n \in \mathbb{N}}$  with  $\kappa_n \rightarrow 0$ . By adding and subtracting  $\lambda\delta$  we have  $D_n(\lambda, \hat{\vartheta}_n) - D_n(\lambda, \vartheta_0) = \tilde{D}_n(\lambda, \hat{\vartheta}_n) - \tilde{D}_n(\lambda, \vartheta_0) + O_p(n^{-1})$ , where

$$\begin{aligned} \tilde{D}_n(\lambda, \hat{\vartheta}_n) - \tilde{D}_n(\lambda, \vartheta_0) &= \frac{1}{[n\hat{\vartheta}_n]} \sum_{j=1}^{\lfloor \lambda [n\hat{\vartheta}_n] \rfloor} \eta_j^{(1)} - \frac{1}{[n\vartheta_0]} \sum_{j=1}^{\lfloor \lambda [n\vartheta_0] \rfloor} \eta_j^{(1)} \\ &\quad + \frac{1}{n - [n\vartheta_0]} \sum_{j=[n\vartheta_0]+1}^{[n\vartheta_0] + \lfloor \lambda(n - [n\vartheta_0]) \rfloor} \eta_j^{(2)} - \frac{1}{n - [n\hat{\vartheta}_n]} \sum_{j=[n\hat{\vartheta}_n]+1}^{[n\hat{\vartheta}_n] + \lfloor \lambda(n - [n\hat{\vartheta}_n]) \rfloor} \eta_j^{(2)}. \end{aligned}$$

For the first difference we obtain by a similar argument as in the proof of part (1) that

$$\begin{aligned} &\sup_{\frac{1}{n} \leq \lambda \leq 1} \sqrt{n} \left\| \frac{1}{[n\hat{\vartheta}_n]} \sum_{j=1}^{\lfloor \lambda [n\hat{\vartheta}_n] \rfloor} \eta_j^{(1)} - \frac{1}{[n\vartheta_0]} \sum_{j=1}^{\lfloor \lambda [n\vartheta_0] \rfloor} \eta_j^{(1)} \right\| \\ &\leq \frac{n}{[n\hat{\vartheta}_n]} \sup_{\frac{1}{n} \leq \lambda \leq 1} \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor \lambda [n\hat{\vartheta}_n] \rfloor} \eta_j^{(1)} - \Gamma_n \left( \cdot, \cdot, \lambda \frac{[n\hat{\vartheta}_n]}{n} \right) \right\| \\ &\quad + \frac{n}{[n\vartheta_0]} \sup_{\frac{1}{n} \leq \lambda \leq 1} \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor \lambda [n\vartheta_0] \rfloor} \eta_j^{(1)} - \Gamma_n \left( \cdot, \cdot, \lambda \frac{[n\vartheta_0]}{n} \right) \right\| \\ &\quad + \frac{n}{[n\hat{\vartheta}_n]} \sup_{\frac{1}{n} \leq \lambda \leq 1} \left\| \Gamma_n \left( \cdot, \cdot, \lambda \frac{[n\hat{\vartheta}_n]}{n} \right) - \Gamma_n(\cdot, \cdot, \lambda \hat{\vartheta}_n) \right\| \\ &\quad + \frac{n}{[n\vartheta_0]} \sup_{\frac{1}{n} \leq \lambda \leq 1} \left\| \Gamma_n \left( \cdot, \cdot, \lambda \frac{[n\vartheta_0]}{n} \right) - \Gamma_n(\cdot, \cdot, \lambda \vartheta_0) \right\| \\ &\quad + \sup_{\frac{1}{n} \leq \lambda \leq 1} \left\| \frac{n}{[n\hat{\vartheta}_n]} \Gamma_n(\cdot, \cdot, \lambda \hat{\vartheta}_n) - \frac{n}{[n\vartheta_0]} \Gamma_n(\cdot, \cdot, \lambda \vartheta_0) \right\| \\ &\leq \sup_{\substack{v, \lambda \in [0, 1]: \\ |v - \lambda| \leq \frac{1}{n}}} \left\| \Gamma_n(\cdot, \cdot, v) - \Gamma_n(\cdot, \cdot, \lambda) \right\| \\ &\quad + O(1) \sup_{\substack{v, \lambda \in [0, 1]: \\ |v - \lambda| \leq |\hat{\vartheta}_n - \vartheta_0|}} \left\| \Gamma_n(\cdot, \cdot, v) - \Gamma_n(\cdot, \cdot, \lambda) \right\| + \frac{|\vartheta_0 - \hat{\vartheta}_n|}{\vartheta_0 \hat{\vartheta}_n} O_p(1) + o_p(1) \\ &= o_p(1), \end{aligned}$$

where we have used (A3) in the last equality. Assertion (2) of Lemma A.2 now follows by using a similar argument for the second difference. ■

We conclude our preparations recalling the definition of the inner product in (4.4) and state a lemma regarding the weak convergence of the process

$$\begin{aligned} (Z_n^{(1)}, Z_n^{(2)})^\top &:= \left\{ (Z_n^{(1)}(\lambda), Z_n^{(2)}(\lambda))^\top \right\}_{\lambda \in [0, 1]} \\ &:= \left\{ \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor \lambda n \rfloor} \left( \langle \eta_j^{(k)} \otimes \eta_j^{(k)} - c^{(k)}, \xi_k \rangle_2 + \langle \eta_j^{(k)}, \zeta_k \rangle \right) \right)_{k=1, 2} \right\}_{\lambda \in [0, 1]}, \end{aligned} \tag{A4}$$

where  $\{\eta_j^{(1)}\}_{j \in \mathbb{Z}}$  and  $\{\eta_j^{(2)}\}_{j \in \mathbb{Z}}$  are centered processes in  $L^2(S \times [0, 1])$  and  $\zeta_1, \zeta_2 \in L^2(S \times [0, 1])$ ,  $\xi_1, \xi_2 \in L^2((S \times [0, 1])^2)$  are given functions. We emphasize that we consider the process  $(Z_n^{(1)}, Z_n^{(2)})^\top$  with different parameters  $\zeta_1, \zeta_2 \in L^2(S \times [0, 1])$ ,  $\xi_1, \xi_2 \in L^2((S \times [0, 1])^2)$  in the proofs of the results of Section 3 and 4. The proof of the following result is similar to the proof of Lemma B.1 in Dette *et al.* (2020c) and therefore omitted.

**Lemma A.3.** Let  $\zeta_1, \zeta_2 \in L^2(S \times [0, 1])$ ,  $\xi_1, \xi_2 \in L^2((S \times [0, 1])^2)$  be fixed but arbitrary functions and let  $\{\eta_j^{(1)}\}_{j \in \mathbb{Z}}$ ,  $\{\eta_j^{(2)}\}_{j \in \mathbb{Z}}$  denote centered processes satisfying Assumption 4.1. Then the process defined in (A4) converges weakly in  $\ell^\infty([0, 1])^2$ , that is

$$(Z_n^{(1)}, Z_n^{(2)})^\top \rightsquigarrow \Sigma^{1/2}(\mathbb{B}_1, \mathbb{B}_2)^\top, \tag{A5}$$

where  $\mathbb{B}_1, \mathbb{B}_2$  are independent Brownian motions and  $\Sigma = (\Sigma_{kl})_{k,l=1,2}$  is a  $2 \times 2$  matrix with entries

$$\begin{aligned} \Sigma_{kl} = \sum_{h \in \mathbb{Z}} & \left( \text{Cov}(\langle \eta_0^{(k)} \otimes \eta_0^{(k)}, \xi_k \rangle_2, \langle \eta_h^{(l)} \otimes \eta_h^{(l)}, \xi_l \rangle_2) + \text{Cov}(\langle \eta_0^{(k)} \otimes \eta_0^{(k)}, \xi_k \rangle_2, \langle \eta_h^{(l)}, \zeta_l \rangle) \right. \\ & \left. + \text{Cov}(\langle \eta_0^{(k)}, \zeta_k \rangle, \langle \eta_h^{(l)} \otimes \eta_h^{(l)}, \xi_l \rangle_2) + \text{Cov}(\langle \eta_0^{(k)}, \zeta_k \rangle, \langle \eta_h^{(l)}, \zeta_l \rangle) \right). \end{aligned} \tag{A6}$$

Moreover, in the case  $\xi_1 \equiv \xi_2 \equiv 0$  Assumption 2.1 instead of Assumption 4.1 is sufficient for the weak convergence in (A5).

A.2. Proof of Theorem 4.1

We will show that

$$\sup_{0 \leq \lambda \leq 1} \sqrt{\lambda} \left\| \hat{\vartheta}_n \hat{c}_\lambda^{(1)} - \vartheta_0 E \left[ \eta_0^{(1)} \otimes \eta_0^{(1)} \right] \right\|_2 = O_p \left( \frac{\log^{2/\kappa}(n)}{\sqrt{n}} \right), \tag{A7}$$

$$\sup_{0 \leq \lambda \leq 1} \sqrt{\lambda} \left\| (1 - \hat{\vartheta}_n) \hat{c}_\lambda^{(2)} - (1 - \vartheta_0) E \left[ \eta_0^{(2)} \otimes \eta_0^{(2)} \right] \right\|_2 = O_p \left( \frac{\log^{2/\kappa}(n)}{\sqrt{n}} \right). \tag{A8}$$

For the sake of brevity, we restrict ourselves to a proof of (A7); the proof of (A8) follows by similar arguments.

First, we assume that  $\lambda < [n\hat{\vartheta}_n]^{-1}$ , then  $\hat{c}_\lambda^{(1)} = 0$  and

$$\sqrt{\lambda} \left\| \hat{\vartheta}_n \hat{c}_\lambda^{(1)} - \vartheta_0 E \left[ \eta_0^{(1)} \otimes \eta_0^{(1)} \right] \right\|_2 < \frac{\vartheta_0}{\sqrt{[n\hat{\vartheta}_n]}} \left| E \left[ \eta_0^{(1)} \otimes \eta_0^{(1)} \right] \right| = O_p(n^{-1/2})$$

uniformly in  $\lambda \in [0, [n\hat{\vartheta}_n]^{-1}]$ . If  $[n\hat{\vartheta}_n]^{-1} \leq \lambda \leq 1$ , we use the representation

$$\hat{c}_\lambda^{(1)} = \frac{1}{[\lambda [n\hat{\vartheta}_n]]} \sum_{j=1}^{[\lambda [n\hat{\vartheta}_n]]} X_j \otimes X_j - \frac{1}{[\lambda [n\hat{\vartheta}_n]]^2} \sum_{i,j=1}^{[\lambda [n\hat{\vartheta}_n]]} X_i \otimes X_j \tag{A9}$$

and consider the first term. For this we get

$$\begin{aligned} & \frac{1}{[\lambda [n\hat{\vartheta}_n]]} \sum_{j=1}^{[\lambda [n\hat{\vartheta}_n]]} X_j \otimes X_j \\ &= \frac{1}{[\lambda [n\hat{\vartheta}_n]]} \sum_{j=1}^{[\lambda [n\hat{\vartheta}_n]] \wedge [n\vartheta_0]} X_j \otimes X_j + \frac{1}{[\lambda [n\hat{\vartheta}_n]]} \sum_{j=[\lambda [n\hat{\vartheta}_n]] \wedge [n\vartheta_0] + 1}^{[\lambda [n\hat{\vartheta}_n]]} X_j \otimes X_j \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{[\lambda[n\hat{\vartheta}_n]]} \sum_{j=1}^{[\lambda[n\hat{\vartheta}_n]] \wedge [n\vartheta_0]} \left( \eta_j^{(1)} \otimes \eta_j^{(1)} + \mu \otimes \eta_j^{(1)} + \eta_j^{(1)} \otimes \mu + \mu \otimes \mu \right) \\
 &\quad + \frac{1}{[\lambda[n\hat{\vartheta}_n]]} \sum_{j=[\lambda[n\hat{\vartheta}_n]] \wedge [n\vartheta_0] + 1}^{[\lambda[n\hat{\vartheta}_n]]} \left( \eta_j^{(2)} \otimes \eta_j^{(2)} + \tilde{\mu} \otimes \eta_j^{(2)} + \eta_j^{(2)} \otimes \tilde{\mu} + \tilde{\mu} \otimes \tilde{\mu} \right), \tag{A10}
 \end{aligned}$$

where we use the notation  $\tilde{\mu} := \mu + \delta$ . The second term in (A9) can be rewritten as follows:

$$\begin{aligned}
 &\frac{1}{[\lambda[n\hat{\vartheta}_n]]^2} \sum_{i,j=1}^{[\lambda[n\hat{\vartheta}_n]]} X_i \otimes X_j \\
 &= \frac{1}{[\lambda[n\hat{\vartheta}_n]]^2} \sum_{i,j=1}^{[\lambda[n\hat{\vartheta}_n]] \wedge [n\vartheta_0]} \left( \eta_i^{(1)} \otimes \eta_j^{(1)} + \mu \otimes \eta_j^{(1)} + \eta_i^{(1)} \otimes \mu + \mu \otimes \mu \right) \\
 &\quad + \frac{\mathbb{1} \{ [n\vartheta_0] \leq [\lambda[n\hat{\vartheta}_n]] \}}{[\lambda[n\hat{\vartheta}_n]]^2} \left\{ \sum_{i=1}^{[n\vartheta_0]} \sum_{j=[n\vartheta_0] + 1}^{[\lambda[n\hat{\vartheta}_n]]} \left( \eta_i^{(1)} \otimes \eta_j^{(2)} + \mu \otimes \eta_j^{(2)} + \eta_i^{(1)} \otimes \tilde{\mu} + \mu \otimes \tilde{\mu} \right) \right. \\
 &\quad \quad \quad + \sum_{i=[n\vartheta_0] + 1}^{[\lambda[n\hat{\vartheta}_n]]} \sum_{j=1}^{[n\vartheta_0]} \left( \eta_i^{(2)} \otimes \eta_j^{(1)} + \eta_i^{(2)} \otimes \mu + \tilde{\mu} \otimes \eta_j^{(1)} + \tilde{\mu} \otimes \mu \right) \\
 &\quad \quad \quad \left. + \sum_{i,j=[n\vartheta_0] + 1}^{[\lambda[n\hat{\vartheta}_n]]} \left( \eta_i^{(2)} \otimes \eta_j^{(2)} + \eta_i^{(2)} \otimes \tilde{\mu} + \tilde{\mu} \otimes \eta_j^{(2)} + \tilde{\mu} \otimes \tilde{\mu} \right) \right\}.
 \end{aligned}$$

We now consider the cases  $[\lambda[n\hat{\vartheta}_n]] < [n\vartheta_0]$  and  $[\lambda[n\hat{\vartheta}_n]] \geq [n\vartheta_0]$  separately. For this purpose, we define the set  $\Lambda_n := \{ [n\hat{\vartheta}_n]^{-1} \leq \lambda \leq 1 | [\lambda[n\hat{\vartheta}_n]] < [n\vartheta_0] \}$  and note that

$$\begin{aligned}
 &\sup_{\frac{1}{n} \leq \lambda \leq 1} \sqrt{\lambda} \left\| \hat{\vartheta}_n \hat{c}_\lambda^{(1)} - \vartheta_0 E \left[ \eta_0^{(1)} \otimes \eta_0^{(1)} \right] \right\|_2 \\
 &= \max \left\{ \sup_{\lambda \in \Lambda_n} \sqrt{\lambda} \left\| \hat{\vartheta}_n \hat{c}_\lambda^{(1)} - \vartheta_0 E \left[ \eta_0^{(1)} \otimes \eta_0^{(1)} \right] \right\|_2, \sup_{\lambda \in \Lambda_n^c} \sqrt{\lambda} \left\| \hat{\vartheta}_n \hat{c}_\lambda^{(1)} - \vartheta_0 E \left[ \eta_0^{(1)} \otimes \eta_0^{(1)} \right] \right\|_2 \right\}.
 \end{aligned}$$

We consider each case individually, starting with the first term, where the supremum is taken over the set  $\Lambda_n$ , that is, the inequality  $[\lambda[n\hat{\vartheta}_n]] < [n\vartheta_0]$  holds. Observing that the sum in (A10) vanishes, we obtain in this case that

$$\hat{c}_\lambda^{(1)} = \frac{1}{[\lambda[n\hat{\vartheta}_n]]} \sum_{j=1}^{[\lambda[n\hat{\vartheta}_n]]} \eta_j^{(1)} \otimes \eta_j^{(1)} - \frac{1}{[\lambda[n\hat{\vartheta}_n]]^2} \sum_{i,j=1}^{[\lambda[n\hat{\vartheta}_n]]} \eta_i^{(1)} \otimes \eta_j^{(1)}. \tag{A11}$$

In the subsequent discussion, we will repeatedly make use of the following inequality, without explicitly mentioning it. It is obtained by expanding the set over which the supremum is calculated and substituting

$$\lambda' = \lambda \frac{\lfloor n\hat{\vartheta}_n \rfloor}{n}:$$

$$\sup_{\lambda \in \Lambda_n} \frac{1}{\sqrt{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor}} \left\| \sum_{j=1}^{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor} \eta_j^{(1)} \right\| \leq \sup_{\frac{1}{n} \leq \lambda' \leq \frac{\lfloor n\hat{\vartheta}_n \rfloor}{n}} \frac{1}{\sqrt{\lfloor \lambda' n \rfloor}} \left\| \sum_{j=1}^{\lfloor \lambda' n \rfloor} \eta_j^{(1)} \right\| \leq \sup_{\frac{1}{n} \leq \lambda \leq 1} \frac{1}{\sqrt{\lfloor \lambda n \rfloor}} \left\| \sum_{j=1}^{\lfloor \lambda n \rfloor} \eta_j^{(1)} \right\|.$$

Regarding the second term in (A11) we have with  $\kappa$  from (A4')

$$\begin{aligned} \sup_{\lambda \in \Lambda_n} \sqrt{\lambda \hat{\vartheta}_n} \left\| \frac{1}{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor^2} \sum_{i,j=1}^{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor} \eta_i^{(1)} \otimes \eta_j^{(1)} \right\|_2 &= \sup_{\lambda \in \Lambda_n} \frac{\lambda \hat{\vartheta}_n}{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor} \frac{1}{\sqrt{\lambda}} \left\| \frac{1}{\sqrt{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor}} \sum_{j=1}^{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor} \eta_j^{(1)} \right\|_2^2 \\ &\leq O_P(n^{-1/2}) \left( \sup_{\frac{1}{n} \leq \lambda \leq 1} \frac{1}{\sqrt{\lfloor \lambda n \rfloor}} \left\| \sum_{j=1}^{\lfloor \lambda n \rfloor} \eta_j^{(1)} \right\| \right)^2 \\ &= O_P\left(\frac{\log^{2/\kappa}(n)}{\sqrt{n}}\right), \end{aligned}$$

by (A1). This implies

$$\begin{aligned} &\sup_{\lambda \in \Lambda_n} \sqrt{\lambda} \left\| \hat{\vartheta}_n \hat{c}_\lambda^{(1)} - \vartheta_0 E \left[ \eta_0^{(1)} \otimes \eta_0^{(1)} \right] \right\|_2 \\ &\leq \sup_{\lambda \in \Lambda_n} \sqrt{\lambda} \left\| \frac{\hat{\vartheta}_n}{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor} \sum_{j=1}^{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor} \left( \eta_j^{(1)} \otimes \eta_j^{(1)} - E \left[ \eta_0^{(1)} \otimes \eta_0^{(1)} \right] \right) \right\|_2 \\ &\quad + |\hat{\vartheta}_n - \vartheta_0| E \left[ \eta_0^{(1)} \otimes \eta_0^{(1)} \right] + O_P\left(\frac{\log^{2/\kappa}(n)}{\sqrt{n}}\right) \\ &\leq O_P(n^{-1/2}) \sup_{\frac{1}{n} \leq \lambda \leq 1} \frac{1}{\sqrt{\lfloor \lambda n \rfloor}} \left\| \sum_{j=1}^{\lfloor \lambda n \rfloor} \left( \eta_j^{(1)} \otimes \eta_j^{(1)} - E \left[ \eta_0^{(1)} \otimes \eta_0^{(1)} \right] \right) \right\|_2 + O_P\left(\frac{\log^{2/\kappa}(n)}{\sqrt{n}}\right) \\ &\leq O_P\left(\frac{\log^{2/\kappa}(n)}{\sqrt{n}}\right), \end{aligned}$$

where we used (3.2) and again a variation of (A1), this time for  $L^2((S \times [0, 1])^2)$ -valued random variables. It remains to calculate the supremum over the set  $\Lambda_n^C$ . If  $\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor \geq \lfloor n\vartheta_0 \rfloor$  is true, then a tedious but straightforward calculation gives

$$\begin{aligned} \hat{c}_\lambda^{(1)} &= \frac{1}{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor} \sum_{j=1}^{\lfloor n\vartheta_0 \rfloor} \eta_j^{(1)} \otimes \eta_j^{(1)} - \frac{1}{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor^2} \sum_{i,j=1}^{\lfloor n\vartheta_0 \rfloor} \eta_i^{(1)} \otimes \eta_j^{(1)} \\ &\quad + \frac{1}{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor} \sum_{j=\lfloor n\vartheta_0 \rfloor+1}^{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor} (\eta_j^{(2)} + \delta) \otimes (\eta_j^{(2)} + \delta) \end{aligned}$$

$$\begin{aligned}
 & - \frac{1}{[\lambda \lfloor n \hat{\vartheta}_n \rfloor]^2} \sum_{i=1}^{\lfloor n \vartheta_0 \rfloor} \sum_{j=\lfloor n \vartheta_0 \rfloor+1}^{\lfloor \lambda \lfloor n \hat{\vartheta}_n \rfloor \rfloor} \eta_i^{(1)} \otimes (\eta_j^{(2)} + \delta) - \frac{1}{[\lambda \lfloor n \hat{\vartheta}_n \rfloor]^2} \sum_{i=\lfloor n \vartheta_0 \rfloor+1}^{\lfloor \lambda \lfloor n \hat{\vartheta}_n \rfloor \rfloor} \sum_{j=1}^{\lfloor n \vartheta_0 \rfloor} (\eta_i^{(2)} + \delta) \otimes \eta_j^{(1)} \\
 & + \frac{1}{[\lambda \lfloor n \hat{\vartheta}_n \rfloor]^2} \sum_{i,j=\lfloor n \vartheta_0 \rfloor+1}^{\lfloor \lambda \lfloor n \hat{\vartheta}_n \rfloor \rfloor} (\eta_i^{(2)} + \delta) \otimes (\eta_j^{(2)} + \delta)
 \end{aligned}$$

and we claim that

$$\begin{aligned}
 & \sup_{\lambda \in \Lambda_n^c} \sqrt{\lambda} \left\| \hat{\vartheta}_n \hat{c}_\lambda^{(1)} - \vartheta_0 E \left[ \eta_0^{(1)} \otimes \eta_0^{(1)} \right] \right\|_2 \\
 & \leq \sup_{\frac{1}{n} \leq \lambda \leq 1} \sqrt{\lambda} \left\| \frac{\hat{\vartheta}_n}{[\lambda \lfloor n \hat{\vartheta}_n \rfloor]} \sum_{j=1}^{\lfloor n \vartheta_0 \rfloor} \eta_j^{(1)} \otimes \eta_j^{(1)} - \vartheta_0 E \left[ \eta_0^{(1)} \otimes \eta_0^{(1)} \right] \right\|_2 + O_P \left( \frac{\log^{2/\kappa}(n)}{\sqrt{n}} \right). \tag{A12}
 \end{aligned}$$

For a proof of (A12), we show that

$$\sup_{\lambda \in \Lambda_n^c} \left\| \frac{1}{[\lambda \lfloor n \hat{\vartheta}_n \rfloor]} \sum_{j=1}^{\lfloor n \vartheta_0 \rfloor} \eta_j^{(1)} \right\| = O_P(n^{-1/2}), \tag{A13}$$

$$\sup_{\lambda \in \Lambda_n^c} \left\| \frac{1}{[\lambda \lfloor n \hat{\vartheta}_n \rfloor]} \sum_{j=\lfloor n \vartheta_0 \rfloor+1}^{\lfloor \lambda \lfloor n \hat{\vartheta}_n \rfloor \rfloor} (\eta_j^{(2)} + \delta) \right\| = O_P \left( \frac{\log^{1/\kappa}(n)}{\sqrt{n}} \right), \tag{A14}$$

$$\sup_{\lambda \in \Lambda_n^c} \sqrt{\lambda} \hat{\vartheta}_n \left\| \frac{1}{[\lambda \lfloor n \hat{\vartheta}_n \rfloor]} \sum_{j=\lfloor n \vartheta_0 \rfloor+1}^{\lfloor \lambda \lfloor n \hat{\vartheta}_n \rfloor \rfloor} (\eta_j^{(2)} + \delta) \right\|^2 = O_P \left( \frac{\log^{2/\kappa}(n)}{\sqrt{n}} \right), \tag{A15}$$

$$\sup_{\lambda \in \Lambda_n^c} \sqrt{\lambda} \hat{\vartheta}_n \left\| \frac{1}{[\lambda \lfloor n \hat{\vartheta}_n \rfloor]} \sum_{j=\lfloor n \vartheta_0 \rfloor+1}^{\lfloor \lambda \lfloor n \hat{\vartheta}_n \rfloor \rfloor} (\eta_j^{(2)} + \delta) \otimes (\eta_j^{(2)} + \delta) \right\| = O_P \left( \frac{\log^{1/\kappa}(n)}{\sqrt{n}} \right). \tag{A16}$$

The estimate (A13) follows from

$$\sup_{\lambda \in \Lambda_n^c} \left\| \frac{1}{[\lambda \lfloor n \hat{\vartheta}_n \rfloor]} \sum_{j=1}^{\lfloor n \vartheta_0 \rfloor} \eta_j^{(1)} \right\| \leq \frac{1}{\sqrt{\lfloor n \vartheta_0 \rfloor}} \left\| \frac{1}{\sqrt{\lfloor n \vartheta_0 \rfloor}} \sum_{j=1}^{\lfloor n \vartheta_0 \rfloor} \eta_j^{(1)} \right\| = O_P(n^{-1/2})$$

because of  $\lfloor n \vartheta_0 \rfloor \leq \lfloor \lambda \lfloor n \hat{\vartheta}_n \rfloor \rfloor$  and the central limit theorem in Hilbert spaces in the last step.

For the estimate (A14) recall that by (3.2) we have  $\hat{\vartheta}_n = \vartheta_0 + o_P(n^{-1/2})$ . Moreover, this implies

$$\sup_{\lambda \in \Lambda_n^c} \frac{\lfloor n \vartheta_0 \rfloor}{[\lambda \lfloor n \hat{\vartheta}_n \rfloor]} = 1 + o_P(n^{-1/2}), \tag{A17}$$

because

$$1 + o_P(n^{-1/2}) = \frac{\lfloor n \vartheta_0 \rfloor}{\lfloor n \hat{\vartheta}_n \rfloor} \leq \sup_{\lambda \in \Lambda_n^c} \frac{\lfloor n \vartheta_0 \rfloor}{[\lambda \lfloor n \hat{\vartheta}_n \rfloor]} \leq \sup_{\lambda \in \Lambda_n^c} \frac{[\lambda \lfloor n \hat{\vartheta}_n \rfloor]}{[\lambda \lfloor n \hat{\vartheta}_n \rfloor]} = 1.$$

Using again  $\lfloor n\vartheta_0 \rfloor \leq \lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor$  and the estimate (A1), we find that

$$\begin{aligned} & \sup_{\lambda \in \Lambda_n^c} \left\| \frac{1}{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor} \sum_{j=\lfloor n\vartheta_0 \rfloor+1}^{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor} (\eta_j^{(2)} + \delta) \right\| \\ & \leq \frac{1}{\sqrt{\lfloor n\vartheta_0 \rfloor}} \sup_{\lambda \in \Lambda_n^c} \left\| \frac{1}{\sqrt{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor}} \sum_{j=1}^{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor} \eta_j^{(2)} \right\| \\ & \quad + \frac{1}{\sqrt{\lfloor n\vartheta_0 \rfloor}} \left\| \frac{1}{\sqrt{\lfloor n\vartheta_0 \rfloor}} \sum_{j=1}^{\lfloor n\vartheta_0 \rfloor} \eta_j^{(2)} \right\| + \sup_{\lambda \in \Lambda_n^c} \frac{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor - \lfloor n\vartheta_0 \rfloor}{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor} \|\delta\| \\ & \leq O_P(n^{-1/2}) \sup_{\frac{1}{n} \leq \lambda \leq 1} \left\| \frac{1}{\lfloor \lambda n \rfloor} \sum_{j=1}^{\lfloor \lambda n \rfloor} \eta_j^{(2)} \right\| + \frac{\lfloor n\hat{\vartheta}_n \rfloor - \lfloor n\vartheta_0 \rfloor}{\lfloor n\vartheta_0 \rfloor} \|\delta\| + O_P(n^{-1/2}) \\ & = O_P\left(\frac{\log^{1/\kappa}(n)}{\sqrt{n}}\right). \end{aligned}$$

These calculations also yield the estimate (A15). For (A16) we have, using a version of (A1) for  $L^2((S \times [0, 1])^2)$ -valued random variables and the fact that  $\lfloor n\vartheta_0 \rfloor \leq \lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor$ , that

$$\begin{aligned} & \sup_{\lambda \in \Lambda_n^c} \sqrt{\lambda \hat{\vartheta}_n} \left\| \frac{1}{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor} \sum_{j=\lfloor n\vartheta_0 \rfloor+1}^{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor} (\eta_j^{(2)} + \delta) \otimes (\eta_j^{(2)} + \delta) \right\|_2 \\ & \leq \frac{1}{\sqrt{\lfloor n\vartheta_0 \rfloor}} \sup_{\lambda \in \Lambda_n^c} \left\| \frac{1}{\sqrt{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor}} \sum_{j=1}^{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor} \tilde{\eta}_j^{\otimes} \right\|_2 + \frac{\hat{\vartheta}_n}{\sqrt{\lfloor n\vartheta_0 \rfloor}} \left\| \frac{1}{\sqrt{\lfloor n\vartheta_0 \rfloor}} \sum_{j=1}^{\lfloor n\vartheta_0 \rfloor} \tilde{\eta}_j^{\otimes} \right\|_2 \\ & \quad + o_P(n^{-1/2}) \cdot E \left[ (\eta_0^{(2)} + \delta) \otimes (\eta_0^{(2)} + \delta) \right] \\ & = O_P\left(\frac{\log^{1/\kappa}(n)}{\sqrt{n}}\right), \end{aligned}$$

where  $\tilde{\eta}_j^{\otimes} := (\eta_j^{(2)} + \delta) \otimes (\eta_j^{(2)} + \delta) - E \left[ (\eta_0^{(2)} + \delta) \otimes (\eta_0^{(2)} + \delta) \right]$ . Finally, observing (A12), the assertion (A7) follows from

$$\begin{aligned} & \sup_{\lambda \in \Lambda_n^c} \sqrt{\lambda} \left\| \frac{\hat{\vartheta}_n}{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor} \sum_{j=1}^{\lfloor n\vartheta_0 \rfloor} \eta_j^{(1)} \otimes \eta_j^{(1)} - \vartheta_0 E \left[ \eta_0^{(1)} \otimes \eta_0^{(1)} \right] \right\|_2 \\ & \leq \frac{1}{\sqrt{\lfloor n\vartheta_0 \rfloor}} \left\| \frac{1}{\sqrt{\lfloor n\vartheta_0 \rfloor}} \sum_{j=1}^{\lfloor n\vartheta_0 \rfloor} \left( \eta_j^{(1)} \otimes \eta_j^{(1)} - E \left[ \eta_0^{(1)} \otimes \eta_0^{(1)} \right] \right) \right\|_2 \\ & \quad + \sup_{\lambda \in \Lambda_n^c} \left| \frac{\lfloor n\vartheta_0 \rfloor}{\lfloor \lambda \lfloor n\hat{\vartheta}_n \rfloor \rfloor} \hat{\vartheta}_n - \vartheta_0 \right| E \left[ \eta_0^{(1)} \otimes \eta_0^{(1)} \right] \\ & = O_P(n^{-1/2}) \end{aligned}$$

by (A1) and (A17). This completes the proof of Theorem 4.1.

A.3. Proof of Theorem 4.2

We begin stating an expansion for the difference  $\hat{c}_{j,\lambda}\hat{w}_{j,\lambda} - w_j$ , which is proved by similar arguments as given in the proof of Proposition 2.1 in Aue *et al.* (2023). The details are therefore omitted.

**Proposition A.4.** Suppose Assumption 4.1 and 4.2 hold. If further  $\vartheta_0 \in (\varepsilon, 1 - \varepsilon)$ , then for any  $k \leq d$  and some  $\alpha > 0$  we have

$$\sup_{0 \leq \lambda \leq 1} \left\| \lambda(\hat{c}_{k,\lambda}\hat{w}_{k,\lambda} - w_k) - \frac{1}{\sqrt{n}} \sum_{i \neq k} \frac{w_i}{\tau_k - \tau_i} \langle \hat{Z}_{n,\lambda}, w_k \otimes w_i \rangle_2 \right\| = O_P \left( \frac{\log^\alpha(n)}{n} \right)$$

and

$$\sup_{0 \leq \lambda \leq 1} \sqrt{\lambda} \|\hat{c}_{k,\lambda}\hat{w}_{k,\lambda} - w_k\| = O_P \left( \frac{\log^{1/\alpha}(n)}{\sqrt{n}} \right),$$

where  $\hat{c}_{k,\lambda} = \text{sign}\langle \hat{w}_{k,\lambda}, w_k \rangle$  and

$$\hat{Z}_{n,\lambda} := \frac{1}{\sqrt{n}} \left( \sum_{j=1}^{\lfloor \lambda \lfloor n\vartheta_0 \rfloor \rfloor} (\eta_j^{(1)} \otimes \eta_j^{(1)} - c^{(1)}) + \sum_{j=\lfloor n\vartheta_0 \rfloor + 1}^{\lfloor n\vartheta_0 \rfloor + \lfloor \lambda(n - \lfloor n\vartheta_0 \rfloor) \rfloor} (\eta_j^{(2)} \otimes \eta_j^{(2)} - c^{(2)}) \right).$$

For the proof of the first assertion (4.6) of Theorem 4.2 we add and subtract  $\lambda\delta$ , use (A2) and obtain uniformly in  $\lambda \in [0, 1]$

$$\langle D_n(\lambda, \vartheta_0), \hat{w}_{k,\lambda} \rangle^2 - \lambda^2 \langle \delta, w_k \rangle^2 = I_{k,n}(\lambda) + J_{k,n}(\lambda) + K_{k,n}(\lambda) + O_P(n^{-1}), \tag{A18}$$

where  $\hat{c}_{k,\lambda}$  is defined in Proposition A.4 and

$$\begin{aligned} I_{k,n}(\lambda) &:= \langle \tilde{D}_n(\lambda, \vartheta_0), \hat{c}_{k,\lambda}\hat{w}_{k,\lambda} \rangle^2, \\ J_{k,n}(\lambda) &:= 2\lambda \langle \tilde{D}_n(\lambda, \vartheta_0), \hat{c}_{k,\lambda}\hat{w}_{k,\lambda} \rangle \langle \delta, \hat{c}_{k,\lambda}\hat{w}_{k,\lambda} \rangle, \\ K_{k,n}(\lambda) &:= \lambda^2 \langle \delta, \hat{c}_{k,\lambda}\hat{w}_{k,\lambda} \rangle^2 - \lambda^2 \langle \delta, w_k \rangle^2. \end{aligned}$$

By Theorem A.1  $\sup_{0 \leq \lambda \leq 1} \|\sqrt{n}\tilde{D}_n(\lambda, \vartheta_0)\|$  is bounded. Therefore, the Cauchy–Schwarz inequality yields

$$\sup_{0 \leq \lambda \leq 1} \sqrt{n}I_{k,n}(\lambda) \leq \frac{1}{\sqrt{n}} \sup_{0 \leq \lambda \leq 1} \|\sqrt{n}\tilde{D}_n(\lambda, \vartheta_0)\|^2 \|\hat{w}_{k,\lambda}\|^2 = o_P(1)$$

for all  $k = 1, \dots, d$ . Moving to the second term  $J_{k,n}$ , we see that

$$\begin{aligned} \sqrt{n}J_{k,n}(\lambda) &= \sqrt{n}2\lambda \langle \tilde{D}_n(\lambda, \vartheta_0), \hat{c}_{k,\lambda}\hat{w}_{k,\lambda} - w_k \rangle \langle \delta, \hat{c}_{k,\lambda}\hat{w}_{k,\lambda} - w_k \rangle \\ &\quad + \langle \tilde{D}_n(\lambda, \vartheta_0), w_k \rangle \langle \delta, \hat{c}_{k,\lambda}\hat{w}_{k,\lambda} - w_k \rangle \\ &\quad + \langle \tilde{D}_n(\lambda, \vartheta_0), \hat{c}_{k,\lambda}\hat{w}_{k,\lambda} - w_k \rangle \langle \delta, w_k \rangle + \langle \tilde{D}_n(\lambda, \vartheta_0), w_k \rangle \langle \delta, w_k \rangle \\ &= 2\lambda \langle \sqrt{n}\tilde{D}_n(\lambda, \vartheta_0), w_k \rangle \langle \delta, w_k \rangle + o_P(1), \end{aligned}$$

uniformly in  $\lambda$  by the second part of Proposition A.4. Lastly, we have by the first part of Proposition A.4 uniformly in  $\lambda$

$$\begin{aligned} \sqrt{n}K_{k,n}(\lambda) &= \langle \delta, \lambda(\hat{c}_{k,\lambda}\hat{w}_{k,\lambda} - w_k) \rangle^2 + 2\lambda \langle \delta, \lambda(\hat{c}_{k,\lambda}\hat{w}_{k,\lambda} - w_k) \rangle \langle \delta, w_k \rangle \\ &= 2\lambda \langle \delta, w_k \rangle \left\langle \delta, \frac{1}{\sqrt{n}} \sum_{i \neq k} \frac{w_i}{\tau_k - \tau_i} \langle \hat{Z}_{n,\lambda}, w_k \otimes w_i \rangle_2 \right\rangle + o_P(1). \end{aligned} \tag{A19}$$

Recalling the notations of  $w$  and  $f$  in (4.8) and (4.9) and combining (A18) – (A19), we can rewrite uniformly in  $\lambda$

$$\begin{aligned} \mathbb{T}_n(\lambda) &:= \sqrt{n} \sum_{k=1}^d (\langle D_n(\lambda, \vartheta_0), \hat{w}_{k,\lambda} \rangle^2 - \lambda^2 \langle \delta, w_k \rangle^2) \\ &= \frac{2\lambda}{\sqrt{n}} \sum_{j=1}^{\lfloor \lambda \lfloor n\vartheta_0 \rfloor \rfloor} \langle \eta_j^{(1)} \otimes \eta_j^{(1)} - c^{(1)}, f \rangle_2 + \frac{2\lambda}{\sqrt{n}} \sum_{j=\lfloor n\vartheta_0 \rfloor + 1}^{\lfloor n\vartheta_0 \rfloor + \lfloor \lambda(n - \lfloor n\vartheta_0 \rfloor) \rfloor} \langle \eta_j^{(2)} \otimes \eta_j^{(2)} - c^{(2)}, f \rangle_2 \\ &\quad - \frac{2\lambda\sqrt{n}}{\lfloor n\vartheta_0 \rfloor} \sum_{j=1}^{\lfloor \lambda \lfloor n\vartheta_0 \rfloor \rfloor} \langle \eta_j^{(1)}, w \rangle - \frac{2\lambda\sqrt{n}}{n - \lfloor n\vartheta_0 \rfloor} \sum_{j=\lfloor n\vartheta_0 \rfloor + 1}^{\lfloor n\vartheta_0 \rfloor + \lfloor \lambda(n - \lfloor n\vartheta_0 \rfloor) \rfloor} \langle \eta_j^{(2)}, w \rangle + o_P(1) \\ &= 2\lambda \left( Z_n^{(1)} \left( \frac{\lambda \lfloor n\vartheta_0 \rfloor}{n} \right) + Z_n^{(2)} \left( \frac{\lfloor n\vartheta_0 \rfloor + \lfloor \lambda(n - \lfloor n\vartheta_0 \rfloor) \rfloor}{n} \right) - Z_n^{(2)}(\vartheta_0) \right) + o_P(1), \end{aligned}$$

where for  $\ell = 1, 2$

$$Z_n^{(\ell)}(\lambda) = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor \lambda n \rfloor} \left( \langle \eta_j^{(\ell)} \otimes \eta_j^{(\ell)} - c^{(\ell)}, \xi_\ell \rangle_2 + \langle \eta_j^{(\ell)}, \zeta_\ell \rangle \right)$$

with  $\xi_1 = \xi_2 = f$ ,  $\zeta_1 = -\frac{1}{\vartheta_0}w$  and  $\zeta_2 = -\frac{1}{1-\vartheta_0}w$ . Now by Lemma A.3 we obtain

$$\begin{aligned} \{\mathbb{T}_n(\lambda)\}_{\lambda \in [0,1]} &\rightsquigarrow \{\mathbb{Z}(\lambda)\}_{\lambda \in [0,1]} \\ &:= \{2\lambda \mathbb{Z}^{(1)}(\lambda\vartheta_0) + 2\lambda (\mathbb{Z}^{(2)}(\vartheta_0 + \lambda(1 - \vartheta_0)) - \mathbb{Z}^{(2)}(\vartheta_0))\}_{\lambda \in [0,1]} \end{aligned}$$

in  $\ell^\infty([0, 1])$ , where  $\mathbb{Z}^{(1)}(\lambda) := \tilde{\Sigma}_{11}\mathbb{B}_1(\lambda) + \tilde{\Sigma}_{12}\mathbb{B}_2(\lambda)$  and  $\mathbb{Z}^{(2)}(\lambda) := \tilde{\Sigma}_{21}\mathbb{B}_1 + \tilde{\Sigma}_{22}\mathbb{B}_2$ .  $\tilde{\Sigma}_{ij}$  denotes the  $ij$ -th entry of the matrix  $\Sigma^{1/2}$  and the entries of  $\Sigma = (\Sigma_{ij})_{i,j=1,2}$  are defined by (A6) with  $\xi_1 = \xi_2 = f$ ,  $\zeta_1 = -\frac{1}{\vartheta_0}w$  and  $\zeta_2 = -\frac{1}{1-\vartheta_0}w$ . Inspecting the covariance structure of the limiting process, we see that

$$\text{Cov}(\mathbb{Z}(\lambda_1), \mathbb{Z}(\lambda_2)) = 4\lambda_1\lambda_2(\lambda_1 \wedge \lambda_2)(\vartheta_0\Sigma_{11} + (1 - \vartheta_0)\Sigma_{22})$$

and hence the weak convergence in (4.6) follows.

To prove the second assertion (4.7) we note that we have by part 1 and 2 of Lemma A.2

$$\begin{aligned} &\sup_{0 \leq \lambda \leq 1} (\langle D_n(\lambda, \hat{\vartheta}_n), \hat{w}_{k,\lambda} \rangle^2 - \langle D_n(\lambda, \vartheta_0), \hat{w}_{k,\lambda} \rangle^2) \\ &= \sup_{0 \leq \lambda \leq 1} (\langle D_n(\lambda, \hat{\vartheta}_n) - D_n(\lambda, \vartheta_0), \hat{w}_{k,\lambda} \rangle^2 + 2\langle D_n(\lambda, \hat{\vartheta}_n) - D_n(\lambda, \vartheta_0), \hat{w}_{k,\lambda} \rangle \langle D_n(\lambda, \vartheta_0), \hat{w}_{k,\lambda} \rangle) \\ &\leq \sup_{0 \leq \lambda \leq 1} \|D_n(\lambda, \hat{\vartheta}_n) - D_n(\lambda, \vartheta_0)\|^2 \|\hat{w}_{k,\lambda}\|^2 \\ &\quad + 2 \sup_{0 \leq \lambda \leq 1} \|D_n(\lambda, \hat{\vartheta}_n) - D_n(\lambda, \vartheta_0)\| \|D_n(\lambda, \vartheta_0)\| \|\hat{w}_{k,\lambda}\|^2 \\ &= o_P(n^{-1/2}). \end{aligned}$$

Consequently, assertion (4.7) follows from (4.6).