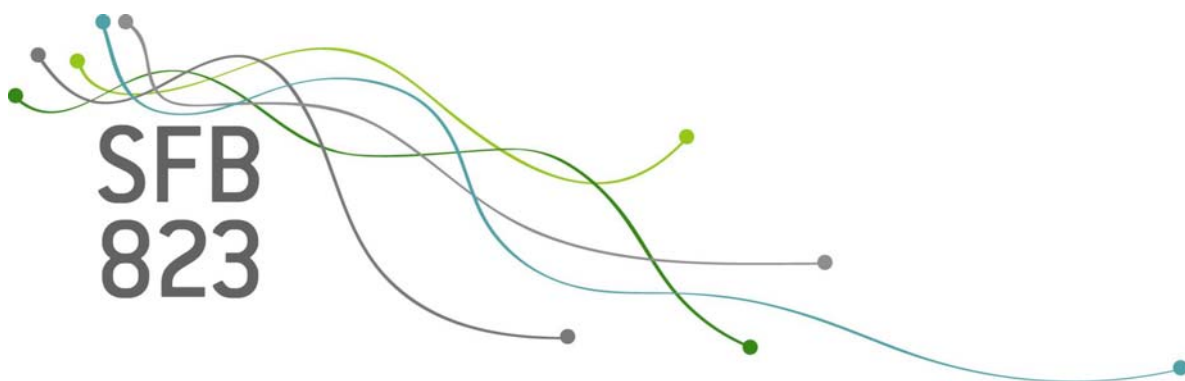# Functional data analysis with increasing number of projections

Stefan Fremdt, Lajos Horváth, Piotr Kokoszka, Josef G. Steinebach

SFB
823

FUNCTIONAL DATA ANALYSIS WITH INCREASING NUMBER OF PROJECTIONS

STEFAN FREMDT, LAJOS HORVÁTH, PIOTR KOKOSZKA, AND JOSEF G. STEINEBACH

ABSTRACT. Functional principal components (FPC's) provide the most important and most extensively used tool for dimension reduction and inference for functional data. The selection of the number, d, of the FPC's to be used in a specific procedure has attracted a fair amount of attention, and a number of reasonably effective approaches exist. Intuitively, they assume that the functional data can be sufficiently well approximated by a projection onto a finite–dimensional subspace, and the error resulting from such an approximation does not impact the conclusions. This has been shown to be a very effective approach, but it is desirable to understand the behavior of many inferential procedures by considering the projections on subspaces spanned by an increasing number of the FPC's. Such an approach reflects more fully the infinite–dimensional nature of functional data, and allows to derive procedures which are fairly insensitive to the selection of d. This is accomplished by considering limits as $d \to \infty$ with the sample size.

We propose a specific framework in which we let $d \to \infty$ by deriving a normal approximation for the partial sum process

$$\sum_{j=1}^{\lfloor du \rfloor} \sum_{i=1}^{\lfloor Nx \rfloor} \xi_{i,j}, \quad 0 \le u \le 1, \ \ 0 \le x \le 1,$$

where N is the sample size and $\xi_{i,j}$ is the score of the ith function with respect to the jth FPC. Our approximation can be used to derive statistics that use segments of observations and segments of the FPC's. We apply our general results to derive two inferential procedures for the mean function: a change–point test and a two–sample test. In addition to the asymptotic theory, the tests are assessed through a small simulation study and a data example.

1. **Introduction**

Functional data analysis has grown into a comprehensive and useful field of statistics which provides a convenient framework to handle some high–dimensional data structures, including curves and images. The monograph of Ramsay and Silverman (2005) has done a lot to introduce its ideas to the statistics community and beyond. Several other monographs and thousands of papers followed. This paper focuses on a specific aspect of the mathematical foundations of functional data analysis, which is however of fairly central importance. We first describe the contribution of this paper in broad terms, and provide some more detailed background and discussion in the latter part of this section. Perhaps the most important, and definitely the most commonly used, tool for dimension reduction of functional data is the principal component analysis. Suppose we observe a sample of functions, $X_1, X_2, \ldots, X_N$, and denote by

$$\hat{\eta}_{i,j} = \int \left(X_i(t) - \bar{X}_N(t) \right) \hat{v}_j(t) dt, \quad i = 1, 2, \ldots, N, \quad j = 1, 2, \ldots, d,$$

the scores of the X_i with respect to the estimated functional principal components $\hat{v}_j$. The scores $\hat{\eta}_{i,j}$ depend on two variables i and j, and to reflect the infinite–dimensional nature of the data, it may be desirable to consider asymptotics in which both N and d increase. This paper establishes results that allow us to study the two–dimensional partial sum process

$$\sum_{j=1}^{\lfloor du \rfloor} \sum_{i=1}^{\lfloor Nx \rfloor} \int \left(X_i(t) - \mu_X(t) \right) v_j(t) dt, \quad 0 \le u \le 1, \quad 0 \le x \le 1.$$

More specifically, we derive a uniform normal approximation and apply it to two problems related to testing the null hypothesis that all observed curves have the same mean function. We obtain new test statistics in which the number of the functional principal components, d, increases slowly with the sample size N. We hope that our general approach will be used to derive similar results in other settings.

Statistical procedures for functional data which use functional principal components (FPC's) often depend on the number d of the components used to compute various statistics. The selection of an optimal d has received a fair deal of attention. Commonly used approaches include the cumulative variance method, the scree plot, and several forms of cross–validation and pseudo information criteria. By now, most of these approaches are implemented in several R packages and in the Matlab package PACE. A related direction of research has focused on the identification of the dimension d assuming that the functional data actually live in a finite–dimensional space of this dimension, see Hall and Vial (2006) and Bathia et al. (2010). The research presented in this paper is concerned with functional data which cannot be reduced to finite–dimensional data in an obvious and easy way. Such data are typically characterized by a slow decay of the eigenvalues of the empirical covariance operator. Figure 1 shows the eigenvalues of the empirical covariance operator of the annual temperature curves obtained over the period 1856–2011 in Melbourne, Australia, while Figure 2 shows the cumulative variance plot for the same data set. It is seen that the eigenfunctions decay at a slow rate, and neither
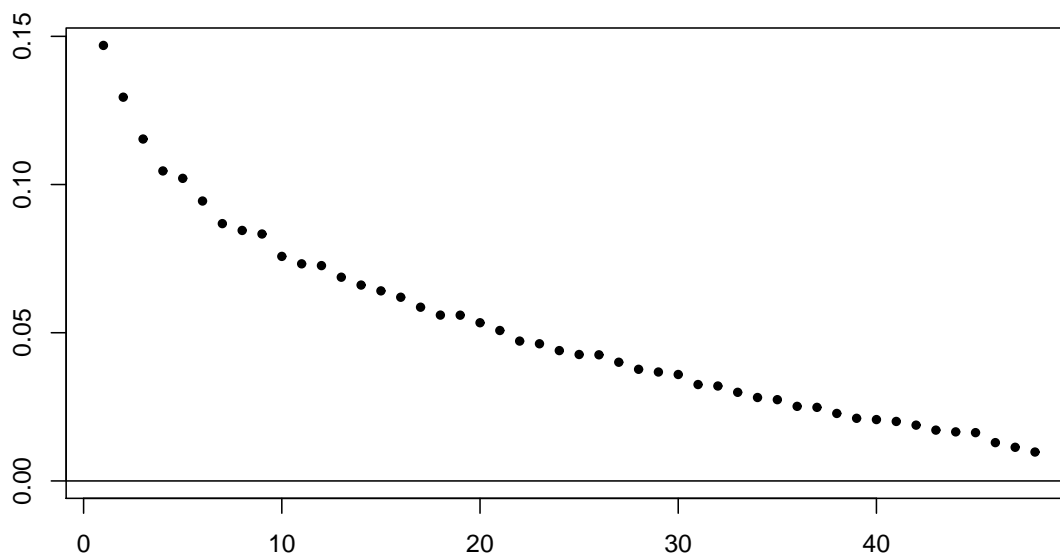
FIGURE 1. Melbourne temperature data: eigenvalues $\hat{\lambda}_2, \ldots, \hat{\lambda}_{49}$.
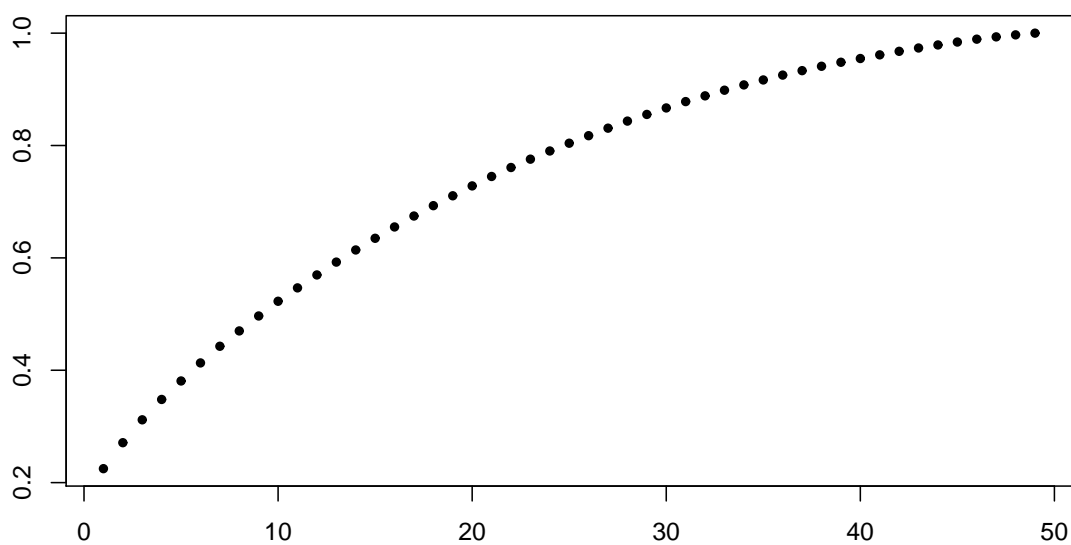

FIGURE 2. Melbourne temperature data: percentage of variance explained by the first k eigenvalues, i.e. $f_k = \sum_{i=1}^{k} \hat{\lambda}_i / \sum_{j=1}^{N} \hat{\lambda}_j$, $k = 1, 2, \ldots, 49$.

their visual inspection nor the analysis of cumulative variance provide a clear guidance on how to select d. This data set is analyzed in greater detail in Section 5.

In situations when the choice of d is difficult, two approaches seem reasonable. In the first approach, one can apply a test using several values of d in a reasonable range. If the conclusion does not depend on d, we can be confident that it is correct. This approach has been used in applied research, see Gromenko et al. (2012) for a recent analysis of this type. The second approach, would be to let d increase with the sample size N, and derive a test statistic based on the limit. In a sense, the second approach is a formalization of the first one because if a limit as $d \to \infty$ exists, then the conclusions should not depend on the choice of d, if it is reasonably large. In the FDA community there is a well grounded intuition that d should increase much slower than N, so asymptotically large d need not be very large in practice. It is also known that the rate at which d increases should depend on the manner in which the eigenvalues decay. We obtain specific conditions that formalize this intuition in the framework we consider. In more specific settings, contributions in this directions were made by Cardot et al. (2003) and Panaretos et al. (2010). The work of Cardot et al. (2003) is more closely related to our research: as part of the justification of their testing procedure, they establish conditions under which a limiting chi–square distribution with d degrees of freedom can be approximated by a normal distribution as $d = d(N) \to \infty$. Panaretos et al. (2010) are concerned with a test of the equality of the covariance operators in two samples of Gaussian curves. In the supplemental material, they derive asymptotics in which d is allowed to increase with the sample size. Our theory is geared toward testing the equality of mean functions, but we do not assume the normality of the functional observations, so we cannot use arguments that use the equivalence of independence and zero covariances. We develop a new technique based on the estimation of the Prokhorov–Lévy distance between the underlying processes and the corresponding normal partial sums.

The paper is organized as follows. In Section 2, we set the framework and state a general normal approximation result in Theorem 2.1. This result is then used in Sections 3 and 4 to derive, respectively, change–point and two–sample tests based on an increasing number of FPC's. Section 5 contains a small simulation study and an application to the annual Melbourne temperature curves. All proofs are collected in the appendices.

2. Uniform normal approximation

We consider functional observations $X_i(t)$, $t \in \mathcal{I}$, $i = 1, 2, \ldots, N$, defined over a compact interval $\mathcal{I}$. We can and shall assume without loss of generality that $\mathcal{I} = [0, 1]$. Throughout the paper, we use the notation $\int = \int_0^1$ and

$$\langle f, g \rangle = \int f(t) g(t) dt, \quad \|f\|^2 = \langle f, f \rangle.$$

All functions we consider will be elements of the Hilbert space L^2 of square integrable functions on $[0, 1]$.

In the testing problems that motivate this research, under the null hypothesis, the observations follow the model

(2.1)
$$X_i(t) = \mu(t) + Z_i(t), \quad 1 \le i \le N,$$

where $EZ_i(t) = 0$ and $\mu(t)$ is the common mean. We impose the following standard assumptions.

Assumption 2.1. $Z_1, Z_2, \ldots, Z_N$ *are independent and identically distributed.*

Assumption 2.2. $\int \mu^2(t)dt < \infty$ *and* $E\|Z_1\|^2 < \infty$.

Under these assumptions, the covariance function

$$\mathfrak{c}(t, s) = EZ_1(t)Z_1(s),$$

is square integrable on the unit square and therefore it has the representation

$$\mathfrak{c}(t, s) = \sum_{k=1}^{\infty} \lambda_k v_k(t)v_k(s),$$

where $\lambda_1 \ge \lambda_2 \ge \ldots$ are the eigenvalues and $v_1, v_2, \ldots$ are the orthonormal eigenfunctions of the covariance operator, i.e. they satisfy the integral equation

(2.2)
$$\lambda_j v_j(t) = \int \mathfrak{c}(t, s)v_j(s)ds.$$

One of the most important dimension reduction techniques of functional data analysis is to project the observations $X_1(t), \ldots, X_N(t)$ onto the space spanned by $v_1, \ldots, v_d$, the eigenfunctions associated with the d largest eigenvalues. Since the covariance function $\mathfrak{c}$, and therefore $v_1, \ldots, v_d$, are unknown, we use the empirical eigenfunctions $\hat{v}_1, \ldots, \hat{v}_d$ and eigenvalues $\hat{\lambda}_1 \ge \hat{\lambda}_2 \ge \ldots \ge \hat{\lambda}_d$ defined by

(2.3)
$$\hat{\lambda}_j \hat{v}_j(t) = \int \hat{\mathfrak{c}}_N(t, s)\hat{v}_j(s)ds,$$

where

$$\hat{\mathfrak{c}}_N(t, s) = \frac{1}{N} \sum_{i=1}^{N} \left(X_i(t) - \bar{X}_N(t)\right) \left(X_i(s) - \bar{X}_N(s)\right)$$

with $\bar{X}_N(t) = N^{-1} \sum_{i=1}^{N} X_i(t)$.

In this section, we require only two more assumptions, namely

Assumption 2.3. $\lambda_1 > \lambda_2 > \ldots$

Assumption 2.4. $E\|Z_1\|^3 < \infty$.

Assumption 2.3 is needed to ensure that the FPC's v_j are uniquely defined. In Theorem 2.1 it could, of course, be replaced by requiring only that the first d eigenvalues are positive and different, but since in the applications we let $d \to \infty$, we just assume that all eigenvalues are positive and distinct. If $\lambda_{d^*+1} = 0$ for some d^*, then the observations are in the linear span of $v_1, \ldots, v_{d^*}$, i.e. they are elements of a d^*–dimensional space,

so in this case we cannot consider $d = d(N) \to \infty$. Assumption 2.3 means that the observations are in an infinite–dimensional space. Assumption 2.4 is weaker than the usual assumption $E\|Z_1\|^4 < \infty$. As will be seen in the proofs, subtle arguments of the probability theory in Banach spaces are needed to dispense with the fourth moment. To state the main result of this section, define

$$\boldsymbol{\xi}_i = (\xi_{i,1}, \ldots, \xi_{i,d})^T \quad \text{and} \quad \xi_{i,j} = \lambda_j^{-1/2} \langle Z_i, v_j \rangle, \ \ 1 \le i \le N, \ 1 \le j \le d,$$

where $\cdot^T$ denotes the transpose of vectors and matrices. Set

$$(2.4) \qquad S_{j,N}(x) = \frac{1}{N^{1/2}} \sum_{i=1}^{\lfloor Nx \rfloor} \xi_{i,j}, \quad 0 \le x \le 1, \ \ 1 \le j \le d.$$

We now provide an approximation for the partial sum processes $S_{j,N}(x)$ defined in (2.4) with suitably constructed Wiener processes (standard Brownian motions).

Theorem 2.1. *If Assumptions 2.1, 2.3 and 2.4 hold, then for every N we can define independent Wiener processes $W_{1,N}, \ldots, W_{d,N}$ such that*

$$(2.5) \qquad P\left\{ \max_{1 \le j \le d} \sup_{0 \le x \le 1} |S_{j,N}(x) - W_{j,N}(x)| \ge N^{1/2 - 1/80} \right\}$$

$$\le c_* N^{-1/80} \left\{ d^{1/12} \left(\sum_{\ell=1}^d 1/\lambda_\ell \right)^{1/8} + \sum_{j=1}^d 1/\lambda_j^{3/2} \right\},$$

where c_* only depends on λ_1 and $E\|Z_1\|^3$.

The constant $1/80$ in (2.5) is not crucial, it is a result of our calculations. Theorem 2.1 is related to the results of Einmahl (1987, 1989) who obtained strong approximations for partial sums of independent and identically distributed random vectors with zero mean and with identity covariance matrix. In our setting, for any fixed d, the covariance matrix is not the identity, but this is not the central difficulty. The main value of Theorem 2.1 stems from the fact that it shows how the rate of the approximation depends on d; no such information is contained in the work of Einmahl (1987, 1989), who did not need to consider the dependence on d. The explicit dependence of the right hand side of (2.5) on d is crucial in the applications presented in the following sections in which the dimension of the projection space depends on the sample size N.

Very broadly speaking, Theorem 2.1 implies that in all reasonable statistics based on averaging the scores, even in those based on an increasing number of FPC's, the partial sums of scores can be replaced by Wiener processes to obtain a limit distribution. The right hand side of (2.5) allows us to derive assumptions on the eigenvalues required to obtain a specific result. Replacing the unobservable scores $\xi_{i,j}$ by the sample scores $\hat{\eta}_{i,j}$ is relatively easy. We will illustrate these ideas in Sections 3 and 4.

3. Change–point detection

Over the past four decades, the investigation of the asymptotic properties of partial sum processes has to a large extent been motivated by change–point detection procedures, and

this is the most natural application of Theorem 2.1. The research on the change–point problem in various contexts is very extensive, some aspects of the asymptotic theory are presented in Csörgő and Horváth (1997). Detection of a change in the mean function was studied by Berkes et al. (2009) who considered a procedure in which the number of the FPC's, d, was fixed, and the asymptotic distribution of the test statistic depended on d. We show in this section that it is possible to derive tests with a standard normal limiting distribution by allowing the d to depend on the sample size N.

We want to test whether the mean of the observations remained the same during the observation period, i.e. we test the null hypothesis

$$H_0: \quad EX_1(\cdot) = EX_2(\cdot) = \cdots = EX_N(\cdot)$$

("=" means equality in L^2). Under the null hypothesis, the X_i follow model (2.1) in which $\mu(\cdot)$ is an unknown common mean function under H_0. The alternative hypothesis is

$$H_A : \text{there is } k^* \in [1, 2, \ldots, N) \text{ such that}$$
$$EX_1(\cdot) = \cdots = EX_{k^*}(\cdot) \neq EX_{k^*+1}(\cdot) = \cdots = EX_N(\cdot).$$

Under H_A the mean changes at an unknown time k^*.

To derive a new class of tests, we introduce the process

$$\hat{Z}_N(u, x) = \frac{1}{d^{1/2}} \sum_{j=1}^{\lfloor du \rfloor} \left\{ \frac{1}{N} \left[\hat{S}_j(\lfloor Nx \rfloor) - x\hat{S}_j(N) \right]^2 - x(1-x) \right\}, \quad 0 \leq u, x \leq 1,$$

where

$$\hat{S}_j(k) = \frac{1}{\hat{\lambda}_j^{1/2}} \sum_{i=1}^{k} \hat{\eta}_{i,j}.$$

The process $\hat{Z}_N(u, x)$ contains the cumulative sums $\hat{S}_j(\lfloor Nx \rfloor) - x\hat{S}_j(N)$ which measure the deviation of the partial sums from their "trend" under H_0, and a correction term $x(1 - x)$ needed to ensure convergence as $d \to \infty$.

To obtain a limit which does not depend on any unknown quantities, we need to impose assumptions on the rate at which $d = d(N)$ increases with N. Intuitively, the assumptions below state that d is much smaller than the sample size N, the d largest eigenvalues are not too small, and that the difference between the consecutive eigenvalues tends to zero slowly. Very broadly speaking, these assumptions mean that the distribution of the observations must sufficiently fill the whole infinite–dimensional space L^2.

Assumption 3.1. $d = d(N) \to \infty$

Assumption 3.2. $(d \log N)^{1/2} N^{-1/80} \to 0,$

Assumption 3.3. $d^{1/12} N^{-1/80} \left(\sum_{j=1}^{d} 1/\lambda_j \right)^{1/8} \to 0.$

Assumption 3.4. $N^{-1/80} \sum_{j=1}^{d} 1/\lambda_j^{3/2} \to 0.$

Assumption 3.5.

$$\frac{1}{d^{1/2}N^{1/3}} \sum_{j=1}^{d} \frac{1}{\lambda_j \zeta_j} \to 0,$$

where $\zeta_1 = \lambda_2 - \lambda_1$, $\zeta_j = \min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})$, $j \geq 2$.

With these preparations, we can state the main result of this section.

Theorem 3.1. *If Assumptions 2.1–2.3 and 3.1–3.5 are satisfied, then*

$$\hat{Z}_N(u,x) \to \Gamma(u,x) \quad in \ \mathcal{D}[0,1]^2,$$

where $\Gamma(u,x)$ is a mean zero Gaussian process with

$$E[\Gamma(u,x)\Gamma(v,y)] = 2ux^2(1-y)^2, \quad 0 \leq u \leq v \leq 1, \ \ 0 \leq x \leq y \leq 1.$$

One can verify by computing the covariance functions that

(3.1) $\{\Gamma(u,x), \ 0 \leq u,x \leq 1\} \stackrel{\mathcal{D}}{=} \{\sqrt{2}(1-x)^2 W(u,x^2/(1-x)^2), \ 0 \leq u,x \leq 1\},$

where $\{W(v,y), v,y \geq 0\}$ is a bivariate Wiener process, i.e. $W(v,y)$ is a Gaussian process with $EW(v,y) = 0$ and $E[W(v,y)W(v',y')] = \min(v,v')\min(y,y')$. Representation (3.1) means that continuous functionals of the process $\Gamma(\cdot,\cdot)$ can be simulated with arbitrary precision, so Monte Carlo tests can be used. One would choose the number of projections in the CUSUM procedure such that the test would give the largest rejection if the alternative holds. The statistic $\max_u \max_x |\hat{Z}_N(u,x)|$ is maximizing the CUSUM statistics $\max_x |\hat{Z}_N(k/d,x)|$, where $k = 1, 2, \ldots, d$ projections are used. It is however possible to obtain a number of simple asymptotic tests by examining closer the structure of the process $\Gamma(\cdot,\cdot)$. We list some of them in Corollary 3.1, and we will see in Section 5 that the Cramér-von-Mises type tests have very good finite sample properties. Let B denote a Brownian bridge and define

$$\mu_0 = E\left(\sup_{0 \leq x \leq 1} B^2(x)\right) \quad and \quad \sigma_0^2 = \text{var}\left(\sup_{0 \leq x \leq 1} B^2(x)\right).$$

Corollary 3.1. *If the assumptions of Theorem 3.1 are satisfied, then*

(3.2) $\dfrac{1}{d^{1/2}\sigma_0} \left\{ \displaystyle\sum_{j=1}^{d} \sup_{0 \leq x \leq 1} \dfrac{1}{N} \left(\hat{S}_j(\lfloor Nx \rfloor) - x\hat{S}_j(N)\right)^2 - d\mu_0 \right\} \stackrel{\mathcal{D}}{\to} N(0,1),$

(3.3) $\dfrac{1}{(d/45)^{1/2}} \left\{ \displaystyle\sum_{j=1}^{d} \dfrac{1}{N} \int (\hat{S}_j(\lfloor Nx \rfloor) - x\hat{S}_j(N))^2 dx - \dfrac{d}{6} \right\} \stackrel{\mathcal{D}}{\to} N(0,1),$

(3.4) $\dfrac{1}{(d/8)^{1/2}} \left\{ \displaystyle\sup_{0 \leq x \leq 1} \sum_{j=1}^{d} \dfrac{1}{N} (\hat{S}_j(\lfloor Nx \rfloor) - x\hat{S}_j(N))^2 - \dfrac{d}{4} \right\} \stackrel{\mathcal{D}}{\to} N(0,1),$

where $N(0,1)$ stands for a standard normal random variable.

We conclude this section with two examples which show that Assumptions 3.2–3.5 hold under both power law and exponential decay of the eigenvalues.

Example 3.1. *If the eigenvalues satisfy*

$$\lambda_j = \frac{c_1}{(j - c_2)^\alpha} + o\left(\frac{1}{j^{\alpha+1}}\right), \quad as \quad j \to \infty,$$

with some $c_1 > 0$, $0 \le c_2 < 1$ and $\alpha > 0$, then Assumptions 3.2–3.5 hold if $d/(\log N)^\beta \to 0$ with some $\beta > 0$.

Under the conditions of Example 3.1, one could choose $d_N = O(N^\zeta)$, where ζ depends on α. In case of a fixed samplle size N, the power of the test would decrease if d_N is too large. Hence we recommend choosing $d_N \approx (\log N)^\beta$, where $\beta > 0$ can be arbitrarily chosen.

Example 3.2. *If the eigenvalues satisfy*

$$\lambda_j = c_0 e^{-\alpha j} + o(e^{-\alpha j}), \quad as \quad j \to \infty,$$

with some $c_0 > 0$ and $\alpha > 0$, then Assumptions 3.2–3.5 hold if $d/(\log \log N)^\beta \to 0$ with some $\beta > 0$.

4. Two–sample problem

The two–sample problem for functional data was perhaps first discussed in depth by Benko et al. (2009) who were motivated by a problem related to implied volatility curves. It has recently attracted a fair amount of attention motivated by problems arising in space physics, see Horváth et al. (2009), genetics, see Panaretos et al. (2010), and finance, see Horváth et al. (2012). The above list does not include many other important contributions. In its simplest, but most important form, it is about testing if curves obtained from two populations have the same mean functions. The most direct approach, developed into a bootstrap procedure by Benko et al. (2009), is to look at the norm of the difference of the estimated mean functions. In this section, we show that the normal approximation of Section 2 leads to an asymptotic test whose limit distribution is standard normal.

Suppose we have two random samples of functions: $X_1, \ldots, X_N$ and $Y_1, \ldots, Y_M$. We assume the X sample satisfies (2.1) and Assumptions 2.1, 2.2 and 2.4. Similarly, the Y sample is a location model given by

$$(4.1) \qquad\qquad Y_i(t) = \mu_*(t) + Q_i(t), \quad 1 \le i \le M,$$

where $\mu_*(t)$ is the common mean of the Y sample and $EQ_i(t) = 0$. As in the case of the X sample, the Y sample satisfies the following conditions:

Assumption 4.1. *$Q_1, Q_2, \ldots, Q_M$ are independent and identically distributed.*

Assumption 4.2. *$\int \mu_*^2(t)dt < \infty$ and $E\|Q_1\|^3 < \infty$.*

Assumption 4.2 yields that

$$\mathfrak{c}_*(t, s) = EQ_1(t)Q_1(s)$$

is a square integrable function on the unit square.

In this section we are interested in testing the null hypothesis

$$H_0^*: \quad \mu(\cdot) = \mu_*(\cdot).$$

The statistical inference to test H_0 is based on the difference $\bar{X}_N - \bar{Y}_M$, where $\bar{X}_N$ and $\bar{Y}_M$ denote the sample means. We assume

Assumption 4.3.

$$\frac{N}{M} = \lambda + O(N^{-1/4}) \quad as \quad \min(M, N) \to \infty$$

with some $0 < \lambda < \infty$.

Now we define the pooled covariance function

$$\mathfrak{c}_P(t, s) = \mathfrak{c}(t, s) + \lambda \mathfrak{c}_*(t, s).$$

Since $\mathfrak{c}_P(t, s)$ is a positive–definite, symmetric, square integrable function, there are real numbers $\kappa_1 \geq \kappa_2 \geq \ldots$ and orthonormal functions $u_1, u_2, \ldots$ satisfying

$$\kappa_i u_i(t) = \int \mathfrak{c}_P(t, s) u_i(s) ds, \quad i = 1, 2, \ldots.$$

We wish to project $\bar{X}_N - \bar{Y}_M$ into the space spanned by $u_1, \ldots, u_d$, where $d = d(N) \to \infty$, so similarly to Assumption 2.3 we require

Assumption 4.4. $\kappa_1 > \kappa_2 > \kappa_3 > \ldots$

Assumption 4.5.

$$N^{-3/32} d^{1/4} \left(\sum_{\ell=1}^{d} 1/\kappa_\ell \right)^{3/8} \to 0.$$

Our test statistic is

$$D_{N,M} = \sum_{i=1}^{d} N \langle \bar{X}_N - \bar{Y}_M, u_i \rangle^2 / \kappa_i.$$

As in Section 3, we need additional assumptions balancing the rate of growth of $d = d(N)$ and the rate of decay of the κ_ℓ and the differences between them.

Assumption 4.6.

$$\frac{1}{d^{1/2} N^{1/4}} \sum_{\ell=1}^{d} \frac{1}{\kappa_\ell^2} \to 0 \quad and \quad \frac{1}{d^{1/2} N^{1/4}} \sum_{\ell=1}^{d} \frac{1}{\kappa_\ell \iota_\ell} \to 0,$$

where $\iota_1 = \kappa_2 - \kappa_1$, $\iota_\ell = \min(\iota_{\ell-1} - \iota_\ell, \iota_\ell - \iota_{\ell+1})$, $\ell \geq 2$.

Since $u_1, u_2, \ldots$ are unknown, we replace them with the corresponding empirical eigenfunctions $\hat{u}_1, \hat{u}_2, \ldots$ defined by the integral operator

$$\hat{\kappa}_i \hat{u}_i(t) = \int \hat{\mathfrak{c}}_P(t, s) \hat{u}_i(s) ds, \quad i = 1, 2, \ldots,$$

α	0.01	0.05	0.10
	0.109256	0.0726292	0.0578267

TABLE 5.1. Critical values for the distribution of (5.3).

where $\hat{\kappa}_1 \geq \hat{\kappa}_2 \geq \ldots$ and

$$\hat{\mathfrak{c}}_P(t, s) = \hat{\mathfrak{c}}_N(t, s) + \frac{N}{M}\hat{\mathfrak{c}}_{*M}(t, s),$$

with

$$\hat{\mathfrak{c}}_{*M}(t, s) = \frac{1}{M}\sum_{\ell=1}^{M}(Y_\ell(t) - \bar{Y}_M(t))(Y_\ell(s) - \bar{Y}_M(s)).$$

The empirical version of $D_{N,M}$ is

$$\widehat{D}_{N,M} = \sum_{i=1}^{d} N\langle \bar{X}_N - \bar{Y}_M, \hat{u}_i\rangle^2/\hat{\kappa}_i.$$

Theorem 4.1. *If H_0^*, Assumptions 2.1, 2.2 and 4.1–4.6 hold, then*

$$(2d)^{-1/2}(\widehat{D}_{N,M} - d) \quad \overset{\mathcal{D}}{\to} \quad N(0, 1),$$

where $N(0, 1)$ stands for a standard normal random variable.

5. **A small simulation study and a data example**

The main contribution of this paper lies in the statistical theory, but it is of interest to check if the new tests derived in Sections 3 and 4 perform well in finite samples. We report the results for the test based on Theorem 3.1 in some detail, as it utilizes the convergence of the two–parameter process in full force, and such an approach has not been used before. We also comment on the tests based on Corollary 3.1 and Theorem 4.1. We conclude this section with an illustrative data example.

The simulated data which satisfy the null hypotheses of Sections 3 and 4 are generated as independent Brownian motions on the interval $[0, 1]$. We generate them by using iid normal increments on 1,000 equispaced points in $[0, 1]$ (random walk approximation). (Example 3.1 shows that for the Brownian motion the assumptions of Theorem 3.1 are satisfied.) Alternatives are obtained by adding the curve $at(1 - t)$ after a change–point or to the observations in the second sample. The parameter a regulates the size of the change or the difference in the means in two samples.

Many tests can be obtained from Theorem 3.1 by applying functionals continuous on $\mathcal{D}[0, 1]^2$. It is not our objective to provide a systematic comparison, we consider only the test based on the weak convergence

(5.1) $$\int_0^1\int_0^1 \hat{Z}_N^2(u, x)dudx \;\to\; \int_0^1\int_0^1 \Gamma^2(u, x)dudx.$$

$N = 100$

d	$\alpha = 0.05$		$\alpha = 0.1$	
	$\hat{p}$	$[a, b]$	$\hat{p}$	$[a, b]$
2	0.047	[0.0360,0.0580]	0.058	[0.0458,0.0702]
3	0.056	[0.0440,0.0680]	0.074	[0.0604,0.0876]
4	0.060	[0.0476,0.0724]	0.081	[0.0668,0.0952]
5	0.059	[0.0467,0.0713]	0.089	[0.0742,0.1038]
6	0.057	[0.0449,0.0691]	0.089	[0.0742,0.1038]
7	0.056	[0.0440,0.0680]	0.089	[0.0742,0.1038]
8	0.059	[0.0467,0.0713]	0.091	[0.0760,0.1060]
9	0.051	[0.0396,0.0624]	0.090	[0.0751,0.1049]
10	0.050	[0.0387,0.0613]	0.082	[0.0677,0.0963]
11	0.054	[0.0422,0.0658]	0.083	[0.0687,0.0973]
12	0.057	[0.0449,0.0691]	0.079	[0.0650,0.0930]
13	0.059	[0.0467,0.0713]	0.075	[0.0613,0.0887]
14	0.057	[0.0449,0.0691]	0.076	[0.0622,0.0898]
15	0.056	[0.0440,0.0680]	0.075	[0.0613,0.0887]

$N = 200$

d	$\alpha = 0.05$		$\alpha = 0.1$	
	$\hat{p}$	$[a, b]$	$\hat{p}$	$[a, b]$
2	0.039	[0.0289,0.0491]	0.055	[0.0431,0.0669]
3	0.048	[0.0369,0.0591]	0.070	[0.0567,0.0833]
4	0.049	[0.0378,0.0602]	0.075	[0.0613,0.0887]
5	0.053	[0.0413,0.0647]	0.076	[0.0622,0.0898]
6	0.057	[0.0449,0.0691]	0.085	[0.0705,0.0995
7	0.057	[0.0449,0.0691]	0.085	[0.0705,0.0995]
8	0.053	[0.0413,0.0647]	0.085	[0.0705,0.0995]
9	0.051	[0.0396,0.0624]	0.083	[0.0687,0.0973]
10	0.051	[0.0378,0.0602]	0.081	[0.0668,0.0952]
11	0.054	[0.0496,0.0624]	0.083	[0.0687,0.0973]
12	0.052	[0.0405,0.0635]	0.086	[0.0714,0.1006]
13	0.050	[0.0387,0.0613]	0.087	[0.0723,0.1017]
14	0.054	[0.0422,0.0658]	0.086	[0.0714,0.1006]
15	0.052	[0.0405,0.0635]	0.079	[0.0650,0.0930]

TABLE 5.2. Empirical sizes and 90% confidence intervals for the probability of rejection for the change–point test based on convergence (5.1).
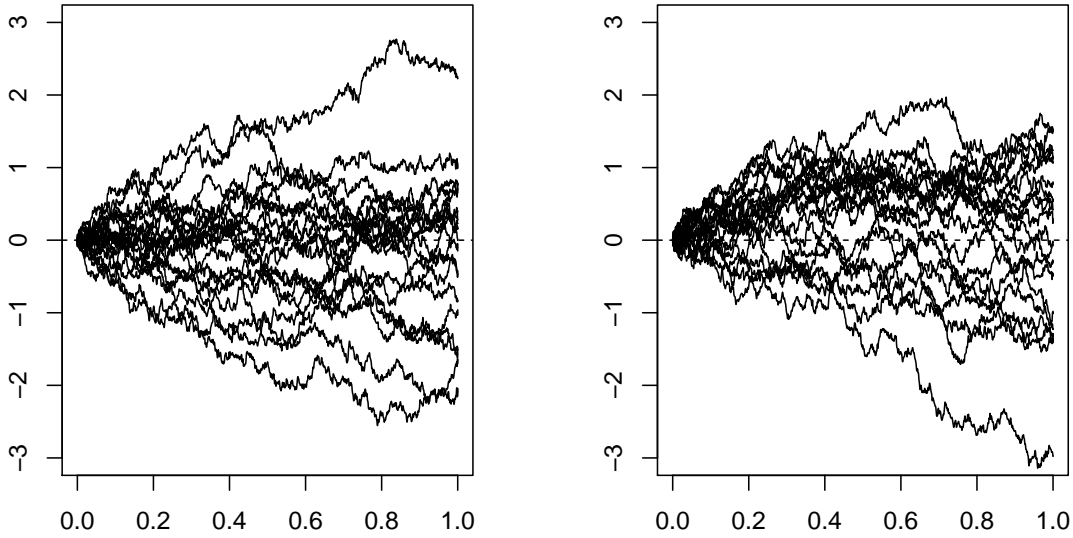
FIGURE 3. Left panel: 20 realizations of the Brownian motion; Right panel: independent 20 realizations of the Brownian motion with the curve $at(1-t), a = 1.5$ added.

To compute the critical values, we use the following representation of the limit

$$(5.2) \qquad \int_0^1 \int_0^1 \Gamma^2(u, x) du dx \overset{\mathcal{D}}{=} \sum_{1 \le k, \ell < \infty} \lambda_k \nu_\ell N_{k,\ell}^2.$$

In (5.2), the $\lambda_k = (\pi(k - 1/2))^{-2}$ are the eigenvalues of the Wiener process, the ν_ℓ are the eigenvalues of the covariance operator with kernel $2(\min(s, t) - st)^2$, and $\{N_{k,\ell}\}$ is an array of independent standard normal random variables. The critical values were determined for a truncated version of the right–hand side of (5.2) with truncation level 49, i.e. for

$$(5.3) \qquad \sum_{1 \le k, \ell \le 49} \lambda_k \nu_\ell N_{k,\ell}^2.$$

Since the eigenvalues ν_ℓ are difficult to determine explicitly, they were calculated numerically using the R package `fda`, cf.Ramsay et al. (2009). The simulated critical values based on 100,000 replications of (5.3) are provided in Table 5.1.
Table 5.2 shows the empirical sizes $\hat{p}$, i.e. the fraction of rejections, as well as asymptotic 90% confidence intervals

$$(5.4) \qquad \left[\hat{p} - 1.654 \sqrt{\frac{\hat{p}(1 - \hat{p})}{R}}, \quad \hat{p} + 1.654 \sqrt{\frac{\hat{p}(1 - \hat{p})}{R}} \right].$$

$N = 100$

d	$a = 1$		$a = 1.5$	
	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$
2	0.168	0.192	0.356	0.398
3	0.456	0.517	0.819	0.851
4	0.501	0.564	0.843	0.875
5	0.496	0.564	0.855	0.887
6	0.481	0.552	0.847	0.883
7	0.473	0.543	0.843	0.881
8	0.465	0.530	0.834	0.874
9	0.461	0.519	0.823	0.870
10	0.453	0.504	0.812	0.859
11	0.441	0.501	0.802	0.853
12	0.431	0.496	0.793	0.844
13	0.420	0.484	0.791	0.834
14	0.400	0.472	0.782	0.822
15	0.388	0.467	0.767	0.817

$N = 200$

d	$a = 1$		$a = 1.5$	
	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$
2	0.327	0.370	0.620	0.660
3	0.784	0.814	0.984	0.991
4	0.808	0.849	0.988	0.994
5	0.823	0.860	0.992	0.994
6	0.825	0.863	0.991	0.996
7	0.819	0.864	0.992	0.994
8	0.814	0.859	0.990	0.994
9	0.802	0.846	0.990	0.993
10	0.791	0.837	0.990	0.993
11	0.766	0.830	0.988	0.992
12	0.754	0.821	0.987	0.992
13	0.740	0.800	0.987	0.991
14	0.734	0.794	0.987	0.991
15	0.726	0.787	0.986	0.990

TABLE 5.3. Power of the test based on convergence (5.1). The change–point is at $k^* = \lfloor N/2 \rfloor$.

for the probability p of rejection. The entries are based on $R = 1,000$ replications. The table shows that the test based on convergence (5.1) has correct empirical size at the 5% level and is a bit too conservative at the 10% level. However even at the 10% level the empirical sizes for $d \geq 3$ are not significantly different; they all fall into each others 90% confidence intervals. This illustrates the main point that for the tests that use the asymptotics with $d \to \infty$ developed in the paper, selecting d is not essential; every sufficiently large d gives the same conclusion on the significance.

The empirical power of the test is reported in Table 5.3. Again, for $d \geq 3$, the power remains statistically the same. We note that the change in mean equal to the function $at(1-t)$ with $a = 1.5$ is fairly small if the "noise curves" are Brownian motions. This is illustrated in Figure 3 which shows 20 Brownian motions in the left panel and another independent sample of 20 Brownian motions with the curve $at(1-t), a = 1.5$ added. If one knows that this curve was added, one can discern it in the plot in the right panel, but the difference would have been much less obvious if individual curves were observed, as in the change–point setting relevant to Table 5.3.

Regarding Corollary 3.1, we found out that the test based on convergence (3.3) has empirical size only slightly higher than nominal (about 1% at 5% level). For $d \geq 3$, the empirical size does not depend on d. The test based on (3.4) severely overrejects for $N = 100$, and we do not recommend it. The test based on Theorem 4.1 overrejects by about 2% at the 5% level, and by about 1% at the 10% level. The power of the test is above 95% for $N, M = 100$ and $a = 1.0$, and practically 100% for larger a or N, M. For $d \geq 2$, the rejection probabilities do not depend on d.

Change–point analysis of annual temperature profiles. The goal of this section is to illustrate the application of the change–point test based on convergence (5.1). Change–point analysis is an important field of statistics with a large number of applications, the recent monographs of Chen and Gupta (2011) and Basseville et al. (2012) provide numerous references. The change–point problem in the context of functional data has also received some attention, we refer to Horváth and Kokoszka (2012) for the references, Aston and Kirch (2012) report some most recent research.

The data set we study consists of 156 years (1856-2011) of minimum daily temperatures in Melbourne. These data are available at `www.bom.gov.au` (the Australian Bureau of Meteorology website). The original data can be viewed as 156 curves with 365 measurements on each curve. We converted them to functional objects in `R` using 49 Fourier basis functions. Five consecutive functions are shown in Figure 4. It is important to emphasize the difference between the data we use and the Canadian temperature data made popular by the books of Ramsay and Silverman (2005) and Ramsay et al. (2009). The Canadian temperature curves are the curves at 35 locations in Canada obtained by averaging annual temperature over forty years. Since each such curve is an average of forty curves like those shown in Figure 4, those curves are much smoother, and the first two FPC's are sufficient to describe their variability. Even after smoothing with 49 Fourier functions, the annual temperature curves exhibit noticeable year to year variability, and a larger number of FPC's is needed to capture it, see Table 5.4. The goals
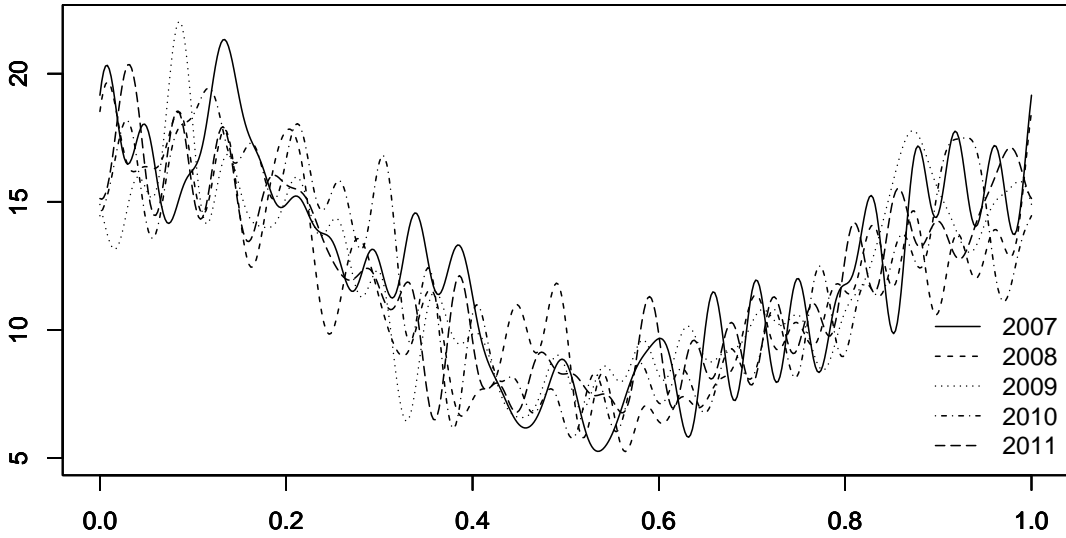
FIGURE 4. Five annual temperature curves represented as functional objects.

of our analysis are also different from those of Ramsay and Silverman (2005). We are interested in detecting a change in the mean function using a sequence of noisy curves; the examples in Ramsay and Silverman (2005) used the averaged curves to describe static regression type dependencies between climatic variables.

The analysis proceeds through the usual binary segmentation procedure. The test is first applied to the whole data set. If the P–value is small, the change–point is estimated as

$$\hat{\theta}_N = \inf\{k : I_N(k) = \sup_{1 \le j \le N} I_N(j)\},$$

where

$$I_N(\ell) = \frac{1}{d^2} \sum_{i=1}^{d-1} \left(\sum_{j=1}^{i} \left\{ \frac{1}{N} \left[\hat{S}_j(\ell) - \frac{\ell}{N} \hat{S}_j(N) \right]^2 - \frac{\ell}{N} \left(\frac{N-\ell}{N} \right) \right\} \right)^2.$$

(I_N is a discretization of $\hat{Z}_N$.) The test is then applied to the two segments, and the procedure continues until no change–points are detected. In practice, a procedure of this type detects only a few change–points (four in our case), so the problems of multiple testing are not an issue. We applied the test using many values of d, and we were pleased to see that the final segmentation does not depend on d. Table 5.5 shows the outcome. The estimated change–points are the years 1892, 1960, 1967, 1996. It is clear that the change–point model is not an exact climatological model for the evolution of annual temperature curves, but it is popular in climate studies, see e.g. Gallagher et al. (2012), as it allows us to attach statistical significance to conclusions and provides
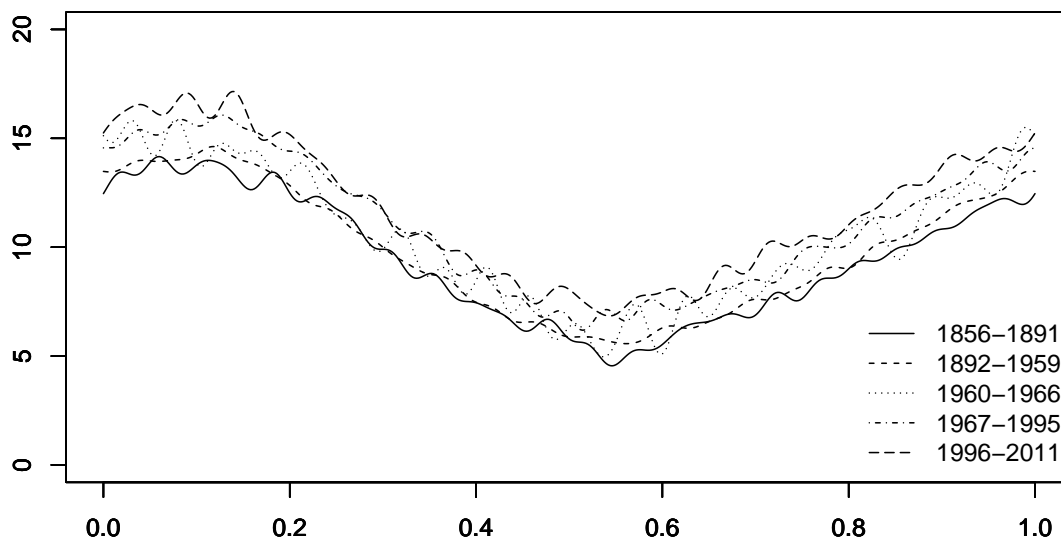
FIGURE 5. Average temperature functions in the estimated partition segments.

periods of approximately constant mean temperature profiles. In this light, the weak evidence for a change–point in 1967 could be viewed as indicating an accelerated change in the period 1960–1995. The estimated mean temperature curves over the segments of approximately constant mean are shown in Figure 5. An increasing pattern of the mean temperature is seen; the mean curve shifted upwards by about two degrees Celsius over the last 150 years. This could be due to the conjectured global temperature increase or the urbanization of the Melbourne area, or a combination of both. A discussion of such issues is however beyond the intended scope of this paper.

k	1	2	3	4	5	6	7	8
$\hat{\lambda}_k$	0.7151	0.1469	0.1295	0.1154	0.1046	0.1021	0.0944	0.0868
f_p	0.2248	0.2711	0.3118	0.3480	0.3809	0.4130	0.4427	0.4700
k	9	10	11	12	13	14	15	16
$\hat{\lambda}_k$	0.0845	0.0833	0.0758	0.0732	0.0726	0.0687	0.0661	0.0641
f_p	0.4966	0.5228	0.5466	0.5696	0.5925	0.6141	0.6349	0.6550
k	17	18	19	20	21	22	23	24
$\hat{\lambda}_k$	0.0620	0.0586	0.0559	0.0559	0.0534	0.0508	0.0472	0.0463
f_p	0.6745	0.6930	0.7105	0.7281	0.7449	0.7609	0.7757	0.7903
k	25	26	27	28	29	30	31	32
$\hat{\lambda}_k$	0.0440	0.0427	0.0426	0.0400	0.0377	0.0367	0.0359	0.0325
f_p	0.8041	0.8175	0.8309	0.8435	0.8553	0.8669	0.8782	0.8884
k	33	34	35	36	37	38	39	40
$\hat{\lambda}_k$	0.0320	0.0299	0.0281	0.0274	0.0252	0.0248	0.0228	0.0211
f_p	0.8985	0.9079	0.9167	0.9253	0.9332	0.9410	0.9482	0.9548
k	41	42	43	44	45	46	47	48
$\hat{\lambda}_k$	0.0207	0.0201	0.0188	0.0171	0.0166	0.0163	0.0129	0.0114
f_p	0.9614	0.9677	0.9736	0.9790	0.9842	0.9893	0.9934	0.9969

TABLE 5.4. Eigenvalues and percentage of variance explained by the first k eigenvalues, i.e. $f_k = \sum_{i=1}^{k} \hat{\lambda}_i / \sum_{j=1}^{N} \hat{\lambda}_j$, for $k = 1, 2, \ldots, 49$.

It.	Segment	Estimated	P-value			
		change–point	$d = 3$	$d = 4$	$d = 5$	$d = 6$
1	1856-2011	1960	0.0000	0.0000	0.0000	0.0000
2	1856-1959	1892	0.0000	0.0000	0.0000	0.0000
3	1856-1891	—	0.1865	0.2323	0.3524	0.4822
4	1892-1959	—	0.9522	0.9690	0.9256	0.6561
5	1960-2011	1996	0.0000	0.0000	0.0000	0.0000
6	1960-1995	1967	0.0013	0.0011	0.0025	0.0017
7	1960-1966	—	0.9568	0.9549	0.9818	0.9935
8	1967-1995	—	0.2927	0.4305	0.1786	0.1348
9	1996-2011	—	0.4285	0.5345	0.6413	0.7365
It.	Segment	Estimated	P-value			
		change–point	$d = 7$	$d = 8$	$d = 9$	$d = 10$
1	1856-2011	1960	0.0000	0.0000	0.0000	0.0000
2	1856-1959	1892	0.0000	0.0000	0.0000	0.0000
3	1856-1891	—	0.4235	0.4325	0.4901	0.5667
4	1892-1959	—	0.4646	0.4348	0.4696	0.5068
5	1960-2011	1996	0.0000	0.0000	0.0000	0.0000
6	1960-1995	1967	0.0026	0.0038	0.0058	0.0067
7	1960-1966	—	0.9992	—	—	—
8	1967-1995	—	0.1245	0.0690	0.0571	0.0586
9	1996-2011	—	0.8243	0.9118	0.9618	0.9779

TABLE 5.5. Segmentation procedure of the data into periods with constant mean function

APPENDIX A. **Proof of Theorem 2.1**

We start with some elementary properties of the projections $\xi_{i,j}$. Let $|\cdot|$ denote the Euclidean norm of vectors.

Lemma A.1. *If Assumptions 2.1, 2.3 and 2.4 hold, then*

(A.1)
$$E\boldsymbol{\xi}_1 = \mathbf{0},$$

(A.2)
$$E\boldsymbol{\xi}_1\boldsymbol{\xi}_1^T = \mathbf{I}_d,$$

where $\mathbf{I}_d$ is the $d \times d$ identity matrix. Moreover,

(A.3)
$$E|\boldsymbol{\xi}_1|^3 \leq E\|Z_1\|^3 \left(\sum_{j=1}^{d} 1/\lambda_j\right)^{3/2}$$

and for all $1 \leq j \leq d$

(A.4)
$$E|\xi_{1,j}|^3 \leq E\|Z_1\|^3/\lambda_j^{3/2}.$$

Proof. Since $EZ_1(t) = 0$, the relation in (A.1) is obvious. The orthonormal functions v_k and v_ℓ satisfy (2.2), so we get

$$E\xi_{i,k}\xi_{i,\ell} = \frac{1}{(\lambda_k\lambda_\ell)^{1/2}} \iint \mathfrak{c}(t,s)v_k(s)v_\ell(s)dtds = \begin{cases} 0, & \text{if } k \neq \ell \\ 1, & \text{if } k = \ell, \end{cases}$$

proving (A.2). Using the definition of the Euclidean norm and the Cauchy–Schwarz inequality we conclude

$$|\boldsymbol{\xi}_1|^3 = \left(\sum_{j=1}^{d}\langle Z_1, v_j\rangle^2/\lambda_j\right)^{3/2} \leq \left(\sum_{j=1}^{d}\|Z_1\|^2\|v_j\|^2/\lambda_j\right)^{3/2} = \|Z_1\|^3\left(\sum_{j=1}^{d}1/\lambda_j\right)^{3/2},$$

since $\|v_j\| = 1$. Taking the expected value of the equation above we obtain (A.3). Clearly,

$$E|\xi_{1,j}|^3 = \lambda_j^{-3/2}E|\langle Z_1, v_j\rangle|^3 \leq \lambda_j^{-3/2}E\|Z_1\|^3.$$

$\square$

The next lemma plays a central role in the proof of Theorem 2.1.

Lemma A.2. *If Assumptions 2.1, 2.3 and 2.4 hold, then for all n we can define independent identically distributed standard normal vectors $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_n$ in R^d such that*

$$P\left\{\left|\sum_{i=1}^{n}\boldsymbol{\xi}_i - \sum_{i=1}^{n}\boldsymbol{\gamma}_i\right| \geq cn^{3/8}d^{1/4}(E|\boldsymbol{\xi}_1|^3 + E|\boldsymbol{\gamma}_1|^3)^{1/4}\right\} \leq cn^{-1/8}d^{1/4}(E|\boldsymbol{\xi}_1|^3 + E|\boldsymbol{\gamma}_1|^3)^{1/4},$$

where c is an absolute constant.

Proof. The result is a consequence of Theorem 6.4.1 on p. 207 of Senatov (1998) and the corollary to Theorem 11 in Strassen (1965).

$\square$

We note that

$$(A.5) \qquad (E|\boldsymbol{\xi}_1|^3 + E|\boldsymbol{\gamma}_1|^3)^{1/4} \leq (E|\boldsymbol{\xi}_1|^3)^{1/4} + (E|\boldsymbol{\gamma}_1|^3)^{1/4}.$$

Also, since $|\boldsymbol{\gamma}_1|^2$ is the sum of the squares of d independent standard normal random variables, Minkowski's inequality implies

$$(A.6) \qquad E|\boldsymbol{\gamma}_1|^3 \leq c_1 d^{3/2},$$

with some constant c_1, and clearly

$$(A.7) \qquad d^{3/2} \leq \lambda_1^{3/2} \left(\sum_{\ell=1}^d 1/\lambda_\ell \right)^{3/2}.$$

Combining Lemma A.2 with (A.5)–(A.7), we conclude that

$$(A.8) \quad P\left\{ \left| \sum_{i=1}^n \boldsymbol{\xi}_i - \sum_{i=1}^n \boldsymbol{\gamma}_i \right| \geq c_2 n^{3/8} d^{1/4} \left(\sum_{j=1}^d 1/\lambda_j \right)^{3/8} \right\} \leq c_2 n^{-1/8} d^{1/4} \left(\sum_{j=1}^d 1/\lambda_j \right)^{3/8},$$

where c_2 does not depend on d.

In the next lemma we provide an upper bound for the variance of $\sum_i^n (\xi_{i,j} - \gamma_{i,j})$, where $\boldsymbol{\gamma}_i = (\gamma_{i,1}, \dots, \gamma_{i,d})^T$ is defined in Lemma A.2.

Lemma A.3. *If Assumptions 2.1, 2.3 and 2.4 hold, then for any $1 \leq j \leq d$ we get*

$$E \left(\sum_{i=1}^n \xi_{i,j} - \sum_{i=1}^n \gamma_{i,j} \right)^2 \leq c_3 n^{23/24} \frac{1}{\lambda_j} \left(d^{1/4} \left(\sum_{\ell=1}^d 1/\lambda_\ell \right)^{3/8} \right)^{1/3},$$

where c_3 does not depend on d.

Proof. Let

$$U_n(j) = n^{-1/2} \sum_{i=1}^n (\xi_{i,j} - \gamma_{i,j}) \quad \text{and} \quad r_n = c_2 n^{-1/8} d^{1/4} \left(\sum_{\ell=1}^d 1/\lambda_\ell \right)^{3/8}.$$

First we write

$$EU_n^2(j) = E[U_n^2(j)I\{|U_n(j)| \leq r_n\}] + E[U_n^2(j)I\{|U_n(j)| > r_n\}]$$

$$\leq r_n^2 + \frac{2}{n} E\left[\left(\sum_{i=1}^n \xi_{i,j} \right)^2 I\{|U_n(j)| > r_n\} \right] + \frac{2}{n} E\left[\left(\sum_{i=1}^n \gamma_{i,j} \right)^2 I\{|U_n(j)| > r_n\} \right].$$

Using Hölder's inequality we get that

$$E\left[\left(\sum_{i=1}^n \xi_{i,j} \right)^2 I\{|U_n(j)| > r_n\} \right] \leq E\left[\left| \sum_{i=1}^n \xi_{i,j} \right|^3 \right]^{2/3} \left[P\{|U_n(j)| > r_n\} \right]^{1/3}$$

$$\leq E\left[\left| \sum_{i=1}^n \xi_{i,j} \right|^3 \right]^{2/3} r_n^{1/3}$$

by (A.8). Applying now Rosenthal's inequality (cf. Petrov (1995), p. 59) we obtain

$$E\left|\sum_{i=1}^{n}\xi_{i,j}\right|^{3} \leq c_4\left\{\sum_{i=1}^{n}E|\xi_{i,j}|^{3} + \left(\sum_{i=1}^{n}E\xi_{i,j}^{2}\right)^{3/2}\right\},$$

where c_4 is an absolute constant. Hence

$$E\left|\sum_{i=1}^{n}\xi_{i,j}\right|^{3} \leq c_5\{n\lambda_j^{-3/2} + n^{3/2}\} \leq c_6(n/\lambda_j)^{3/2}$$

and therefore

$$E\left[\left(\sum_{i=1}^{n}\xi_{i,j}\right)^{2}I\{|U_n(j)| > r_n\}\right] \leq c_7(n/\lambda_j)r_n^{1/3}$$

$$\leq c_8 n^{23/24}\frac{1}{\lambda_j}\left(d^{1/4}\left(\sum_{\ell=1}^{d}1/\lambda_\ell\right)^{3/8}\right)^{1/3}.$$

Following the previous arguments one can show that

$$E\left[\left(\sum_{i=1}^{n}\gamma_{i,j}\right)^{2}I\{|U_n(j)| > r_n\}\right] \leq c_9 n^{23/24}\frac{1}{\lambda_j}\left(d^{1/4}\left(\sum_{\ell=1}^{d}1/\lambda_\ell\right)^{3/8}\right)^{1/3}.$$

The constants c_8 and c_9 do not depend on d. Since in view of Assumption 3.3, nr_n^2 is smaller than the latter rates, this completes the proof of Lemma A.3.   □

Proof of Theorem 2.1. We use a blocking argument to construct a Wiener process which is close to the partial sums $\sum_{1\leq i\leq k}\xi_{i,j}, 1 \leq k \leq N, 1 \leq j \leq d$. Let K be the length of the blocks to be chosen later. Let $M = \lfloor N/K \rfloor$. For $k = \ell M, 1 \leq \ell \leq K$ we write

$$\sum_{i=1}^{k}\xi_{i,j} = \sum_{v=1}^{\ell}\left(\sum_{i=(v-1)M+1}^{vM}\xi_{i,j}\right).$$

Using the $\gamma_{i,j}$'s, the independent standard normal random variables constructed in Lemma A.2, we define

$$(A.9) \qquad W_j(k) = \sum_{i=1}^{k}\gamma_{i,j}, \quad 1 \leq j \leq d, \ 1 \leq k \leq N.$$

By Lemma A.3 we get for any $0 < \delta < 1/2$ and $1 \leq j \leq d$ via Kolmogorov's inequality (cf. Petrov (1995)), p. 54)

$$(A.10) \qquad P\left\{\max_{1\leq\ell\leq K}\left|\sum_{i=1}^{\ell M}\xi_{i,j} - W_j(\ell M)\right| \geq N^{1/2-\delta}\right\}$$

$$= P\left\{\max_{1\leq\ell\leq K}\left|\sum_{v=1}^{\ell}\left(\sum_{i=(v-1)M+1}^{vM}(\xi_{i,j} - \gamma_{i,j})\right)\right| \geq N^{1/2-\delta}\right\}$$

$$\leq \frac{1}{N^{1-2\delta}}\sum_{v=1}^{K}E\left(\sum_{i=(v-1)M+1}^{vM}(\xi_{i,j} - \gamma_{i,j})\right)^{2}$$

$$\leq \frac{c_3}{N^{1-2\delta}} K M^{23/24} \frac{1}{\lambda_j} \left(d^{1/4} \left(\sum_{\ell=1}^{d} 1/\lambda_\ell \right)^{3/8} \right)^{1/3}$$

$$\leq c_3 N^{2\delta-1/24} K^{1/24} \frac{1}{\lambda_j} \left(d^{1/4} \left(\sum_{\ell=1}^{d} 1/\lambda_\ell \right)^{3/8} \right)^{1/3}.$$

One can define independent Wiener processes (standard Brownian motions) $W_j(x), x \geq 0, 1 \leq j \leq d$ such that (A.9) holds. We obtained approximations for the partial sums of the $\xi_{i,j}$'s at the points $k = \ell M, 1 \leq \ell \leq K$. Next we show that neither the partial sums of the $\xi_{i,j}$'s nor the Wiener processes $W_j(x)$ can oscillate too much between ℓM and $(\ell + 1)M$.

Using again Rosenthal's inequality (cf. Petrov (1995), p. 59) we obtain for all $1 \leq j \leq d$ that

$$(A.11) \qquad E \left| \sum_{i=1}^{M} \xi_{i,j} \right|^3 \leq c_{10} \left\{ \sum_{i=1}^{M} E |\xi_{i,j}|^3 + \left(\sum_{i=1}^{M} E \xi_{i,j}^2 \right)^{3/2} \right\}$$

$$\leq c_{11} \{ M/\lambda_j^{3/2} + M^{3/2} \}$$

$$\leq c_{11} (1 + \lambda_1^{3/2})(M/\lambda_j)^{3/2}$$

on account of Lemma A.1. Combining the Marcinkiewicz–Zygmund inequality (cf. Petrov (1995), p. 82) with (A.11) we conclude

$$(A.12) \qquad E \left(\max_{1 \leq h \leq M} \left| \sum_{i=1}^{h} \xi_{i,j} \right| \right)^3 \leq c_{12} (M/\lambda_j)^{3/2}.$$

Applying (A.12) we get

$$(A.13) \qquad P \left\{ \max_{0 \leq \ell \leq K+1} \max_{1 \leq h \leq M} \left| \sum_{i=1}^{\ell M} \xi_{i,j} - \sum_{i=1}^{\ell M+h} \xi_{i,j} \right| \geq N^{1/2-\delta} \right\}$$

$$\leq (K+2) P \left\{ \max_{1 \leq h \leq M} \left| \sum_{i=1}^{h} \xi_{i,j} \right| > N^{1/2-\delta} \right\}$$

$$\leq \frac{c_{13}}{N^{3/2-3\delta}} K (M/\lambda_j)^{3/2}$$

$$\leq c_{13} N^{3\delta} K^{-1/2} \lambda_j^{-3/2}.$$

Lemma 1.2.1 of Csörgő and Révész (1981) yields

$$(A.14) \qquad P \left\{ \max_{0 \leq \ell \leq K} \sup_{|h| \leq M} |W_j(\ell M) - W_j(\ell M + h)| \geq c_{14} M^{1/2} (\log N)^{1/2} \right\} \leq \frac{c_{15}}{N^2}.$$

Now choosing $\delta = 1/80$ and $K = \lfloor N^\beta \rfloor$ with $\beta = 1/10$, it follows from (A.10), (A.13) and (A.14) for all $1 \leq j \leq d$ that

$$(A.15) \qquad P \left\{ \sup_{0 \leq y \leq N} \left| \sum_{1 \leq i \leq y} \xi_{i,j} - W_j(y) \right| > N^{1/2-\delta} \right\}$$

$$\leq c_{15}N^{-\delta}\left\{\frac{1}{\lambda_j}\left(d^{1/4}\left(\sum_{\ell=1}^{d}1/\lambda_\ell\right)^{3/8}\right)^{1/3}+\frac{1}{\lambda_j^{3/2}}\right\}.$$

The result now follows from (A.15) with $W_{j,N}(x) = N^{-1/2}W_j(Nx), 0 \leq x \leq 1$.   □

APPENDIX B. **Proofs of the results of Section 3**

We first investigate the weak convergence of the process

$$Z_N(u,x) = \frac{1}{d^{1/2}}\sum_{j=1}^{\lfloor du \rfloor}\left\{(S_{j,N}(x) - xS_{j,N}(1))^2 - x(1-x)\right\}, \ \ 0 \leq u, x \leq 1,$$

with $S_{j,N}(x)$ given by (2.4). The difference between $\hat{Z}_N(u,x)$ and $Z_N(u,x)$ is that $\hat{Z}_N$ is computed from the empirical projections $\hat{v}_1, \ldots, \hat{v}_d$, while Z_N is based on the unknown population eigenfunctions $v_1, \ldots, v_d$.

Theorem B.1. *If Assumptions 2.1, 2.3, 2.4 and 3.1–3.4 hold, then*

$$Z_N(u,x) \ \to \ \Gamma(u,x) \ \text{in} \ \mathcal{D}[0,1]^2,$$

where the Gaussian process $\Gamma(u,x)$ is defined in Theorem 3.1.

To prove Theorem B.1, we need several lemmas and some additional notation. Let

$$V_{j,N}(x) = S_{j,N}(x) - xS_{j,N}(1) \ \ \text{and} \ \ B_{j,N}(x) = W_{j,N}(x) - xW_{j,N}(1),$$

where $S_{j,N}$ is defined in (2.4) and the $W_{j,N}$'s are the Wiener processes of Theorem 2.1. It follows from the definition that for each N the processes $B_{j,N}, 1 \leq j \leq d$, are independent Brownian bridges.

Lemma B.1. *If Assumptions 2.1, 2.3 and 2.4 hold, then*

$$P\left\{\sup_{0\leq x\leq 1}\sum_{j=1}^{d}\left|V_{j,N}^2(x) - B_{j,N}^2(x)\right| \geq 20dN^{-1/80}(\log N)^{1/2}\right\}$$

$$\leq c_*N^{-1/80}\left\{d^{1/12}\left(\sum_{\ell=1}^{d}1/\lambda_\ell\right)^{1/8} + \sum_{j=1}^{d}1/\lambda_j^{3/2}\right\}+c_{**}dN^{-2},$$

where c_* and c_{**} only depend on λ_1 and $E\|Z_1\|^3$.

Proof. First we write

$$V_{j,N}^2(x) - B_{j,N}^2(x) = (V_{j,N}(x) - B_{j,N}(x))^2 + 2B_{j,N}(x)(V_{j,N}(x) - B_{j,N}(x)).$$

Since the $B_{j,N}$'s are Brownian bridges, the distribution of the supremum functional of the Brownian bridge (cf. Csörgő and Révész (1981)) gives

$$P\left\{\max_{1\leq j\leq d}\sup_{0\leq x\leq 1}|B_{j,N}(x)| \geq 4(\log N)^{1/2}\right\}\leq c_{**}\frac{d}{N^2},$$

where c_{**} is an absolute constant. Now the result follows immediately from Theorem 2.1.   □

Now we prove the weak convergence of the partial sums of the squares of independent Brownian bridges. Let $B_1, B_2, \ldots, B_d$ be independent Brownian bridges.

Lemma B.2. *As $d \to \infty$, we have that*

$$\frac{1}{d^{1/2}} \sum_{j=1}^{\lfloor du \rfloor} \left(B_j^2(x) - x(1-x) \right) \to \Gamma(u, x) \quad in \ \mathcal{D}[0,1]^2,$$

where the Gaussian process $\Gamma(u, x)$ is defined in Theorem 3.1.

Proof. The proof is based on Theorem 2 of Hahn (1978). Let B denote a Brownian bridge and $\theta_1 = \sup_{0 \le t \le 1} |B(t)|$. It is clear that $E\theta_1^m < \infty$ for all $m \ge 1$. According to Garsia (1970), there is a random variable θ_2 such that $E\theta_2^m < \infty$ for all $m \ge 1$ and

$$|B(t) - B(s)| \le \theta_2 (|t-s| \log(1/|t-s|))^{1/2}, \quad 0 \le t, s \le 1.$$

Let $V(t) = B^2(t) - t(1-t)$. We note

$$|V(t) - V(s)| \le 2\theta_1 \theta_2 (|t-s| \log(1/|t-s|))^{1/2} + |t-s|.$$

Thus we get

(B.1) $\qquad E(V(t) - V(s))^2 \le c_{16} |t-s| \log(1/|t-s|) \quad \text{for all } 0 \le t, s \le 1$

and

(B.2) $\qquad E[(V(t) - V(z))^2 (V(z) - V(s))^2] \le c_{17} (|t-s| \log(1/|t-s|))^2$

for all $0 \le s \le z \le t \le 1$. The estimates in (B.1) and (B.2) yield that the conditions of Theorem 2 of Hahn (1978) are satisfied, completing the proof Lemma B.2. $\qquad \square$

Proof of Theorem B.1. It follows immediately from Lemmas B.1 and B.2. $\qquad \square$

The transition from Theorem B.1 to Theorem 3.1 is based on the following lemma, in which the norm is the Hilbert–Schmidt norm.

Lemma B.3. *If Assumptions 2.1, 2.2 and 2.3 hold, then*

(B.3) $\qquad\qquad\qquad\qquad |\lambda_j - \hat{\lambda}_j| \le \|\mathfrak{c} - \hat{\mathfrak{c}}\|$

and

(B.4) $\qquad\qquad\qquad\qquad \|v_j - \hat{c}_j \hat{v}_j\| \le \frac{2\sqrt{2}}{\zeta_j} \|\mathfrak{c} - \hat{\mathfrak{c}}\|,$

where $\hat{c}_j = \text{sign}(\langle \hat{v}_j, v_j \rangle)$ are random signs, and $\zeta_1, \zeta_2, \ldots$ are defined in Assumption 3.5.

Proof. Inequality (B.3) can be deduced from the general results presented in Section VI.1 of Gohberg et al. (1990) or in Dunford and Schwartz (1988). These results are presented in a convenient form in Lemma 2.2 in Horváth and Kokoszka (2012). Finally Lemma 2.3 in Horváth and Kokoszka (2012) gives (B.4). $\qquad \square$

Proof of Theorem 3.1. Introducing

$$U_N(x) = U_N(x,t) = \frac{1}{N^{1/2}}\left\{ \sum_{i=1}^{\lfloor Nx \rfloor} Z_i(t) - x \sum_{i=1}^{N} Z_i(t) \right\}$$

we can write

$$\hat{Z}_N(u,x) = \frac{1}{d^{1/2}} \sum_{j=1}^{\lfloor du \rfloor} \left\{ \frac{1}{\hat{\lambda}_j} \langle U_N(x), \hat{v}_j \rangle^2 - x(1-x) \right\}.$$

Elementary arguments give

$$\sum_{j=1}^{\lfloor du \rfloor} \frac{1}{\hat{\lambda}_j} \langle U_N(x), \hat{v}_j \rangle^2 = \sum_{j=1}^{\lfloor du \rfloor} \frac{1}{\lambda_j} \langle U_N(x), \hat{c}_j v_j \rangle^2 + \sum_{j=1}^{\lfloor du \rfloor} \left\{ \frac{1}{\hat{\lambda}_j} - \frac{1}{\lambda_j} \right\} \langle U_N(x), \hat{v}_j \rangle^2$$

$$+ \sum_{j=1}^{\lfloor du \rfloor} \frac{1}{\lambda_j} (\langle U_N(x), \hat{v}_j \rangle^2 - \langle U_N(x), \hat{c}_j v_j \rangle^2).$$

By the Cauchy–Schwarz inequality we have

(B.5) $$\frac{1}{d^{1/2}} \sum_{j=1}^{d} \left| \frac{1}{\hat{\lambda}_j} - \frac{1}{\lambda_j} \right| \langle U_N(x), \hat{v}_j \rangle^2 \le \|U_N(x)\|^2 \frac{1}{d^{1/2}} \sum_{j=1}^{d} \frac{|\lambda_j - \hat{\lambda}_j|}{\hat{\lambda}_j \lambda_j}$$

and since $|a^2 - b^2| = |a+b||a-b|$,

(B.6) $$\frac{1}{d^{1/2}} \sum_{j=1}^{d} \frac{1}{\lambda_j} (\langle U_N(x), \hat{v}_j \rangle^2 - \langle U_N(x) - \hat{c}_j v_j \rangle^2) \le \|U_N(x)\|^2 \frac{2}{d^{1/2}} \sum_{j=1}^{d} \frac{1}{\lambda_j} \|\hat{v}_j - \hat{c}_j v_j\|^2.$$

It follows from the results of Kuelbs (1973) (for a shorter proof we refer to Theorem 6.3 in Horváth and Kokoszka (2012)) that

$$\sup_{0 \le x \le 1} \|U_N(x)\|^2 = O_P(1).$$

Due to Assumption 2.4 we can use a Marcinkiewicz–Zygmund type law of large numbers for sums of independent and identically distributed random functions in Banach spaces (cf., e.g., Woyczynski (1978) or Howell and Taylor (1980)) to conclude

$$\|\mathfrak{c} - \hat{\mathfrak{c}}\| = O_P(N^{-1/3}).$$

Assumption 3.4 gives that $N^{-1/120}/\lambda_d \to 0$ and therefore by Lemma B.3

$$\max_{1 \le i \le d} \frac{\lambda_i}{\hat{\lambda}_i} = O_P(1).$$

So by Lemma B.3 and (B.5) we have

$$\frac{1}{d^{1/2}} \sum_{j=1}^{d} \left| \frac{1}{\hat{\lambda}_j} - \frac{1}{\lambda_j} \right| \langle U_N(x), \hat{v}_j \rangle^2 = O_P(1) \frac{1}{d^{1/2} N^{1/3}} \sum_{i=1}^{d} 1/\lambda_i^2$$

$$= O_P(1) \frac{d^{1/2}}{N^{1/3}} \frac{1}{\lambda_d^2}$$

$$= O_P(1)\frac{N^{1/80}}{N^{1/3}}N^{1/60}$$
$$= o_P(1)$$

on account of Assumptions 3.2 and 3.4. Similarly, (B.6) and Assumption 3.5 yield

$$\text{(B.7)}\qquad \frac{1}{d^{1/2}}\sum_{j=1}^{d}\frac{1}{\lambda_j}\langle U_N(x),\hat{v}_j-\hat{c}_j v_j\rangle^2 = O_P(1)\frac{1}{d^{1/2}N^{1/3}}\sum_{j=1}^{d}\frac{1}{\lambda_j\zeta_j} = o_P(1).$$

Theorem 3.1 now follows from Theorem B.1. $\qquad\square$

Proof of Corollary 3.1. By Lemma B.1 and (B.7), relation (3.2) is proven if we show that

$$\text{(B.8)}\qquad \frac{1}{d^{1/2}\sigma_0}\left\{\sum_{i=1}^{d}\sup_{0\leq x\leq 1}B_i^2(x)-d\kappa_0\right\} \xrightarrow{\mathcal{D}} N(0,1),$$

where $B_1, B_2, \ldots, B_d$ are independent Brownian bridges. Clearly, (B.8) is an immediate consequence of the central limit theorem. Similarly, to establish (3.3), we need to show only that

$$\frac{1}{(d/45)^{1/2}}\left\{\sum_{i=1}^{d}\int B_i^2(x)dx-\frac{d}{6}\right\} \xrightarrow{\mathcal{D}} N(0,1).$$

The above result is known, see Remark 2.1 in Aue et al. (2009). The same argument can be used to prove (3.4). $\qquad\square$

Appendix C. **Proofs of the results of Section 4**

We note that under the null hypothesis $\bar{X}_N-\bar{Y}_M = \bar{Z}_N-\bar{Q}_M$. Define

$$F_{N,M} = \sum_{j=1}^{N}Z_j - \frac{N}{M}\sum_{j=1}^{M}Q_j.$$

The proof of Theorem 4.1 is based on Lemma A.2, we need to write $F_{N,M}$ as a single sum of independent identically distributed random processes and an additional small remainder term. Let K be an integer and define the integers $R = \lfloor N/K \rfloor$ and $L = \lfloor M/K \rfloor$. Next we define

$$A_i = \sum_{\ell=R(i-1)+1}^{iR}Z_\ell - \sum_{\ell=L(i-1)+1}^{iL}\frac{N}{M}Q_\ell, \quad i=1,2,\ldots,K.$$

Clearly,

$$F_{N,M} = \sum_{i=1}^{K}A_i + \tilde{A},$$

where

$$\tilde{A} = \sum_{\ell=KR+1}^{N}Z_\ell - \frac{N}{M}\sum_{\ell=KL+1}^{M}Q_\ell.$$

We will show first if v is a function with $\|v\| = 1$, then for every n

$$(C.1) \qquad E\left|\sum_{\ell=1}^{n}\langle Z_\ell, v\rangle\right|^3 \leq c_1 n^{3/2}$$

and

$$(C.2) \qquad E\left|\sum_{\ell=1}^{n}\langle Q_\ell, v\rangle\right|^3 \leq c_2 n^{3/2},$$

where c_1 and c_2 only depends on $E\|Z_1\|^3$ and $E\|Q_1\|^3$, respectively. Using Rosenthal's inequality (cf. Petrov (1995), p. 59) we get

$$E\left|\sum_{\ell=1}^{n}\langle Z_\ell, v\rangle\right|^3 \leq c_3\left\{nE|\langle Z_1, v\rangle|^3 + (nE\langle Z_1, v\rangle^2)^{3/2}\right\},$$

where c_3 is an absolute constant. It is easy to see that

$$|\langle Z_1, v\rangle| \leq \|Z_1\|,$$

which implies (C.1). The same argument can be used to prove (C.2).

Next we define the function

$$\mathfrak{c}_{N,M}(t,s) = \mathfrak{c}(t,s) + \frac{N^2 L}{M^2 R}\mathfrak{c}_*(t,s).$$

It is clear that $\mathfrak{c}_{N,M}$ is a covariance function and therefore we can find $\bar{\kappa}_1 = \bar{\kappa}_1(N,M) \geq \bar{\kappa}_2 = \bar{\kappa}_2(N,M) \geq \ldots$ and orthonormal functions $\bar{u}_1(t) = \bar{u}_1(N,M), \bar{u}_2(t) = \bar{u}_2(N,M), \ldots$ satisfying

$$\bar{\kappa}_i \bar{u}_i(t) = \int \mathfrak{c}_{N,M}(t,s)\bar{u}_i(s)ds, \quad 1 \leq i < \infty.$$

Now we define the vector

$$\boldsymbol{\psi}_i = (\langle A_i, \bar{u}_1\rangle/(R\bar{\kappa}_1)^{1/2}, \langle A_i, \bar{u}_2\rangle/(R\bar{\kappa}_2)^{1/2}, \ldots, \langle A_i, \bar{u}_d\rangle/(R\bar{\kappa}_d)^{1/2})^T, \quad 1 \leq i \leq K.$$

It is easy to see that $\boldsymbol{\psi}_i$, $1 \leq i \leq K$, are independent and identically distributed random vectors with mean $\mathbf{0}$ and $E\boldsymbol{\psi}_1\boldsymbol{\psi}_1^T = \mathbf{I}_d$, where $\mathbf{I}_d$ is the $d \times d$ identity matrix. Also, (C.1) and (C.2) imply that

$$E|\boldsymbol{\psi}_1| \leq c_4\left(\sum_{\ell=1}^{d} 1/\bar{\kappa}_\ell\right)^{3/2},$$

where c_4 only depends on $E\|Z_1\|^3$ and $E\|Q_1\|^3$. Using Lemma A.2 we obtain similarly to (A.8) that there are independent standard normal random vectors $\boldsymbol{\gamma}_i = \boldsymbol{\gamma}_i(N,M), 1 \leq i \leq K$, in R^d such that

$$(C.3) \qquad P\left\{\left|\sum_{i=1}^{K}\boldsymbol{\psi}_i - \sum_{i=1}^{K}\boldsymbol{\gamma}_i\right| \geq c_5 K^{3/8} d^{1/4}\left(\sum_{\ell=1}^{d} 1/\bar{\kappa}_\ell\right)^{3/8}\right\}$$

$$\leq c_5 K^{-1/8} d^{1/4}\left(\sum_{\ell=1}^{d} 1/\bar{\kappa}_\ell\right)^{3/8},$$

where c_5 does not depend on d. Let

$$\tilde{\boldsymbol{\psi}} = (\langle \tilde{A}, \bar{u}_1\rangle/\sqrt{\bar{\kappa}_1}, \langle \tilde{A}, \bar{u}_2\rangle/\sqrt{\bar{\kappa}_2}, \ldots, \langle \tilde{A}, \bar{u}_d\rangle/\sqrt{\bar{\kappa}_d})^T.$$

It follows from (C.1) and (C.2) that with some constant c_6, not depending on d we have

$$E|\tilde{\boldsymbol{\psi}}|^3 \le c_6 K^{3/2} \left(\sum_{\ell=1}^d 1/\bar{\kappa}_\ell \right)^{3/2}$$

and therefore by Markov's inequality for every $x > 0$

$$(C.4) \qquad P\left\{ N^{-1/2}|\tilde{\boldsymbol{\psi}}| > x \right\} \le c_7 \frac{K^{3/2}}{x^3 N^{3/2}} \left(\sum_{\ell=1}^d 1/\bar{\kappa}_\ell \right)^{3/2}.$$

Let

$$\boldsymbol{\kappa}_{N,M} = (\langle F_{N,M}, \bar{u}_1\rangle/\sqrt{\bar{\kappa}_1}, \langle F_{N,M}, \bar{u}_2\rangle/\sqrt{\bar{\kappa}_2}, \ldots, \langle F_{N,M}, \bar{u}_d\rangle/\sqrt{\bar{\kappa}_d})^T.$$

Next we choose $K = \lfloor N^{3/4} \rfloor$ in (C.3), (C.4) and $x = K^{-1/8}(\sum_{\ell=1}^d 1/\bar{\kappa}_\ell)^{3/8}$ in (C.4) to conclude that there is $\boldsymbol{\gamma}_{N,M}$, a standard normal random vector in R^d such that

$$(C.5) \qquad P\left\{ \left| \frac{1}{\sqrt{N^*}} \boldsymbol{\kappa}_{N,M} - \boldsymbol{\gamma}_{N,M} \right| \ge c_8 N^{-3/32} d^{1/4} \left(\sum_{\ell=1}^d 1/\bar{\kappa}_\ell \right)^{3/8} \right\}$$

$$\le c_8 N^{-3/32} d^{1/4} \left(\sum_{\ell=1}^d 1/\bar{\kappa}_\ell \right)^{3/8},$$

where $N^* = \lfloor N/\lfloor N^{3/4}\rfloor \rfloor \lfloor N^{3/4} \rfloor$. Using the definitions of $\mathfrak{c}_P$ and $\mathfrak{c}_{N,M}$, together with Assumption 4.3, we conclude

$$(C.6) \qquad \|\mathfrak{c}_P - \mathfrak{c}_{N,M}\| = O(N^{-1/4}),$$

so by Lemma 2.3 of Horváth and Kokoszka (2012), cf. Lemma B.3, we have

$$(C.7) \qquad |\kappa_i - \bar{\kappa}_i| \le c_9 \|\mathfrak{c}_P - \mathfrak{c}_{N,M}\| = O(N^{-1/4}).$$

Using Assumption 4.5 we conclude that

$$\sum_{\ell=1}^d 1/\bar{\kappa}_\ell = O\left(\sum_{\ell=1}^d 1/\kappa_\ell \right).$$

Hence it follows from (C.5) and Assumption 4.5 that

$$\frac{1}{N}|\boldsymbol{\kappa}_{N,M}|^2 - \frac{N^*}{N}|\boldsymbol{\gamma}_{N,M}|^2 = o_P(d^{1/2}).$$

Since $|\boldsymbol{\gamma}_{N,M}|^2$ is a χ^2 random variable with d degrees of freedom, Assumption 4.5 yields that

$$\left| \frac{N^*}{N} - 1 \right| |\boldsymbol{\gamma}_{N,M}|^2 = o_P(d^{1/2}).$$

It is well known that $(|\boldsymbol{\gamma}_{N,M}|^2 - d)/(2d)^{1/2}$ converges in distribution to a standard normal random variable, and therefore

$$\frac{1}{\sqrt{2d}}\left\{\frac{1}{N}|\boldsymbol{\kappa}_{N,M}|^2 - d\right\} \xrightarrow{\mathcal{D}} N(0,1),$$

where $N(0,1)$ stands for a standard normal random variable.

The difference between $|\boldsymbol{\kappa}_{N,M}|^2/N$ and $\widehat{D}_{N,M}$ is that the projections are done into the direction of different functions ($\bar{u}_i$'s and $\hat{u}_i$'s, respectively) and the normalizations ($\bar{\kappa}_i$'s and $\hat{\kappa}_i$'s, respectively) are also different. However, using the Marcinkiewicz–Zygmund law of large numbers in a Banach space together with (C.6) and Assumption 4.5, we obtain that

$$\|\hat{\mathfrak{c}}_P - \mathfrak{c}_{N,M}\| = O_P(N^{-1/4}).$$

Hence, in view of (C.7), also

$$\sup_i |\hat{\kappa}_i - \bar{\kappa}_i| = O_P(N^{-1/4}),$$

and there are random signs $\hat{d}_i$ such that

$$\sup_i \left(\sum_{\ell=1}^i 1/\iota_\ell\right)^{-1} \|\hat{u}_i - \hat{d}_i\bar{u}_i\| = O_P(N^{-1/4}).$$

So repeating the arguments used in the proof of Theorem 3.1, we get

$$\left|\widehat{D}_{N,M} - \frac{1}{N}|\boldsymbol{\kappa}_{N,M}|^2\right| = o_P(d^{1/2}),$$

completing the proof.

References

J. A. D. Aston and C. Kirch. Estimation of the distribution of change–points with application to fMRI data. *The Annals of Applied Statistics*, 2012. Forthcoming.

A. Aue, S. Hörmann, L. Horváth, and M. Reimherr. Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, 37:4046–4087, 2009.

M. Basseville, I. V. Nikifirov, and A. Tartakovsky. *Sequential Analysis: Hypothesis Testing and Change–Point Detection*. Chapman & Hall/CRC, 2012.

N. Bathia, Q. Yao, and F. Ziegelmann. Identifying the finite dimensionality of curve time series. *The Annals of Statistics*, 38:3353–3386, 2010.

M. Benko, W. Härdle, and A. Kneip. Common functional principal components. *The Annals of Statistics*, 37:1–34, 2009.

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

H. Cardot, F. Ferraty, A. Mas, and P. Sarda. Testing hypothesis in the functional linear model. *Scandinavian Journal of Statistics*, 30:241–255, 2003.

J. Chen and A. K. Gupta. *Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance*. Birkhäuser, 2011.

M. Csörgő and L. Horváth. *Limit Theorems in Change-Point Analysis*. Wiley, New York, 1997.

M. Csörgő and P. Révész. *Strong Approximations in Probability and Statistics*. Academic Press, New York, 1981.

N. Dunford and J. T. Schwartz. *Linear Operators, Parts I and II*. Wiley, 1988.

U. Einmahl. Strong invariance principles for partial sums of independent random vectors. *The Annals of Probability*, 15:1419–1440, 1987.

U. Einmahl. Extension of results of Komlós, Major and Tusnady to the multivariate case. *Journal of Multivariate Analysis*, 28:20–68, 1989.

C. Gallagher, R. Lund, and M. Robbins. Changepoint detection in daily precipitation data. *Environmetrics*, 23: 407–419, 2012.

A. M. Garsia. Continuity properties of Gaussian processes with multidimensional time parameter. In *Proceedings of the 6th Berkeley Symp. Math. Stat. Probab.*, volume 2, pages 369–374. University of California Press, 1970.

I. Gohberg, S. Golberg, and M. A. Kaashoek. *Classes of Linear Operators*, volume 49 of *Operator Theory: Advances and Applications*. Birkhaüser, 1990.

O. Gromenko, P. Kokoszka, L. Zhu, and J. Sojka. Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. *The Annals of Applied Statistics*, 6:669–696, 2012.

M. G. Hahn. Central limit theorems in $D[0,1]$. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 44:89–101, 1978.

P. Hall and C. Vial. Assessing the finite dimensionality of functional data. *Journal of the Royal Statistical Society (B)*, 68:689–705, 2006.

L. Horváth and P. Kokoszka. *Inference for Functional Data with Applications.* Springer, 2012.

L. Horváth, P. Kokoszka, and M. Reimherr. Two sample inference in functional linear models. *Canadian Journal of Statistics*, 37:571–591, 2009.

L. Horváth, P. Kokoszka, and R. Reeder. Estimation of the mean of functional time series and a two sample problem. *Journal of the Royal Statistical Society (B)*, 75 : 103–122, 2013.

J. O. Howell and R. L. Taylor. Marcinkiewicz–Zygmund weak laws of large numbers for unconditional random elements in banach spaces. In J. Kuelbs, editor, *Probability in Banach Spaces. III. Proceedings of the Third International Conference held at Tufts University, Medford, Mass.*, pages 219–230. Springer, 1980.

J. Kuelbs. The invariance principle for Banach space valued random variables. *Journal of Multivariate Analysis*, 3:161–172, 1973.

V. M. Panaretos, D. Kraus, and J. H. Maddocks. Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *Journal of the American Statistical Association*, 105:670–682, 2010.

V. V. Petrov. *Limit Theorems of Probability Theory: Sequences of Independent Random Varianles.* Clarendon Press, 1995.

J. Ramsay, G. Hooker, and S. Graves. *Functional Data Analysis with R and MATLAB.* Springer, 2009.

J. O. Ramsay and B. W. Silverman. *Functional Data Analysis.* Springer, 2005.

V. V. Senatov. *Normal Approximation: New Results, Methods and Problems.* VSP, 1998.

V. Strassen. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 36:423–439, 1965.

W. Woyczynski. Geometry and martingales in Banach spaces. II. independent increments. In J. Kuelbs, editor, *Probability on Banach Spaces*, pages 267–517. Marcel Dekker, 1978.

STEFAN FREMDT, MATHEMATICAL INSTITUTE, UNIVERSITY OF COLOGNE, WEYERTAL 86–90, D–50931 KÖLN, GERMANY

LAJOS HORVÁTH, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF UTAH, SALT LAKE CITY, UT 84112–0090, USA

PIOTR KOKOSZKA, DEPARTMENT OF STATISTICS, COLORADO STATE UNIVERSITY, FT. COLLINS, CO 80523–1877, USA

JOSEF G. STEINEBACH, MATHEMATICAL INSTITUTE, UNIVERSITY OF COLOGNE, WEYERTAL 86–90, D–50931 KÖLN, GERMANY