Journal of the Royal Statistical Society
SERIES B Statistical Methodology

J. R. Statist. Soc. B (2009)
71, *Part* 5, *pp.* 927–946

Detecting changes in the mean of functional observations

István Berkes,

Graz University of Technology, Austria

Robertas Gabrys,

Utah State University, Logan, USA

Lajos Horváth

University of Utah, Salt Lake City, USA

and Piotr Kokoszka

Utah State University, Logan, USA

[Received December 2007. Final revision March 2009]

Summary. Principal component analysis has become a fundamental tool of functional data analysis. It represents the functional data as $X_i(t) = \mu(t) + \Sigma_{1 \leqslant l < \infty} \eta_{i,l} + v_l(t)$, where μ is the common mean, v_l are the eigenfunctions of the covariance operator and the $\eta_{i,l}$ are the scores. Inferential procedures assume that the mean function $\mu(t)$ is the same for all values of i. If, in fact, the observations do not come from one population, but rather their mean changes at some point(s), the results of principal component analysis are confounded by the change(s). It is therefore important to develop a methodology to test the assumption of a common functional mean. We develop such a test using quantities which can be readily computed in the R package fda. The null distribution of the test statistic is asymptotically pivotal with a well-known asymptotic distribution. The asymptotic test has excellent finite sample performance. Its application is illustrated on temperature data from England.

Keywords: Change-point detection; Functional data analysis; Mean of functional data; Significance test

1. Introduction

Functional data analysis (FDA) has been enjoying increased popularity over the last decade due to its applicability to problems which are difficult to cast into a framework of scalar or vector observations. Even if such standard approaches are available, the functional approach often leads to a more natural and parsimonious description of the data, and to more accurate inference and prediction results. Ramsay and Silverman (2005) has become a standard reference to the ideas and tools of FDA. To name a few recent applications of FDA which illustrate its advantages alluded to above, we cite Antoniadis and Sapatinas (2003), Fernández de Castro *et al.* (2005), Müller and Stadtmüller (2005), Yao *et al.* (2005) and Glendinning and Fleet (2007).

Address for correspondence: Piotr Kokoszka, Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan, UT 84322-3900, USA.
E-mail: Piotr.Kokoszka@usu.edu

 1369–7412/09/71927

A main tool of FDA is principal component analysis (PCA). It represents the functional observations $X_i(t)$, $t \in \mathcal{T}$, $i = 1, 2, \ldots, N$, in the form $X_i(t) = \mu(t) + \Sigma_{1 \leqslant l < \infty} \eta_{i,l} v_l(t)$, where μ is the mean, v_l are the eigenfunctions of the covariance operator and the $\eta_{i,l}$ are the scores. The set $\mathcal{T}$ can be interpreted as a time or a spatial domain, the methodology we develop requiring merely that it be a compact subset of a Euclidean space. To perform the functional PCA, the functional mean $\mu(t)$, approximated by the sample mean of the $X_i(t)$, is first subtracted from the data. The first principal component $v_1(t)$ is then interpreted as the main pattern of deviation of the observations from the mean $\mu(t)$ or, equivalently, as the direction in a function space of the largest variability away from the mean function. The subsequent eigenfunctions define analogous directions orthogonal to the previous eigenfunctions. This interpretation and the inferential procedures that are based on it assume that the mean function $\mu(t)$ is the same for all values of i. If, in fact, the mean changes at some index i, the results of PCA are confounded by the change. Issues of this type are most likely to emerge if the data are collected sequentially over time. Applications that we have in mind abound in climatology, environmental science and economics; detecting and locating changes in mean can be interpreted, for example, as climate shifts, a baseline change in a pollution level or a shift in a long-term rate of growth.

It is thus useful to develop a methodology for the detection of changes in the mean of functional observations that is both easy to apply and justified by a clear large sample argument. We propose a significance test for testing the null hypothesis of a constant functional mean against the alternative of a changing mean. We also show how to locate the change-points if the null hypothesis is rejected. Our methodology is readily implemented by using the R package `fda` (R Development Core Team, 2008; Ramsay *et al.*, 2007). The null distribution of the test statistic is asymptotically pivotal with a well-known asymptotic distribution going back to the work of Kiefer (1959).

The problem of detecting a change in the mean of a sequence of Banach-space-valued random elements has recently been approached from a theoretical angle by Rackauskas and Suquet (2006). Motivated by detecting an epidemic change (the mean changes and then returns to its original value), Rackauskas and Suquet (2006) proposed an interesting statistic based on increasingly fine dyadic partitions of the index interval and derived its limit, which is non-standard.

The change-point problem has been extensively studied in the multivariate setting starting with Srivastava and Worsley (1986), whereas the work of Horváth *et al.* (1999) is most closely related to the present paper. Different multivariate settings with further references are discussed in Lavielle and Teyssiére (2006), Zamba and Hawkins (2006) and Qu and Perron (2007), among others.

Returning to the functional setting, a somewhat related problem has recently been studied by Benko *et al.* (2009), who considered two populations, admitting the PCAs:

$$X_{i,p}(t) = \mu_p(t) + \sum_{1 \leqslant l < \infty} \eta_{i,p,l} v_{p,l}(t), \qquad p = 1, 2.$$

Benko *et al.* (2009) developed a bootstrap test for checking whether the elements of the two decompositions (including the means) are the same. Earlier, Laukaitis and Račkauskas (2005) considered the model

$$X_{i,g}(t) = \mu_g(t) + \varepsilon_{i,g}(t), \qquad g = 1, 2, \ldots, G,$$

with innovations $\varepsilon_{i,g}$ and group means μ_g, and tested $H_0 : \mu_1(t) = \ldots = \mu_G(t)$. Other contributions in this direction include Cuevas *et al.* (2004), Delicado (2007) and Ferraty *et al.* (2007). In these settings, it is known which population or group each observation belongs to. In our setting, we do not have any partition of the data into several sets with possibly different means. The change can occur at any point, and we want to test whether it occurs or not.

The paper is organized as follows. In Section 2, we introduce the required notation and assumptions, and recall several results which will be used in the following sections. Section 3 describes the methodology proposed and contains theorems which provide its asymptotic justification. The finite sample performance is investigated in Section 4, which also contains an illustrative application to the detection of changes in mean patterns of annual temperatures. The proofs of the theorems of Section 3 are presented in Appendix A.

2. Notation and assumptions

We consider functional observations $X_i(t)$, $t \in \mathcal{T}$, $i = 1, 2, \ldots, N$, defined over a compact set $\mathcal{T}$. We assume that the X_i are independent, and we want to test whether their mean remains constant in i. Thus we test the null hypothesis

$$H_0 : \mathbf{E}\{X_1(t)\} = \mathbf{E}\{X_2(t)\} = \ldots = \mathbf{E}\{X_N(t)\}, \qquad t \in \mathcal{T}.$$

Note that, under hypothesis H_0, we do not specify the value of the common mean.

Under the alternative, hypothesis H_0 does not hold. The test that we construct has particularly good power against the alternative in which the data can be divided into several consecutive segments, and the mean is constant within each segment, but changes from segment to segment. The simplest case of only two segments (one change-point) is specified in assumption 4.

Under the null hypothesis, we can represent each functional observation as

$$\begin{aligned} X_i(t) &= \mu(t) + Y_i(t), \\ \mathbf{E}\{Y_i(t)\} &= 0. \end{aligned} \tag{2.1}$$

The following assumption specifies conditions on $\mu(\cdot)$ and the errors $Y_i(\cdot)$ that are needed to establish the asymptotic distribution of the test statistic.

In what follows, unless indicated otherwise, all integrals denote integration over the set $\mathcal{T}$.

Assumption 1. The mean $\mu(\cdot)$ is in $L^2(\mathcal{T})$. The errors $Y_i(\cdot)$ are independent and identically distributed mean 0 random elements of $L^2(\mathcal{T})$ which satisfy

$$\mathbf{E}\|Y_i\|^2 = \int \mathbf{E}\{Y_i^2(t)\}\,\mathrm{d}t < \infty. \tag{2.2}$$

Their covariance function

$$c(t, s) = \mathbf{E}\{Y_i(t)\, Y_i(s)\} \qquad t, s \in \mathcal{T} \tag{2.3}$$

is square integrable, i.e. is in $L^2(\mathcal{T} \times \mathcal{T})$.

Assumption 1 implies the following expansions (see for example chapter 4 of Indritz (1963)):

$$c(t, s) = \sum_{1 \leqslant k < \infty} \lambda_k\, v_k(t)\, v_k(s) \tag{2.4}$$

and

$$Y_i(t) = \sum_{1 \leqslant l < \infty} \lambda_l^{1/2} \xi_{i,l}\, v_l(t), \tag{2.5}$$

where λ_k and v_k are the eigenvalues and eigenfunctions respectively of the covariance operator, i.e. they are defined by

$$\int c(t, s)\, v_l(s)\,\mathrm{d}s = \lambda_l\, v_l(t), \qquad l = 1, 2, \ldots. \tag{2.6}$$

The sequences $\{\xi_{i,l}, l = 1, 2, \ldots\}$ are independent, and within each sequence the $\xi_{i,l}$ are uncorrelated with mean 0 and unit variance. The infinite sum in equation (2.5) converges in $L^2(\mathcal{T})$ with

probability 1. Recall also that v_l, $l=1,2,\ldots$, form an orthonormal basis in $L^2(\mathcal{T})$, and all λ_l are non-negative.

In practice, we work with estimated eigenelements that are defined by

$$\int \hat{c}(t,s)\,\hat{v}_l(s)\,\mathrm{d}s = \hat{\lambda}_l\,\hat{v}_l(t), \qquad l=1,2,\ldots, \tag{2.7}$$

where

$$\hat{c}(t,s)=\frac{1}{N}\sum_{1\leqslant i\leqslant N}\{X_i(t)-\bar{X}_N(t)\}\{X_i(s)-\bar{X}_N(s)\} \qquad \bar{X}_N(t)=\frac{1}{N}\sum_{1\leqslant i\leqslant N}X_i(t).$$

To control the distance between the estimated and the population eigenelements, we need the following assumptions.

Assumption 2. The eigenvalues λ_l satisfy, for some $d>0$,

$$\lambda_1>\lambda_2>\ldots>\lambda_d>\lambda_{d+1}.$$

Assumption 3. The Y_i in expression (2.1) satisfy

$$\mathbf{E}\|Y_i\|^4=\int \mathbf{E}\{Y_i^4(t)\}\,\mathrm{d}t<\infty.$$

The results of Dauxois *et al.* (1982) and Bosq (2000) then imply that, for each $k\leqslant d$,

$$\begin{aligned}&\limsup_{N\to\infty}[N\{\mathbf{E}(\|\hat{c}_k v_k-\hat{v}_k\|^2)\}]<\infty,\\ &\limsup_{N\to\infty}[N\{\mathbf{E}(|\lambda_k-\hat{\lambda}_k|^2)\}]<\infty,\end{aligned} \tag{2.8}$$

where $\hat{c}_k=\mathrm{sgn}\{\int_{\mathcal{T}} v_k(t)\,\hat{v}_k(t)\,\mathrm{d}t\}$. The random sign $\hat{c}_k$ is included because the v_k and $\hat{v}_k$ are defined up to a sign and, since v_k is unknown, it is impossible to ensure that $\int_{\mathcal{T}} v_k(t)\,\hat{v}_k(t)\,\mathrm{d}t\geqslant 0$.

We establish the consistency of the test under the alternative of one change-point formalized in assumption 4. A similar argument can be developed if there are several change-points, but the technical complications then obscure the main idea that is explained in Section 3 and Appendix A.2 (in particular the functions (2.10) and (3.7) would need to be modified). The more general case is studied empirically in Section 4.

Assumption 4. The observations follow the model

$$X_i(t)=\begin{cases}\mu_1(t)+Y_i(t), & 1\leqslant i\leqslant k^*,\\ \mu_2(t)+Y_i(t), & k^*<i\leqslant N,\end{cases} \tag{2.9}$$

in which the Y_i satisfy assumption 1, the mean functions μ_1 and μ_2 are in $L^2(\mathcal{T})$ and

$$k^*=[n\theta] \qquad \text{for some } 0<\theta<1.$$

We shall see in the proof of theorem 2 that under assumption 4 the sample covariances of the functional observations converge to the function

$$\tilde{c}(t,s)=c(t,s)+\theta(1-\theta)\{\mu_1(t)-\mu_2(t)\}\{\mu_1(s)-\mu_2(s)\}. \tag{2.10}$$

This is a symmetric, square integrable function, and it is easy to see that, for any $x, y\in L^2(\mathcal{T})$,

$$\int\int \tilde{c}(t,s)\,x(t)\,x(s)\,\mathrm{d}t\,\mathrm{d}s\geqslant 0,$$

so $\tilde{c}(t,s)$ is a covariance function. Consequently, it has orthonormal eigenfunctions w_k and non-negative eigenvalues γ_k satisfying

$$\int \tilde{c}(t,s)\, w_k(s)\, \mathrm{d}s = \gamma_k\, w_k(t). \tag{2.11}$$

The quantities $\tilde{c}(t,s)$, w_k and γ_k are used in Section 3 to describe the distribution of the test statistic under the alternative of a single change-point.

3. Detection procedure

To explain the idea of the test procedure, denote

$$\hat{\mu}_k(t) = \frac{1}{k} \sum_{1 \leqslant i \leqslant k} X_i(t),$$

$$\tilde{\mu}_k(t) = \frac{1}{N-k} \sum_{k < i \leqslant N} X_i(t).$$

If the mean is constant, the difference $\Delta_k(t) = \hat{\mu}_k(t) - \tilde{\mu}_k(t)$ is small for all $1 \leqslant k < N$ and all $t \in \mathcal{T}$. However, $\Delta_k(t)$ can become large owing to chance variability if k is close to 1 or to N. It is therefore usual to work with the sequence

$$P_k(t) = \sum_{1 \leqslant i \leqslant k} X_i(t) - \frac{k}{N} \sum_{1 \leqslant i \leqslant N} X_i(t) = \frac{k(N-k)}{N} \{\hat{\mu}_k(t) - \tilde{\mu}_k(t)\} \tag{3.1}$$

in which the variability at the end points is attenuated by a parabolic weight function. If the mean changes, the difference $P_k(t)$ is large for some values of k and of t. Since the observations are in an infinite dimensional domain, we work with the projections of the functions $P_k(\cdot)$ on the principal components of the data. These projections can be expressed in terms of functional scores which can be easily computed by using the R package `fda`.

Consider thus the scores corresponding to the largest d eigenvalues:

$$\hat{\eta}_{i,l} = \int \{X_i(t) - \bar{X}_N(t)\}\, \hat{v}_l(t)\, \mathrm{d}t, \qquad i = 1, 2, \ldots, N, \quad l = 1, 2, \ldots, d.$$

Observe that the value of $P_k(t)$ does not change if the $X_i(t)$ are replaced by $X_i(t) - \bar{X}_N(t)$. Consequently, setting $l = [Nx], x \in (0,1)$, we obtain

$$\int \left\{ \sum_{1 \leqslant i \leqslant Nx} X_i(t) - \frac{[Nx]}{N} \sum_{1 \leqslant i \leqslant N} X_i(t) \right\} \hat{v}_l(t)\, \mathrm{d}t = \sum_{1 \leqslant i \leqslant Nx} \hat{\eta}_{i,l} - \frac{[Nx]}{N} \sum_{1 \leqslant i \leqslant N} \hat{\eta}_{i,l}. \tag{3.2}$$

Identity (3.2) shows that functional scores can be used for testing the constancy of the mean function.

The following theorem can be used to derive various test statistics. To state it, introduce the statistic

$$T_N(x) = \frac{1}{N} \sum_{l=1}^{d} \hat{\lambda}_l^{-1} \left(\sum_{1 \leqslant i \leqslant Nx} \hat{\eta}_{i,l} - x \sum_{1 \leqslant i \leqslant N} \hat{\eta}_{i,l} \right)^2 \tag{3.3}$$

and let $B_1(\cdot), \ldots, B_d(\cdot)$ denote independent standard Brownian bridges.

Theorem 1. Suppose that assumptions 1–3 hold. Then, under hypothesis H_0,

$$T_N(x) \xrightarrow{\mathrm{d}} \sum_{1 \leqslant l \leqslant d} B_l^2(x) \qquad 0 \leqslant x \leqslant 1,$$

in the Skorohod topology of $D[0,1]$.

Theorem 1 is proved in Appendix A.

By theorem 1, $U(T_N) \to^{\mathrm{d}} U\{\Sigma_{1\leqslant l\leqslant d} B_l^2(\cdot)\}$, for any continuous functional $U : D[0,1] \to R$. Applying integral or max-functionals, or their weighted versions, leads to useful statistics. In this paper, we focus on the integral of the squared function, i.e. the Cramer–von Mises functional, which is known to produce effective tests (this functional was also selected in a different context by Bugni *et al.* (2006)). Thus, we consider the convergence

$$\int_0^1 T_N(x)\,\mathrm{d}x \xrightarrow{\mathrm{d}} \int_0^1 \sum_{1\leqslant l\leqslant d} B_l^2(x)\,\mathrm{d}x,$$

which can be rewritten as

$$S_{N,d} := \frac{1}{N^2}\sum_{l=1}^{d} \hat{\lambda}_l^{-1} \sum_{k=1}^{N} \left(\sum_{1\leqslant i\leqslant k} \hat{\eta}_{i,l} - \frac{k}{N} \sum_{1\leqslant i\leqslant N} \hat{\eta}_{i,l} \right)^2 \xrightarrow{\mathrm{d}} \int_0^1 \sum_{1\leqslant l\leqslant d} B_l^2(x)\,\mathrm{d}x. \tag{3.4}$$

The distribution of the random variable

$$K_d = \int_0^1 \sum_{1\leqslant l\leqslant d} B_l^2(x)\,\mathrm{d}x \tag{3.5}$$

was derived by Kiefer (1959). Denoting by $c_d(\alpha)$ its $(1-\alpha)$th quantile, the test rejects hypothesis H_0 if $S_{N,d} > c_d(\alpha)$. The critical values $c_d(\alpha)$ are presented in Table 1 in Section 4.

A multivariate analogue of statistic (3.4) that was considered in Horváth *et al.* (1999) is

$$M_{N,d} = \frac{1}{N^2}\sum_{k=1}^{N} \left(\frac{k}{N}\frac{N-k}{N} \right)^2 \boldsymbol{\Delta}(k)\hat{\mathbf{D}}_d{}^{-1}\boldsymbol{\Delta}^{\mathrm{T}}(k), \tag{3.6}$$

where $\boldsymbol{\Delta}(k)$ is the difference of the mean vectors (of dimension d) computed from the first k and the last $N-k$ data vectors, and $\hat{\mathbf{D}}_d$ is the $d\times d$ matrix of estimated residual vectors. If d is large, the inverse of $\hat{\mathbf{D}}_d$ is unstable. In statistic (3.4), this inverse is 'replaced' by inverses of the d largest eigenvalues $\hat{\lambda}_l$, and the whole statistic is properly 'diagonalized' so that only the most important variability of the data is considered, whereas the high dimensional noise is ignored.

We now turn to the behaviour of the test under the alternative hypothesis. We shall show that it is consistent, i.e. $S_{N,d} \to^{\mathrm{P}} \infty$. In fact, we can obtain the rate of divergence: under H_{A}, $S_{n,d}$ grows linearly with N. We formulate these results under the assumption of one change-point.

Under assumption 4, for $1 \leqslant k \leqslant d$, introduce the functions

$$g_k(x) = \begin{cases} x(1-\theta)\displaystyle\int \{\mu_1(t)-\mu_2(t)\}\,w_k(t)\,\mathrm{d}t, & 0 < x \leqslant \theta, \\ \theta(1-x)\displaystyle\int \{\mu_1(t)-\mu_2(t)\}\,w_k(t)\,\mathrm{d}t, & \theta < x < 1. \end{cases} \tag{3.7}$$

Theorem 2. Under assumption 4,

$$\sup_{0\leqslant x\leqslant 1} |N^{-1}T_N - \mathbf{g}^{\mathrm{T}}(x)\Sigma^{*}\mathbf{g}(x)| = o_P(1),$$

where

$$\mathbf{g}(x) = (g_1(x),\ldots,g_d(x))^{\mathrm{T}},$$

$$\Sigma^{*} = \begin{pmatrix} 1/\gamma_1 & 0 & \cdots & 0 \\ 0 & 1/\gamma_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1/\gamma_d \end{pmatrix}.$$

Theorem 2 is proved in Appendix A.

It follows that the test statistic (3.4) satisfies the law of large numbers under the alternative hypothesis, i.e.

$$\frac{1}{N} S_{N,d} \xrightarrow{\mathrm{P}} \sum_{1 \leqslant k \leqslant d} \frac{1}{\gamma_k} \int_0^1 g_k^2(x)\,\mathrm{d}x.$$

If $\int_0^1 g_k^2(x)\,\mathrm{d}x > 0$ for some $1 \leqslant k \leqslant d$, then $S_{N,d} \to^{\mathrm{P}} \infty$.

To understand when the test is consistent, introduce the jump function $\Delta(t) = \mu_1(t) - \mu_2(t)$. By expression (3.7), the condition $\int_0^1 g_k^2(x)\,\mathrm{d}x > 0$ is equivalent to $\int_0^1 \Delta(s)\, w_k(s)\,\mathrm{d}s \neq 0$. Thus the test will have no power if

$$\int_0^1 \Delta(s)\, w_k(s)\,\mathrm{d}s = 0, \qquad \text{for all } 1 \leqslant k \leqslant d. \tag{3.8}$$

By equations (2.10) and (2.11), condition (3.8) is equivalent to

$$\int c(t,s)\, w_k(s)\,\mathrm{d}s = \gamma_k\, w_k(t), \qquad \text{for all } 1 \leqslant k \leqslant d. \tag{3.9}$$

Comparing with condition (2.6), we see that condition (3.8) means that, up to a sign, the w_k and γ_k are equal to v_k and λ_k, for $1 \leqslant k \leqslant d$. This leads us to the following corollary.

Corollary 1. If assumption 4 holds, and the jump function $\Delta(t) = \mu_1(t) - \mu_2(t)$ is not orthogonal to the subspace that is spanned by the first d eigenfunctions of the covariance kernel $c(t,s)$ (2.3), then $S_{N,d} \to^{\mathrm{P}} \infty$, as $N \to \infty$.

To estimate the change-point, we plot the function $T_N(x)$ (3.3) against $0 \leqslant x \leqslant 1$ and estimate θ by the value of x which maximizes $T_N(x)$. The intuition behind this estimator is clear from equations (3.3) and (3.2). To ensure uniqueness, we formally define this estimator as

$$\hat{\theta}_N = \inf\{x : T_N(x) = \sup_{0 \leqslant y \leqslant 1} \{T_N(y)\}\}. \tag{3.10}$$

Its weak consistency is established in the following proposition.

Proposition 1. If the assumptions of corollary 1 hold, then $\hat{\theta}_N \to^{\mathrm{P}} \theta$.

Proof. The argument x maximizing $T_N(x)$ clearly maximizes $A_N(x) = N^{-1} T_N(x)$. Theorem 2 states that $\sup_{0 \leqslant x \leqslant 1} |A_N(x) - A(x)| \to^{\mathrm{P}} 0$, where

$$A(x) = \mathbf{g}^{\mathrm{T}}(x) \Sigma^* \mathbf{g}(x) = \begin{cases} x(1-\theta)A, & 0 \leqslant x \leqslant \theta, \\ \theta(1-x)A, & \theta < x < 1, \end{cases}$$

with

$$A = \sum_{1 \leqslant l \leqslant d} \frac{1}{\gamma_l} \left\{ \int \Delta(t)\, w_l(t)\,\mathrm{d}t \right\}^2.$$

Under the assumptions of corollary 1, $A > 0$, and it is easy to verify that $A(x)$ has then a unique maximum at $x = \theta$.

An important aspect of the procedure is the choice of the number d of the eigenfunctions v_k. This issue is common to all FDA procedures using functional PCA, and several approaches have been proposed. These include an adaptation of the *scree plot* of Cattell (1966) (see Kokoszka *et al.* (2008)), the *cumulative percentage variance* approach that is used in Section 4.2, the *pseudo Akaike information criterion* and *cross-validation* (see Yao *et al.* (2005)). All these methods are implemented in the MATLAB PACE package that was developed at the University of California

at Davis. A general recommendation for the cumulative percentage variance method is to use d which explains 85% of the variance. This choice is suitable in the setting of Section 4.2, where $d = 8$ explains 84% of the variance.

4. Finite sample performance and application to temperature data

In this section, we report the results of a simulation study that examines the finite sample performance of the test. Recall that the test rejects if $S_{N,d}$ (3.4) exceeds the $(1-\alpha)$th quantile of K_d (3.5). For $d \leqslant 5$, these quantiles were computed by Kiefer (1959) using a series expansion of the cumulative distribution function of K_d. Horváth *et al.* (1999) used these expansions to find the critical values for $d = 12$ and noticed that the critical values that were obtained by simulating K_d by discretizing the integral are slightly different but actually lead to more accurate tests. To cover a fuller range of the d-values, Table 1 gives simulated critical values for $d = 1, \ldots, 30$, computed by discretizing the integral over 1000 points and running 100 000 replications.

The simulation study consists of two parts. First we use standard Gaussian processes as the errors Y_i and a number of rather arbitrary mean functions μ. This part assesses the test in some generic cases analogously to assuming a normal distribution of scalar observations. In the second part, we use mean functions and errors that are derived from monthly temperature data.

Table 1. Simulated critical values of the distribution of K_d

Nominal size (%)	*Results for the following values of d:*					
	1	*2*	*3*	*4*	*5*	*6*
10	0.345165	0.606783	0.842567	1.065349	1.279713	1.485200
5	0.460496	0.748785	1.001390	1.239675	1.469008	1.684729
1	0.740138	1.072101	1.352099	1.626695	1.866702	2.125950
	7	*8*	*9*	*10*	*11*	*12*
10	1.690773	1.897365	2.096615	2.288572	2.496635	2.686238
5	1.895557	2.124153	2.322674	2.526781	2.744438	2.949004
1	2.342252	2.589244	2.809778	3.033944	3.268031	3.491102
	13	*14*	*15*	*16*	*17*	*18*
10	2.884214	3.066906	3.268958	3.462039	3.650724	3.837678
5	3.147604	3.336262	3.544633	3.740248	3.949054	4.136169
1	3.708033	3.903995	4.116829	4.317087	4.554650	4.734714
	19	*20*	*21*	*22*	*23*	*24*
10	4.024313	4.214800	4.404677	4.591972	4.778715	4.965613
5	4.327286	4.532917	4.718904	4.908332	5.101896	5.303462
1	4.974172	5.156282	5.369309	5.576596	5.759427	5.973941
	25	*26*	*27*	*28*	*29*	*30*
10	5.159057	5.346543	5.521107	5.714145	5.885108	6.083306
5	5.495721	5.688849	5.866095	6.068351	6.242770	6.444772
1	6.203718	6.393582	6.572949	6.771058	6.977607	7.186491

No assumptions on the marginal distribution of the Y_is or the shape of the μs are made. This part assesses the test in a specific, practically relevant setting.

4.1. Gaussian processes

To investigate the empirical size, without loss of generality, $\mu(t)$ was chosen to be equal to 0 and two different cases of $Y_i(t)$ were considered, namely the trajectories of the standard Brownian motion (BM), and the Brownian bridge (BB). These processes were generated by transforming cumulative sums of independent normal variables computed on a grid of 10^3 equispaced points in $[0, 1]$. Following Ramsay and Silverman (2005) (chapter 3) discrete trajectories were converted to functional observations (functional objects in R) by using *B*-spline and Fourier bases and various numbers of basis functions. No systematic dependence either on the type of the basis or on the number of basis functions was found. The results that are reported in this section were obtained by using a *B*-spline basis with 800 basis functions. We used a wide spectrum of N and d, but for brevity we present the results for $N = 50, 150, 200, 300, 500$ and $d = 1, 2, 3, 4$. All empirical rejection rates are based on 1000 replications.

Table 2 shows the empirical sizes based on critical values reported in Table 1. The empirical sizes are fairly stable. Except for a very few cases of small sample sizes, all deviations from the nominal significance levels do not exceed 2 standard errors computed by using the normal approximation $\sqrt{\{p(1-p)/R\}}$, where p is a nominal level and R the number of repetitions.

Table 2. Empirical size of the test using the *B*-spline basis

Process	*Results (%) for the following values of d and nominal sizes:*											
	d=1			*d=2*			*d=3*			*d=4*		
	10%	*5%*	*1%*	*10%*	*5%*	*1%*	*10%*	*5%*	*1%*	*10%*	*5%*	*1%*
N=50												
BM	10.3	4.6	0.1	9.9	4.8	0.7	8.4	3.3	0.6	9.7	4.8	0.8
BB	11.2	5.5	0.8	10.6	4.9	1.1	8.4	4.0	0.9	8.5	4.3	1.2
N=100												
BM	12.2	5.6	1.3	9.8	5.6	0.9	9.3	4.6	0.9	9.0	5.4	0.9
BB	12.4	5.7	0.7	10.2	4.2	0.6	9.9	4.6	1.0	8.3	4.1	0.8
N=150												
BM	10.8	5.7	1.3	9.7	4.6	1.2	11.8	6.2	0.8	10.8	5.3	1.1
BB	10.5	5.0	1.2	9.8	4.4	1.1	10.4	6.2	0.7	10.5	5.1	1.2
N=200												
BM	9.7	5.4	0.8	9.2	4.3	0.7	9.3	5.8	1.3	10.8	5.5	0.9
BB	9.2	5.1	0.8	10.8	5.6	1.2	10.0	5.2	1.0	9.6	5.2	1.0
N=300												
BM	10.3	5.2	1.5	11.1	6.1	0.6	10.1	4.5	0.6	9.9	5.5	0.7
BB	10.4	5.6	1.1	9.4	4.8	0.9	9.9	4.1	0.8	10.5	5.3	1.3
N=500												
BM	11.6	6.3	1.3	10.6	6.9	1.5	10.9	5.7	1.4	9.0	4.4	0.6
BB	11.7	5.1	1.3	9.7	5.8	1.4	10.3	5.3	1.1	10.0	5.4	1.1

Table 3. Empirical power of the test using the B-spline basis and a change-point at $k^* = [n/2]$

Process	*Results (%) for the following values of d and nominal sizes:*								
	d = 1			*d = 2*			*d = 3*		
	10%	*5%*	*1%*	*10%*	*5%*	*1%*	*10%*	*5%*	*1%*
N = 50									
BM; BM + sin(*t*)	81.5	70.8	43.7	72.6	60.0	33.2	67.7	54.9	27.3
BM; BM + *t*	88.4	78.0	54.1	84.7	74.0	45.4	77.5	64.3	36.0
BB; BB + sin(*t*)	99.8	99.4	97.4	100	100	99.9	100	100	100
BB; BB + *t*	99.9	99.8	98.9	100	100	99.9	100	100	100
N = 100									
BM; BM + sin(*t*)	97.4	95.3	86.3	96.4	91.0	76.5	93.5	88.0	68.7
BM; BM + *t*	99.0	97.5	91.2	98.7	97.1	87.6	97.5	94.9	83.8
BB; BB + sin(*t*)	100	100	100	100	100	100	100	100	100
BB; BB + *t*	100	100	100	100	100	100	100	100	100
N = 150									
BM; BM + sin(*t*)	99.9	99.5	96.6	99.6	98.6	95.1	98.9	97.4	90.3
BM; BM + *t*	100	99.8	98.7	99.8	99.7	98.8	99.9	99.7	97.8
BB; BB + sin(*t*)	100	100	100	100	100	100	100	100	100
BB; BB + *t*	100	100	100	100	100	100	100	100	100
N = 200									
BM; BM + sin(*t*)	100	99.9	99.1	100	99.8	99.0	99.9	99.7	98.2
BM; BM + *t*	100	100	100	100	100	99.9	100	100	99.3
BB; BB + sin(*t*)	100	100	100	100	100	100	100	100	100
BB; BB + *t*	100	100	100	100	100	100	100	100	100

Table 2 shows that, for these Gaussian processes, the empirical size does not depend appreciably either on n or on d.

In the power study, several cases that violate the null hypothesis were considered. We report the power for $k^* = [N/2]$. Several other values of k^* were also considered, and only a small loss of power was observed for $N/4 < k^* \leqslant 3N/4$. A few different mean functions μ before and after change were used, namely $\mu_i(t) = 0, t, t^2, \sqrt{t}, \exp(t), \sin(t), \cos(t)$, $i = 1, 2$, e.g. $\mu_1(t) = t$ and $\mu_2(t) = \cos(t)$, etc.

Table 3 presents selected results of the power study. It shows that the test has overall good power. For small samples, $N \leqslant 100$, in cases where the BB was used the power is slightly higher than for those with the BM. Nonetheless, for $N \geqslant 150$ the power approaches 100% for both processes and all choices of other parameters. The power decreases as the number of principal components d increases. This can be explained as follows: the critical values of $S_{N,d}$ increase with d, but the change-point is mainly captured by a few initial leading principal components explaining the major part of the variance.

4.2. Analysis of central England temperatures

The goal of this section is twofold: to investigate the performance of the test in a real world setting, and to demonstrate the advantages of the functional approach for high dimensional data.

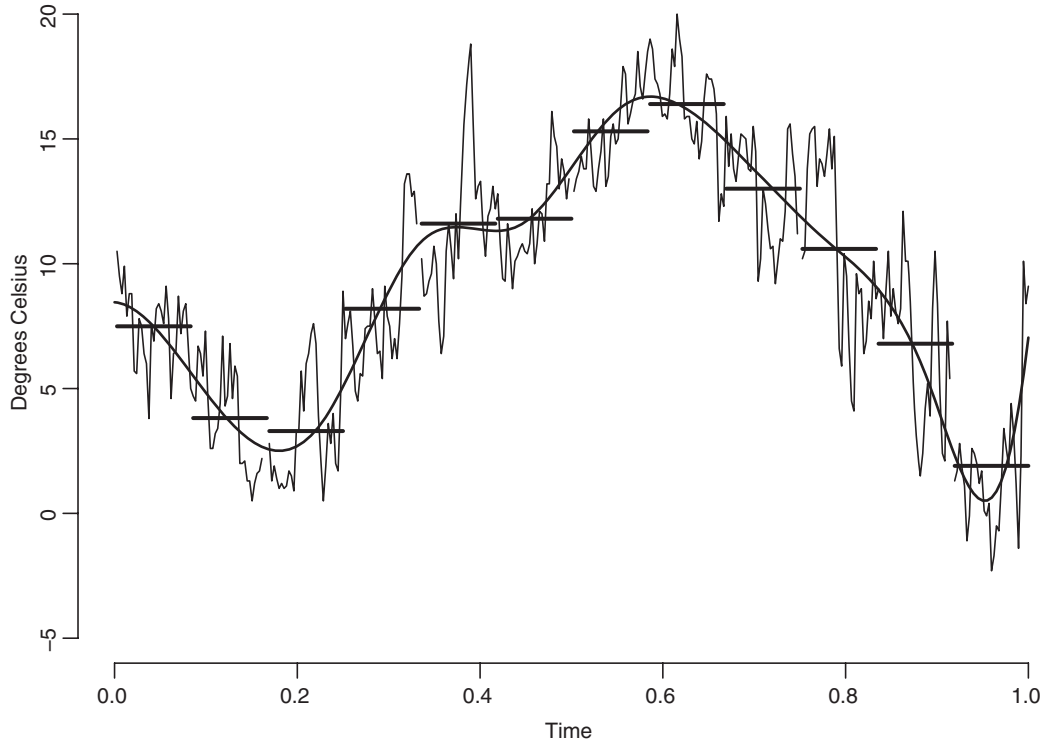

Fig. 1. Daily temperatures in 1916 (∧) with monthly averages (—) and functional observations obtained by smoothing with *B*-splines (———)

The data consist of 228 years (1780–2007) of average daily temperatures in central England. The original data can thus be viewed as 228 curves with 365 measurements on each curve. These data were converted to functional objects in R by using 12 *B*-spline basis functions. Multivariate observations were obtained as in Horváth *et al.* (1999) by computing monthly averages, resulting in 228 vectors of dimension $d = 12$. (We could not even compute statistics (3.6) for vectors of dimension 365 because R reported that $\hat{\mathbf{D}}$ was singular.) These two procedures are illustrated in Fig. 1. Even though we used 12 *B*-splines and 12 averages, the resulting data look quite different, especially in the spring and autumn, when the temperatures change most rapidly. Gregorian months form a somewhat arbitrary fixed partition of the data, whereas the splines adapt to their shapes which differ from year to year.

To compute statistic (3.4), we used $d = 8$ eigenfunctions which explain 84% of variability. If the test indicates a change, we estimate it by the estimator $\hat{\theta}_N$ (3.10). This divides the data set into two subsets. The procedure is then repeated for each subset until periods of constant mean functions are obtained. We proceed in exactly the same manner using statistic (3.6). We refer to these procedures respectively as the FDA and multivariate data analysis (MDA) approaches. The resulting segmentations are shown in Tables 4 and 5.

The functional approach identified two more change-points, 1850 and 1992, which roughly correspond to the beginning of mass industralization and the advent of rapid global warming. The multivariate approach 'almost' identified these change-points with the *P*-values in iterations 4 and 5 being just above the significance level of 5%. This may indicate that the functional method has better power, perhaps due to its greater flexibility in capturing the shape of the data. This

Table 4. Segmentation procedure of the data into periods with constant mean function

Iteration	*Segment*	*Decision*	$S_{N,d}$ *or* $M_{N,d}$	*P-value*	*Estimated change-point*
England temperatures (d = 8) (FDA approach)					
1	1780–2007	Reject	8.020593	0.00000	1926
2	1780–1925	Reject	3.252796	0.00088	1808
3	1780–1807	Accept	0.888690	0.87404	—
4	1808–1925	Reject	2.351132	0.02322	1850
5	1808–1849	Accept	0.890845	0.87242	—
6	1850–1925	Accept	1.364934	0.41087	—
7	1926–2007	Reject	2.311151	0.02643	1993
8	1926–1992	Accept	0.927639	0.84289	—
9	1993–2007	Accept	1.626515	0.21655	—
England temperatures (d = 12) (MDA approach)					
1	1780–2007	Reject	7.971031	0.00000	1926
2	1780–1925	Reject	3.576543	0.00764	1815
3	1780–1814	Accept	1.534223	0.81790	—
4	1815–1925	Accept	2.813596	0.07171	—
5	1926–2007	Accept	2.744801	0.08662	—

Table 5. Summary and comparison of segmentation†

Approach	*Change-points*					
FDA	*1780*	1808	1850	1926	1992	*2007*
MDA	*1780*	1815		1926		*2007*

†The beginning and end of the data period are given in italics.

conjecture is investigated below. Fig. 2 shows average temperatures in the last four segments and clearly illustrates the warming trend.

The analysis that was presented above assumes a simple functional change-point model for the daily temperatures. Obviously, we cannot realistically believe that the mean curves change abruptly in one year; this is merely a modelling assumption that is useful in identifying patterns of change in mean temperature curves. Well-established alternative modelling approaches have been used to study the variability of temperatures. For example, Hosking (1984) fitted a fractionally differenced auto-regressive moving average (ARMA(1,1)) model to the series of *annual* average temperatures in central England in 1659–1976. It is generally very difficult to determine on purely statistical grounds whether a change-point or a long-range dependent model is more suitable for any particular finite length record; see Berkes *et al.* (2006) and Jach and Kokoszka (2008) for recent methodology, discussion and references. It is often more useful to choose a modelling methodology which depends on specific goals, and this is the approach that we use. One way of checking an approximate adequacy of our model is to check whether the residuals that are obtained after subtracting the mean in each segment are approximately independent and identically distributed. This can be done by applying the test that was developed by Gabrys and Kokoszka (2007) which is a functional analogue of the well-known test of Hosking (1980) and Li and McLeod (1981) (see also Hosking (1981)). The *P*-value of 8% indicates the acceptance of the hypothesis that the residuals are independent and identically distributed.

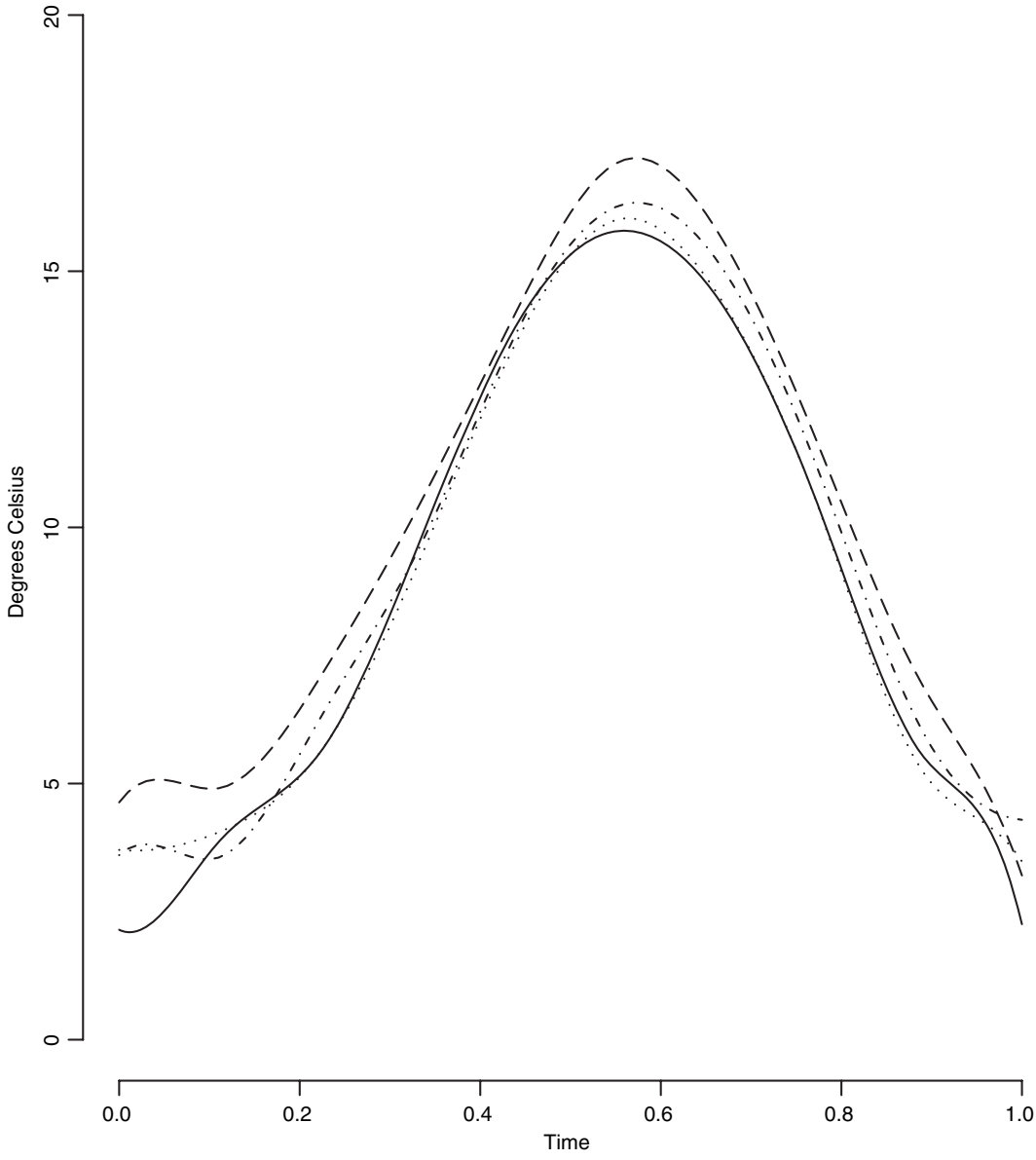

Fig. 2. Average temperature functions in the estimated partition segments: ———, 1808–1849; · · · · · · ·, 1850–1925; · - - - ·, 1926–1991; – – –, 1992–2007

Keeping these *caveats* in mind, we use the partitions that were obtained above to generate realistic synthetic data with and without change-points. We use them to evaluate and compare the size and power properties of the FDA and MDA tests, and to validate our findings. We compute the residuals of every observation in a constant mean segment by subtracting the average of the segment, i.e. $\hat{\mathbf{Y}}_{is} = \mathbf{X}_{is} - \hat{\mu}_s$, where $s = 1, \ldots, S$ denotes the segment, and $i = 1, \ldots, I_s$ indexes observations in the sth segment. The $\hat{\mathbf{Y}}_{is}$ are functional residuals, and their average in each segment is clearly the zero function.

Table 6. Empirical size of the test for models derived from the temperature data

Segment	*Number of functions*	*Sizes for the following cases and nominal sizes:*					
		Case I			*Case II*		
		10%	*5%*	*1%*	*10%*	*5%*	*1%*
FDA approach (d=8)							
1780–1807 (Δ_1)	28	8.0	3.0	0.1	7.6	2.5	0.2
1808–1849 (Δ_2)	42	9.5	3.9	0.4	9.7	4.1	0.4
1850–1925 (Δ_3)	76	10.0	4.7	0.7	10.2	4.3	0.7
1926–1992 (Δ_4)	66	8.8	3.7	0.8	9.2	4.1	1.0
1993–2007 (Δ_5)	16	3.8	0.3	0.0	3.3	0.1	0.0
MDA approach (d=12)							
1780–1807 (Δ_1)	28	3.0	0.5	0.0	2.8	0.4	0.0
1808–1849 (Δ_2)	42	5.3	2.3	0.1	5.4	1.3	0.0
1850–1925 (Δ_3)	76	6.9	1.9	0.0	9.1	4.2	0.6
1926–1992 (Δ_4)	66	7.9	3.3	0.5	7.4	2.7	0.2
1993–2007 (Δ_5)	16	—	—	—	0.0	0.0	0.0

Table 7. Empirical power of the test for change-point models derived from the temperature data for England ($d=8$) (FDA approach)

Segment	*Sample size*	*Change-point(s) θ*	*Powers for the following cases and nominal levels:*								
			Case I			*Case II*			*Case III*		
			10%	*5%*	*1%*	*10%*	*5%*	*1%*	*10%*	*5%*	*1%*
Δ_1, Δ_2	70	0.41	85.6	76.8	49.7	86.4	76.9	46.3	87.0	75.7	45.3
Δ_1, Δ_3	104	0.28	86.2	75.8	47.4	88.6	78.8	50.6	93.1	83.3	58.1
Δ_1, Δ_4	94	0.31	100	100	98.7	100	100	99.3	99.8	99.7	96.3
Δ_1, Δ_5	44	0.66	100	99.9	93.4	100	99.8	92.7	99.8	99.6	92.2
Δ_2, Δ_3	118	0.36	87.9†	78.5	52.8	88.0	78.9	52.1	88.6	79.6	54.0
Δ_2, Δ_4	108	0.40	99.7	99.0	95.6	100	99.6	96.7	100	99.3	95.7
Δ_2, Δ_5	58	0.74	99.2	97.8	86.3	99.4	98.6	85.8	99.6	98.7	86.6
Δ_3, Δ_4	142	0.54	99.9†	99.5†	99.1	100	100	98.9†	99.6	99.1	96.6
Δ_3, Δ_5	92	0.84	99.1	96.7	82.9	99.4	97.4	84.4	98.9	95.4	79.6
Δ_4, Δ_5	82	0.82	93.0	85.0	58.8	94.0	86.3	57.0	77.9	64.9	32.6
$\Delta_1, \Delta_2, \Delta_3$	146	0.20 0.49	99.1	97.9	89.6	99.2	97.0	89.9	99.3	98.5	94.2
$\Delta_1, \Delta_2, \Delta_4$	136	0.21 0.52	100	100	100	100	100	100	100	100	100
$\Delta_1, \Delta_2, \Delta_5$	86	0.34 0.83	100	100	99.7	99.9	99.9	99.2	100	100	99.7
$\Delta_2, \Delta_3, \Delta_4$	184	0.23 0.65	100	100	99.9	100	100	99.9	100	99.9†	99.9
$\Delta_2, \Delta_3, \Delta_5$	134	0.32 0.89	100	99.3†	96.4	99.9	99.8	97.4	100	99.7	97.7
$\Delta_3, \Delta_4, \Delta_5$	158	0.49 0.91	100	100	100	100	100	100	100	100	100
$\Delta_1, \Delta_2, \Delta_3, \Delta_4$	212	0.14 0.33 0.69	100	100	100	100	100	100	100	100	100
$\Delta_1, \Delta_2, \Delta_3, \Delta_5$	162	0.18 0.44 0.91	100	100	99.9	100	100	99.9	100	100	100
$\Delta_2, \Delta_3, \Delta_4, \Delta_5$	200	0.22 0.60 0.93	100	100	100	100	100	100	100	100	100
$\Delta_1, \Delta_2, \Delta_3, \Delta_4, \Delta_5$	228	0.13 0.31 0.64 0.93	100	100	100	100	100	100	100	100	100

†MDA performed better than FDA.

Table 8. Empirical power of the test for change-point models derived from the temperature data for England ($d = 8$) (MDA approach)

Segment	*Sample size*	*Change-point(s) θ*	*Powers for the following cases and nominal levels:*								
			Case I			*Case II*			*Case III*		
			10%	*5%*	*1%*	*10%*	*5%*	*1%*	*10%*	*5%*	*1%*
Δ_1, Δ_2	70	0.41	82.9	70.2	38.2	85.2	73.4	39.3	76.2	59.6	26.8
Δ_1, Δ_3	104	0.28	79.7	63.9	32.6	79.4	64.8	30.5	81.1	67.4	35.1
Δ_1, Δ_4	94	0.31	100	99.4	95.8	99.9	99.0	96.0	99.3	96.9	82.0
Δ_1, Δ_5	44	0.66	98.4	93.8	54.5	99.0	93.0	55.8	98.5	91.8	49.0
Δ_2, Δ_3	118	0.36	88.3	75.9	46.8	86.7	75.6	43.5	82.3	70.7	41.7
Δ_2, Δ_4	108	0.40	97.3	93.3	77.5	97.8	95.6	78.1	98.3	95.7	80.7
Δ_2, Δ_5	58	0.74	93.9	85.5	50.4	94.7	85.2	48.3	96.3	90.9	57.9
Δ_3, Δ_4	142	0.54	100	100	98.5	100	99.8	99.0	99.5	98.9	94.6
Δ_3, Δ_5	92	0.84	98.2	93.9	71.2	99.1	94.2	71.3	96.7	90.2	58.2
Δ_4, Δ_5	82	0.82	78.4	63.1	28.0	79.4	63.4	26.4	60.9	44.1	15.7
$\Delta_1, \Delta_2, \Delta_3$	146	0.20 0.49	97.5	93.2	76.9	97.7	93.1	77.9	97.4	94.9	80.2
$\Delta_1, \Delta_2, \Delta_4$	136	0.21 0.52	100	100	100	100	100	99.9	100	100	99.9
$\Delta_1, \Delta_2, \Delta_5$	86	0.34 0.83	100	99.8	96.2	99.9	99.7	95.7	100	99.8	97.4
$\Delta_2, \Delta_3, \Delta_4$	184	0.23 0.65	100	100	99.1	100	99.9	98.7	100	100	99.5
$\Delta_2, \Delta_3, \Delta_5$	134	0.32 0.89	99.8	99.4	93.7	99.6	99.3	93.8	99.7	98.6	92.1
$\Delta_3, \Delta_4, \Delta_5$	158	0.49 0.91	100	100	100	100	100	100	100	100	100
$\Delta_1, \Delta_2, \Delta_3, \Delta_4$	212	0.14 0.33 0.69	100	100	99.9	100	100	100	100	100	100
$\Delta_1, \Delta_2, \Delta_3, \Delta_5$	162	0.18 0.44 0.91	100	100	99.1	100	99.9	99.1	100	100	98.9
$\Delta_2, \Delta_3, \Delta_4, \Delta_5$	200	0.22 0.60 0.93	100	100	100	100	100	100	100	100	100
$\Delta_1, \Delta_2, \Delta_3, \Delta_4, \Delta_5$	228	0.13 0.31 0.64 0.93	100	100	100	100	100	100	100	100	100

To assess the empirical size, we simulate 'temperature-like' data by considering two cases. *Case I*: for every constant mean segment s, we produce synthetic observations by adding to its mean function $\hat{\mu}_s$ errors drawn from the empirical distribution of the residuals of that segment, i.e. synthetic (bootstrap) observations in the sth segment are generated via $\mathbf{X}^*_{is} = \hat{\mu}_s + \hat{\mathbf{Y}}_{i^*s}$, where i^* indicates that $\hat{\mathbf{Y}}_{i^*s}$ is obtained by drawing with replacement from $\{\hat{\mathbf{Y}}_{is}, i = 1, \ldots, I_s\}$. *Case II*: we compute residuals in each segment and pool them. We use this larger set of residuals to create new observations by adding to the average of a segment the errors drawn with replacement from that pool of residuals. For each segment, we generate 1000 of these bootstrap sequences. Table 6 shows the resulting empirical sizes. As the sample size increases, the FDA rejection rates approach nominal sizes, whereas the MDA test is much more conservative. For the 1993–2007 segment, the size is not reported because the matrix $\mathbf{D}$ was (numerically) singular for most bootstrap replications.

We next investigate the power. Three cases are considered. *Case I*: for each segment, we produce synthetic observations by using the bootstrap procedure and sampling residuals from a corresponding period. This means that the errors in each segment come from possibly different distributions. *Case II*: we pool two, three, four or five sets of residuals (depending on how many constant mean segments we consider) and sample from that pool to produce new observations. This means that the errors in each segment come from the same distribution. *Case III*: we slightly modify case II by combining all residuals from all segments into one population and use it to produce new observations. In both case II and case III, the theoretical assumptions of Section 2 are satisfied (see assumption 4), i.e. the means change, but the errors come from

the same population. Table 7 shows the power of the test for the FDA approach and Table 8 presents results of the discrete MDA method. As seen in Table 7, the differences between the three cases are of the order of chance error. Table 7 shows that the test has excellent power, even in small samples, both for single and for multiple change-points. As for the Gaussian processes, the power is slightly higher if there is a change-point around the middle of the sample. Comparing Tables 7 and 8, it is seen that the FDA approach dominates the MDA approach. There are a handful of cases, indicated with a dagger, when MDA performed better, but their frequency and the difference size suggest that this may be attributable to chance error.

Acknowledgements

István Berkes was partially supported by Ovszánagos Tudományos Kutatási Alapprogramok grants T43037, K61052 and K67961. Lajos Horváth was partially supported by National Science Foundation grant DMS-0604670. Robertas Gabrys and Piotr Kokoszka were partially supported by National Science Foundation grants DMS-0413653 and DMS-0804165. The temperature data that were used in Section 4 were published by the British Atmospheric Data Centre and compiled by Nick Humphreys.

Appendix A: Proof of theorems 1 and 2

A key element of the proofs of theorems 1 and 2 is the bound (A.5), which follows from a functional central limit theorem in a Hilbert space. A result of this type is needed because the observations $X_i(\cdot)$ are elements of a Hilbert space and, to detect a change-point, we must monitor the growth of the partial sums $\Sigma_{1\leqslant i\leqslant Nx}X_i(t)$ which are a function of $0<x<1$ (and of $t\in\mathcal{T}$).

Lemma 1 is particularly noteworthy because it shows that the eigenvalues and the eigenfunctions also converge under the alternative.

A.1. Proof of theorem 1

We shall work with the unobservable projections

$$\tilde{\beta}_{i,k}=\int Y_i(t)\,\hat{v}_k(t)\,\mathrm{d}t,$$

$$\beta_{i,k}=\int Y_i(t)\,v_k(t)\,\mathrm{d}t,$$

$$\beta^*_{i,k}=\hat{c}_k\beta_{i,k}$$

and the vectors

$$\boldsymbol{\beta}_i=(\beta_{i,1},\ldots,\beta_{i,d})^{\mathrm{T}},\qquad \boldsymbol{\beta}^*_i=(\beta^*_{i,1},\ldots,\beta^*_{i,d})^{\mathrm{T}},\qquad 1\leqslant i\leqslant N.$$

Since the Y_i are independent and identically distributed functions with mean 0, the $\boldsymbol{\beta}_i$ are independent and identically distributed mean zero vectors in R^d. A simple calculation using the orthonormality of the v_k shows that each $\boldsymbol{\beta}_i$ has a diagonal covariance matrix

$$\Sigma_d=\begin{pmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \cdots & \lambda_d\end{pmatrix}.$$

The functional central limit theorem thus implies that

$$N^{-1/2}\sum_{1\leqslant i\leqslant Nx}\boldsymbol{\beta}_i\xrightarrow{\mathrm{d}}\boldsymbol{\Delta}_d(x),\qquad 0\leqslant x\leqslant 1,\tag{A.1}$$

where the convergence is in the Skorohod space $D^d[0,1]$. The process $\{\boldsymbol{\Delta}_d(x), 0 \leqslant x \leqslant 1\}$ takes values in R^d, has zero mean and covariance matrix Σ_d. Convergence (A.1) implies in turn that

$$\frac{1}{N}\left(\sum_{1\leqslant i\leqslant Nx}\boldsymbol{\beta}_i - x\sum_{1\leqslant i\leqslant N}\boldsymbol{\beta}_i\right)^{\mathrm{T}}\Sigma_d^{-1}\left(\sum_{1\leqslant i\leqslant Nx}\boldsymbol{\beta}_i - x\sum_{1\leqslant i\leqslant N}\boldsymbol{\beta}_i\right)\xrightarrow{\mathrm{d}}\sum_{1\leqslant i\leqslant d}B_i^2(x) \tag{A.2}$$

in the Skorohod space $D[0,1]$.

The matrix Σ_d is estimated by $\hat{\Sigma}_d$. By expression (2.8) and assumption 2, $\hat{\Sigma}_d^{-1}\to^{\mathrm{P}}\Sigma_d^{-1}$, so result (A.2) yields

$$\frac{1}{N}\left(\sum_{1\leqslant i\leqslant Nx}\boldsymbol{\beta}_i - x\sum_{1\leqslant i\leqslant N}\boldsymbol{\beta}_i\right)^{\mathrm{T}}\hat{\Sigma}_d{}^{-1}\left(\sum_{1\leqslant i\leqslant Nx}\boldsymbol{\beta}_i - x\sum_{1\leqslant i\leqslant N}\boldsymbol{\beta}_i\right)\xrightarrow{\mathrm{d}}\sum_{1\leqslant i\leqslant d}B_i^2(x). \tag{A.3}$$

Note that

$$\sum_{1\leqslant i\leqslant Nx}\beta_{i,k}^* - x\sum_{1\leqslant i\leqslant N}\beta_{i,k}^* = \hat{c}_k\left(\sum_{1\leqslant i\leqslant Nx}\beta_{i,k} - x\sum_{1\leqslant i\leqslant N}\beta_{i,k}\right).$$

Since $\hat{c}_k^2 = 1$, we can replace the $\boldsymbol{\beta}_i$ in result (A.3) by the $\boldsymbol{\beta}_i^*$ and obtain

$$\frac{1}{N}\left(\sum_{1\leqslant i\leqslant Nx}\boldsymbol{\beta}_i^* - x\sum_{1\leqslant i\leqslant N}\boldsymbol{\beta}_i^*\right)^{\mathrm{T}}\hat{\Sigma}_d{}^{-1}\left(\sum_{1\leqslant i\leqslant Nx}\boldsymbol{\beta}_i^* - x\sum_{1\leqslant i\leqslant N}\boldsymbol{\beta}_i^*\right)\xrightarrow{\mathrm{d}}\sum_{1\leqslant i\leqslant d}B_i^2(x). \tag{A.4}$$

We now turn to the effect of replacing the $\beta_{i,k}^*$ by $\tilde{\beta}_{i,k}$. Observe that

$$\begin{aligned}\sup_{0<x<1}\left|N^{-1/2}\sum_{1\leqslant i\leqslant Nx}\beta_{i,k}^* - N^{-1/2}\sum_{1\leqslant i\leqslant Nx}\tilde{\beta}_{i,k}\right| &= \sup_{0<x<1}\left|\int\left\{N^{-1/2}\sum_{1\leqslant i\leqslant Nx}Y_i(t)\right\}\{\hat{c}_k\, v_k(t) - \hat{v}_k(t)\}\,\mathrm{d}t\right| \\ &\leqslant \sup_{0<x<1}\left[\int\left\{N^{-1/2}\sum_{1\leqslant i\leqslant Nx}Y_i(t)\right\}^2\mathrm{d}t\right]^{1/2}\left[\int\{\hat{c}_k\, v_k(t) - \hat{v}_k(t)\}^2\,\mathrm{d}t\right]^{1/2}.\end{aligned}$$

The first factor is bounded in probability, i.e.

$$\sup_{0<x<1}\left[\int\left\{N^{-1/2}\sum_{1\leqslant i\leqslant Nx}Y_i(t)\right\}^2\mathrm{d}t\right] = O_P(1). \tag{A.5}$$

Relationship (A.5) follows from the weak convergence in $D\{[0,1], L^2(\mathcal{T})\}$ of the partial sum process $\Sigma_{1\leqslant i\leqslant Nx}Y_i$, $x\in[0,1]$; see for example Kuelbs (1973).

Combining expressions (A.5) and (2.8), we obtain

$$\sup_{0<x<1}\left|N^{-1/2}\sum_{1\leqslant i\leqslant Nx}\beta_{i,k}^* - N^{-1/2}\sum_{1\leqslant i\leqslant Nx}\tilde{\beta}_{i,k}\right|\xrightarrow{\mathrm{P}}0,$$

which in turn implies that

$$\left\|\left(\sum_{1\leqslant i\leqslant Nx}\boldsymbol{\beta}_i^* - x\sum_{1\leqslant i\leqslant N}\boldsymbol{\beta}_i^*\right) - \left(\sum_{1\leqslant i\leqslant Nx}\hat{\boldsymbol{\eta}}_i - x\sum_{1\leqslant i\leqslant N}\hat{\boldsymbol{\eta}}_i\right)\right\| = o_P(N^{-1/2}), \tag{A.6}$$

where the norm is the Euclidean norm in R^d. Relationships (A.4) and (A.6) yield the claim in theorem 1.

A.2. Proof of theorem 2

Theorem 2 follows from relationship (A.10) and lemma 2. To establish them, we need the following lemma.

Lemma 1. Under assumption 4, for every $1\leqslant k\leqslant d$, as $N\to\infty$,

$$\hat{\lambda}_k\xrightarrow{\mathrm{P}}\gamma_k, \tag{A.7}$$

$$\int\{\hat{v}_k(t) - \hat{c}_k\, w_k(t)\}^2\,\mathrm{d}t\xrightarrow{\mathrm{P}}0, \tag{A.8}$$

where $\hat{v}_k$ and $\hat{\lambda}_k$ are defined by expression (2.7), w_k and γ_k by equation (2.11) and $\hat{c}_k = \mathrm{sgn}\{\int_{\mathcal{T}} v_k(t)\,\hat{v}_k(t)\,\mathrm{d}t\}$.

Proof. It is easy to see that

$$\bar{X}_N(t) = \bar{Y}_N(t) + \frac{k^*}{N}\mu_1(t) + \frac{N-k^*}{N}\mu_2(t)$$

and, denoting $\Delta(t) = \mu_1(t) - \mu_2(t)$,

$$\begin{aligned}
\hat{c}_N(t,s) &= \frac{1}{N}\left(\sum_{1\leqslant i\leqslant k^*} + \sum_{k^*<i\leqslant N}\right)\{X_i(t) - \bar{X}_N(t)\}\{X_i(s) - \bar{X}_N(s)\} \\
&= \frac{1}{N}\sum_{1\leqslant i\leqslant k^*}\left\{Y_i(t) - \bar{Y}_N(t) + \mu_1(t) - \frac{k^*}{N}\mu_1(t) - \frac{N-k^*}{N}\mu_2(t)\right\} \\
&\quad\times\left\{Y_i(s) - \bar{Y}_N(s) + \mu_1(s) - \frac{k^*}{N}\mu_1(s) - \frac{N-k^*}{N}\mu_2(s)\right\} \\
&\quad+\frac{1}{N}\sum_{k^*<i\leqslant N}\left\{Y_i(t) - \bar{Y}_N(t) + \mu_2(t) - \frac{k^*}{N}\mu_1(t) - \frac{N-k^*}{N}\mu_2(t)\right\} \\
&\quad\times\left\{Y_i(s) - \bar{Y}_N(s) + \mu_2(s) - \frac{k^*}{N}\mu_1(s) - \frac{N-k^*}{N}\mu_2(s)\right\} \\
&= \frac{1}{N}\sum_{1\leqslant i\leqslant k^*}\left\{Y_i(t) - \bar{Y}_N(t) + \frac{N-k^*}{N}\Delta(t)\right\}\left\{Y_i(s) - \bar{Y}_N(s) + \frac{N-k^*}{N}\Delta(s)\right\} \\
&\quad+\frac{1}{N}\sum_{k^*<i\leqslant N}\left\{Y_i(t) - \bar{Y}_N(t) - \frac{k^*}{N}\Delta(t)\right\}\left\{Y_i(s) - \bar{Y}_N(s) - \frac{k^*}{N}\Delta(s)\right\}.
\end{aligned}$$

Rearranging terms, we obtain

$$\hat{c}_N(t,s) = \frac{1}{N}\sum_{i=1}^{N}\{Y_i(t) - \bar{Y}_N(t)\}\{Y_i(s) - \bar{Y}_N(s)\} + \frac{k^*}{N}\left(1 - \frac{k^*}{N}\right)\Delta(t)\,\Delta(s) + r_N(t,s),$$

where

$$\begin{aligned}
r_N(t,s) &= \left(1 - \frac{k^*}{N}\right)\frac{1}{N}\sum_{1\leqslant i\leqslant k^*}[\{Y_i(t) - \bar{Y}_N(t)\}\,\Delta(s) + \{Y_i(s) - \bar{Y}_N(s)\}\,\Delta(t)] \\
&\quad+\frac{k^*}{N}\frac{1}{N}\sum_{k^*<i\leqslant N}[\{Y_i(t) - \bar{Y}_N(t)\}\,\Delta(s) + \{Y_i(s) - \bar{Y}_N(s)\}\,\Delta(t)].
\end{aligned}$$

Using the law of large numbers for independent, identically distributed Hilbert-space-valued random variables (see for example theorem 2.4 of Bosq (2000)), we obtain $\int_{\mathcal{T}}\int_{\mathcal{T}} r_N^2(t,s)\,\mathrm{d}t\,\mathrm{d}s \to^{\mathrm{P}} 0$ and

$$\int\int\{\hat{c}_N(t,s) - \tilde{c}_N(t,s)\}^2 \xrightarrow{\mathrm{P}} 0. \tag{A.9}$$

Hence lemmas 4.2 and 4.3 of Bosq (2000) imply respectively results (A.7) and (A.8). □

As an immediate corollary to result (A.7), we obtain

$$\hat{\Sigma}_d^{-1} \xrightarrow{\mathrm{P}} \Sigma^*. \tag{A.10}$$

Lemma 2. Under assumption 4,

$$\sup_{0\leqslant x\leqslant 1}\left|\frac{1}{N}\left(\sum_{1\leqslant i\leqslant Nx}\hat{\eta}_{i,k} - x\sum_{1\leqslant i\leqslant N}\hat{\eta}_{i,k}\right) - \hat{c}_k\,g_k(x)\right| = o_P(1),$$

with the functions g_k defined by expression (3.7).

Proof. Denote

$$\hat{g}_k(x) = \frac{1}{N}\left(\sum_{1\leqslant i\leqslant Nx}\hat{\eta}_{i,k} - x\sum_{1\leqslant i\leqslant N}\hat{\eta}_{i,k}\right), \qquad x\in[0,1],$$

and observe that

$$\hat{\eta}_{i,k} = \int Y_i(t)\,\hat{v}_k(t)\,\mathrm{d}t + \int \mu_1(t)\,\hat{v}_k(t)\,\mathrm{d}t - \int \bar{X}_N(t)\,\hat{v}_k(t)\,\mathrm{d}t, \qquad \text{if } 1 \leqslant i \leqslant k^*$$

and

$$\hat{\eta}_{i,k} = \int Y_i(t)\,\hat{v}_k(t)\,\mathrm{d}t + \int \mu_2(t)\,\hat{v}_k(t)\,\mathrm{d}t - \int \bar{X}_N(t)\,\hat{v}_k(t)\,\mathrm{d}t, \qquad \text{if } k^* < i \leqslant N.$$

We shall use the relationship

$$\sup_{0<x<1} \left| \sum_{1 \leqslant i \leqslant Nx} \int Y_i(t)\,\hat{v}_k(t)\,\mathrm{d}t \right| = O_P(N^{1/2}), \tag{A.11}$$

which follows from equation (A.5).

Suppose first that $0 < x \leqslant \theta$. Then, by results (A.11) and (A.8), uniformly in $x \in [0, 1]$,

$$\begin{aligned}\hat{g}_k(x) &= x(1-\theta)\left\{\int \mu_1(t)\,\hat{v}_k(t)\,\mathrm{d}t - \int \mu_2(t)\,\hat{v}_k(t)\,\mathrm{d}t\right\} + o_P(N^{-1/2}) \\ &= x(1-\theta)\hat{c}_k\left\{\int \mu_1(t)\,w_k(t)\,\mathrm{d}t - \int \mu_2(t)\,w_k(t)\,\mathrm{d}t\right\} + o_P(1).\end{aligned}$$

If $x > \theta$, then, uniformly in $x \in [0, 1]$,

$$\hat{g}_k(x) = \theta(1-x)\hat{c}_k\left\{\int \mu_1(t)\,w_k(t)\,\mathrm{d}t - \int \mu_2(t)\,w_k(t)\,\mathrm{d}t\right\} + o_P(1).$$

References

Antoniadis, A. and Sapatinas, T. (2003) Wavelet methods for continuous time prediction using Hilbert-valued autoregressive processes. *J. Multiv. Anal.*, **87**, 133–158.

Benko, M., Härdle, W. and Kneip, A. (2009) Common functional principal components. *Ann. Statist.*, **37**, 1–34.

Berkes, I., Horváth, L., Kokoszka, P. and Shao, Q.-M. (2006) On discriminating between long-range dependence and changes in mean. *Ann. Statist.*, **34**, 1140–1165.

Bosq, D. (2000) *Linear Processes in Function Spaces*. New York: Springer.

Bugni, F. A., Hall, P., Horowitz, J. L. and Neuman, G. R. (2006) Goodness-of-fit tests for functional data. *Technical Report*. Northwestern University, Evanston.

Cattell, R. B. (1966) The scree test for the number of factors. *J. Multiv. Behav. Res.*, **1**, 245–276.

Cuevas, A., Febrero, M. and Fraiman, R. (2004) An ANOVA test for functional data. *Computnl Statist. Data Anal.*, **47**, 111–122.

Dauxois, J., Pousse, A. and Romain, Y. (1982) Asymptotic theory for principal component analysis of a vector random function. *J. Multiv. Anal.*, **12**, 136–154.

Delicado, P. (2007) Functional k-sample problem when data are density functions. *Computnl Statist.*, **22**, 391–410.

Fernández de Castro, B., Guillas, S. and Gonzáles Manteiga, W. (2005) Functional samples and bootstrap for predicting sulfur dioxide levels. *Technometrics*, **47**, 212–222.

Ferraty, F., Vieu, P. and Viguier-Pla, S. (2007) Factor-based comparison of groups of curves. *Computnl Statist. Data Anal.*, **51**, 4903–4910.

Gabrys, R. and Kokoszka, P. (2007) Portmanteau test of independence for functional observations. *J. Am. Statist. Ass.*, **102**, 1338–1348.

Glendinning, R. H. and Fleet, S. L. (2007) Classifying functional time series. *Signal Process.*, **87**, 79–100.

Horváth, L., Kokoszka, P. S. and Steinebach, J. (1999) Testing for changes in multivariate dependent observations with applications to temperature changes. *J. Multiv. Anal.*, **68**, 96–119.

Hosking, J. R. M. (1980) The multivariate portmanteau statistic. *J. Am. Statist. Ass.*, **75**, 602–608.

Hosking, J. R. M. (1981) Equivalent forms of the multivariate portmanteau statistics. *J. R. Statist. Soc.* B, **43**, 261–262; correction, **51** (1989), 303.

Hosking, J. R. M. (1984) Modeling persistence in hydrological time series using fractional differencing. *Wat. Resour. Res.*, **20**, 1898–1908.

Indritz, J. (1963) *Methods in Analysis*. New York: Macmillan.

Jach, A. and Kokoszka, P. (2008) Wavelet domain test for long-range dependence in the presence of a trend. *Statistics*, **42**, 101–113.

Kiefer, J. (1959) K-sample analogues of the Kolmogorov-Smirnov and Cramér-v. Mises tests. *Ann. Math. Statist.*, **30**, 420–447.

Kokoszka, P., Maslova, I., Sojka, J. and Zhu, L. (2008) Testing for lack of dependence in functional linear model. *Can. J. Statist.*, **36**, 207–222.

Kuelbs, J. (1973) The invariance principle for Banach space valued random variables. *J. Multiv. Anal.*, **3**, 161–172.
Laukaitis, A. and Račkauskas, A. (2005) Functional data analysis for clients segmentation tasks. *Eur. J. Oper. Res.*, **163**, 210–216.
Lavielle, M. and Teyssiére, G. (2006) Detection of multiple change-points in multivariate time series. *Lith. Math. J.*, **46**, 287–306.
Li, W. K. and McLeod, A. I. (1981) Distribution of the residual autocorrelations in multivariate ARMA time series models. *J. R. Statist. Soc.* B, **43**, 231–239.
Müller, H.-G. and Stadtmüller, U. (2005) Generalized functional linear models. *Ann. Statist.*, **33**, 774–805.
Qu, Z. and Perron, P. (2007) Estimating and testing structural changes in multivariate regressions. *Econometrica*, **75**, 459–502.
Rackauskas, A. and Suquet, C. (2006) Testing epidemic changes of infinite dimensional parameters. *Statist. Inf. Stochast. Process.*, **9**, 111–134.
Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*. New York: Springer.
Ramsay, J. O., Wickham, H. and Graves, S. (2007) fda: functional data analysis. *R Package Version 1.2.3*. Vienna: R Foundation for Statistical Computing.
R Development Core Team (2008) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Srivastava, M. S. and Worsley, K. J. (1986) Likelihood ratio tests for a change in the multivariate normal mean. *J. Am. Statist. Ass.*, **81**, 199–204.
Yao, F., Müller, H.-G. and Wang, J.-L. (2005) Functional linear regression analysis for longitudinal data. *Ann. Statist.*, **33**, 2873–2903.
Zamba, K. D. and Hawkins, D. M. (2006) A multivariate change-point model for statistical process control. *Technometrics*, **48**, 539–549.