The Canadian Journal of Statistics
Vol. 37, No. 4, 2009, Pages 571–591
La revue canadienne de statistique

Two sample inference in functional linear models

Lajos HORVÁTH[1]*, Piotr KOKOSZKA[2] and Matthew REIMHERR[3]

[1] *Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA*
[2] *Department of Mathematics and Statistics, Utah State University, Logan, UT 84322, USA*
[3] *Department of Statistics, University of Chicago, Chicago, IL 60637, USA*

Key words and phrases: Functional linear model; significance test.

MSC 2000: Primary 62J15; Secondary 62G08 and 62F03.

Abstract: We propose a method of comparing two functional linear models in which explanatory variables are functions (curves) and responses can be either scalars or functions. In such models, the role of parameter vectors (or matrices) is played by integral operators acting on a function space. We test the null hypothesis that these operators are the same in two independent samples. The complexity of the test statistics increases as we move from scalar to functional responses and relax assumptions on the covariance structure of the regressors. They all, however, have an asymptotic chi-squared distribution with the number of degrees of freedom which depends on a specific setting. The test statistics are readily computable using the `R` package `fda`, and have good finite sample properties. The test is applied to egg-laying curves of Mediterranean flies and to data from terrestrial magnetic observatories. *The Canadian Journal of Statistics* 37: 571–591; 2009 © 2009 Statistical Society of Canada

Résumé: Nous proposons une méthode pour comparer deux modèles linéaires fonctionnels, pour lesquels les variables explicatives sont des fonctions (courbes) et les réponses sont ou bien des scalaires ou bien des fonctions. Pour ces modèles, le rôle des vecteurs ou des matrices de paramètres est joué par des opérateurs d'intégration sur un espace de fonctions. Nous testons l'hypothèse nulle que ces opérateurs sont les mêmes pour les deux échantillons indépendants. La complexité des statistiques de test s'accroit quand nous passons de réponses scalaires à des réponses fonctionnelles, et relachons les hypoth'eses sur la structure de covariance des régresseurs. Ces tests ont cependant une distribution asymptotique qui s uit une loi du Chi-deux, dont le nombre de degrés de liberté dépend d'un cadre particulier. Ces statistiques de test sont calculables avec le programme fda du logiciel R, et ont de bonnes propriétés en petit échantillon. La statistique de test est appliquée à des courbes de pontes de mouches Méditerranéennes, et à des données d'observatoires magnétiques terrestres. *La revue canadienne de statistique* 37: 571–591; 2009 © 2009 Société statistique du Canada

1. INTRODUCTION

In the last decade, methods of functional data analysis have been found useful in a number of applied fields, including environmental science, finance, genetics, biology, public health, geophysics, image and signal processing, to name just a few, see, for example, Besse, Cardot & Stephenson (2000), Damon & Guillas (2002), Laukaitis & Račkauskas (2002), Ando, Imoto & Miyano (2004), Fernández de Castro, Guillas & Gonzáles Manteiga (2005), Morris et al. (2006), Müller, Stadtmüller & Yao (2006), Hlubinka & Prchal (2007), Nerini & Ghattas (2007), Febrero, Galeano & González-Manteiga (2008), and Glendinning & Fleet (2007). The functional linear model, in its various forms, stands out as a particularly useful tool in many such analyses, and has consequently been thoroughly studied, see Cuevas, Febrero & Fraiman (2002), Malfait & Ramsay

* *Author to whom correspondence may be addressed.*
E-mail: horvath@math.utah.edu

(2003), Cardot et al. (2003), Chiou, Müller & Wang (2004), Müller& Stadtmüller (2005), Yao, Müller & Wang (2005), Cai & Hall (2006), Chiou & Müller (2007), Li & Hsing (2007), and among others.

This paper is concerned with the following problem of functional data analysis. We observe two samples: sample 1: (X_i, Y_i), $1 \le i \le N$, and sample 2: (X_j^*, Y_j^*), $1 \le j \le M$. The explanatory variables X_i and X_j^* are functions over a compact subset of a Euclidean space, whereas the responses Y_i and Y_j^* can be either functions or scalars (the Y_i and Y_j^* are either both functions, or both scalars). We model the dependence of the Y_i (Y_j^*) on the X_i (X_j^*) by the functional regression models

$$Y_i = \Psi X_i + \varepsilon_i, \qquad Y_j^* = \Psi^* X_j^* + \varepsilon_j^*,$$

where Ψ and Ψ^* are linear operators whose domain is a function space, and which take values either in the same function space or in the real line. We wish to test if the operators Ψ and Ψ^* are equal.

The above testing problem is directly motivated by our work with terrestrial magnetometer data. Specifically, we have been interested in finding out if interactions of currents of charged particles flowing in the ionosphere and the magnetosphere in various states of the near Earth space and at various times are statistically significantly different. The problem is truly functional, as the *shapes* of the daily magnetogram curves have been used for almost one hundred years to describe various electrodynamic phenomena in near Earth space. Digital observatories measure the components of the magnetic field every 5 s. In this application, which is described in some detail in Section 4, the model is fully functional (both responses and regressors are curves). The case of scalar responses is also included because of its importance in many applications, see, for example, Cardot et al. (2003), Müller & Stadtmüller (2005) and Ferraty & Vieu (2006). Like classical two sample procedures in various forms, our methodology is likely to be applicable to a wide range of problems, where estimating two significantly different functional linear regressions on subsamples of a larger sample may reveal additional features.

Even in the traditional multivariate setting, except for comparing mean responses, the problem of comparing the regression coefficients for two models based on two different samples is not trivial, and we could not find a ready reference for it. In the functional setting, it becomes even more complex, as the dimension of the data must be reduced, and the resulting approximation error must be controlled.

We achieve it by expanding the data with respect to the functional principal components (FPCs), see for example, Chapter 8 of Ramsay & Silverman (2005), when the covariance structures of the covariates are equal. In the case where the structures are unequal we expand both samples using the same, but arbitrary bases. This provides some useful flexibility in the procedure. We suggest, as one possibility, a pooling procedure to obtain bases to expand the data. Generally, due to different FPCs structures, working with two functional samples is difficult, but important theoretical advances, mostly in the context of canonical correlation analysis, have recently been made, see He, Müller & Wang (2003), Cupidon et al. (2007), and Eubank & Hsing (2008).

Our paper aims at developing a practical testing procedure, but some complexity in the form of the test statistics and the derivation of their asymptotic properties is unavoidable. Our theory is developed using population, rather than estimated FPCs. Estimation of the FPCs introduces additional terms, and would make the arguments even longer. The additional terms are generally asymptotically or practically negligible, and the simulations we present here show that this is indeed the case in our setting. Gabrys & Kokoszka (2007), Horváth, Hušková & Kokoszka (in press), Berkes et al. (in press), and among others, show how the estimation of the FPCs is handled in simpler settings.

Other approaches to comparing two samples of curves, but not samples of dependent pairs of curves, have been explored in Laukaitis & Račkauskas (2005), Ferraty, Vieu & Viguier-Pla (2007), and Benko, Härdle & Kneip (2009).

After introducing the requisite concepts and notation in Section 2, we describe the testing procedures in Section 3. Applications to medfly and magnetometer data are presented in Section 4. The asymptotic results and their proofs are collected in Section 5.

2. PRELIMINARIES

We state the assumptions and introduce the notation for the observations $(X_i, Y_i), i = 1, 2, \ldots, N$. The observations (X_j^*, Y_j^*), $j = 1, 2, \ldots, M$, are assumed to satisfy the same assumptions with the superscript $*$ added to the corresponding quantities. The two samples are assumed to be independent.

Throughout this paper we take $\langle \cdot, \cdot \rangle$ to be the $L^2([0, 1])$ inner product.

Assumption 2.1. *The observations $\{X_n\}$ are i.i.d. mean zero random functions in $L^2([0, 1])$ satisfying*

$$\mathrm{E}\|X_n\|^4 = E\left[\int X_n^2(t)\,\mathrm{d}t\right]^2 < \infty.$$

For the linear model with scalar responses, we formulate the following assumption.

Assumption 2.2. *The scalar responses Y_i satisfy*

$$Y_i = \int_0^1 \psi(s)X_i(s)\,\mathrm{d}s + \varepsilon_i,$$

with i.i.d. mean zero errors ε_i satisfying $\mathrm{E}\varepsilon_i^4 < \infty$, and $\psi \in L^2([0, 1])$. The errors ε_i and the regressors X_i are independent.

In the case of functional responses, we define an analogous assumption.

Assumption 2.3. *The functional responses $Y_i \in L^2([0, 1])$ satisfy*

$$Y_i(t) = \int_0^1 \psi(t, s)X_i(s)\,\mathrm{d}s + \varepsilon_i(t),$$

with i.i.d. mean zero errors ε_i satisfying

$$\mathrm{E}\|\varepsilon_n\|^4 = E\left[\int \varepsilon_n^2(t)\,\mathrm{d}t\right]^2 < \infty,$$

and $\psi \in L^2([0, 1] \times [0, 1])$. The errors ε_i and the regressors X_i are independent.

Our objective is to test

$$H_0 : \|\psi - \psi^*\| = 0$$

against

$$H_A : \|\psi - \psi^*\| \neq 0,$$

where the norm is in $L^2([0, 1])$ under Assumption 2.2 and in $L^2([0, 1] \times [0, 1])$ under Assumption 2.3.

The covariance operator of the X_n is defined by

$$C(x) = E[\langle X_n, x\rangle\, X_n], \quad x \in L^2([0, 1])$$

and its eigenfunctions v_j and eigenvalues λ_j by $C(v_j) = \lambda_j v_j$, $j \geq 1$. The eigenfunctions v_j form an orthonormal basis of $L^2([0, 1])$. We also call them the FPCs.

The empirical covariance operator is defined by

$$\hat{C}(x) = \frac{1}{N} \sum_{n=1}^{N} \langle X_n, x\rangle\, X_n, \quad x \in L^2([0, 1]),$$

and its eigen elements by $\hat{C}(\hat{v}_j) = \hat{\lambda}_j \hat{v}_j$, $j = 1, 2, \ldots, N$.

Since the operators C and $\hat{C}$ are symmetric and nonnegative definite, the eigenvalues λ_j and $\hat{\lambda}_j$ are nonnegative. Following the usual convention, we assume that $\lambda_1 > \lambda_2 > \cdots$ with the same ordering for the empirical eigenvalues. Recall that $\sum_{j=1}^{\infty} \lambda_j = E\|X_n\|^2$, see, for example, Section 1.5 of Bosq (2000), so v_1 explains the largest portion of the variance of X, with the subsequent FPCs explaining decreasing contributions to the variance. Our asymptotic results require that the first p eigenvalues are nonzero and distinct, so we impose an additional assumption.

Assumption 2.4. *The eigenvalues of the covariance operator C satisfy*

$$\lambda_1 > \lambda_2 > \cdots \lambda_p > \lambda_{p+1}.$$

3. TESTING PROCEDURE

In this section we introduce the testing procedure. We start with the easier case of scalar responses, and then generalize to the more technical case of functional response variables, which uses the results developed in the scalar setting. A reader interested only in the practical application can start with Section 3.3, which provides an outline.

3.1. Scalar Responses

Since $\psi \in L^2([0, 1])$, we can expand it as $\psi(s) = \sum_{i=1}^{\infty} \mu_i v_i(s)$, where $\mu_i = \langle \psi, v_i\rangle$. Consequently, the response variables can be expressed as $Y_i = \sum_{k=1}^{\infty} \mu_k \langle X_i, v_k\rangle + \varepsilon_i$. We truncate the above expansion at $1 \leq p < \infty$, and combine the error made by the truncation with the ε_i. Defining $\varepsilon_i' = \varepsilon_i + \sum_{k=p+1}^{\infty} \mu_k \langle X_i, v_k\rangle$, the response is equivalently given by

$$Y_i = \sum_{k=1}^{p} \mu_k \langle X_i, v_k\rangle + \varepsilon_i'. \tag{1}$$

In terms of matrix and vector notation we have

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\varepsilon}',$$

where, for $1 \leq i \leq N$ and $1 \leq j \leq p$,

$$\mathbf{Y}(i) = Y_i, \qquad \mathbf{X}(i, j) = \left\langle X_i, v_j \right\rangle, \qquad \boldsymbol{\mu}(j) = \mu_j, \qquad \boldsymbol{\varepsilon}'(i) = \varepsilon_i'.$$

The least squares estimator for $\boldsymbol{\mu}$ is therefore

$$\hat{\boldsymbol{\mu}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}. \tag{2}$$

By Theorem 5.1, $\hat{\boldsymbol{\mu}}$ is a consistent estimator of $\boldsymbol{\mu}$, and for the second sample, the analogously defined $\hat{\boldsymbol{\mu}}^*$ a consistent estimator of $\boldsymbol{\mu}^*$. Thus we can base a test statistic on the difference $\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}^*$. To motivate our construction, assume first that the covariance operators of the X_i and X_j^* are equal and the errors ε_i and ε_j^* have equal variances, that is,

$$\mathrm{E}(X_1(s)X_1(t)) = \mathrm{E}(X_1^*(s)X_1^*(t)) = c(s, t) \tag{3}$$

and

$$\mathrm{var}(\varepsilon_1) = \mathrm{var}(\varepsilon_1^*). \tag{4}$$

In order to develop meaningful asymptotic results we assume that the two samples sizes are of roughly the same order. That is, we assume there exists a constant $0 < \zeta < \infty$ such that

$$\frac{N}{M} \to \zeta, \quad N \to \infty. \tag{5}$$

Finally we introduce the random variable

$$\Lambda_p = N(1+\zeta)^{-1}(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}^*)^{\mathrm{T}}\boldsymbol{\Sigma}_p^{-1}(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}^*), \tag{6}$$

where $\boldsymbol{\Sigma}_p$ is the common asymptotic covariance matrix of $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\mu}}^*$ defined by

$$\boldsymbol{\Sigma}_p(i, i) = \lambda_i^{-1}\sigma^2 + \lambda_i^{-2}\,\mathrm{var}\left(\langle X_1, v_i\rangle \sum_{k=p+1}^{\infty} \mu_k \langle X_1, v_k\rangle\right), \quad i = 1, \ldots, p; \tag{7}$$

$$\boldsymbol{\Sigma}_p(i, j) = \lambda_i^{-1}\lambda_j^{-1}\mathrm{E}\left(\langle X_1, v_i\rangle \left\langle X_1, v_j\right\rangle \left(\sum_{k=p+1}^{\infty} \mu_k \langle X_1, v_k\rangle\right)^2\right), \quad i \neq j. \tag{8}$$

The estimation of $\boldsymbol{\Sigma}_p$ and Λ_p is discussed later on in this section.

The following theorem, which follows immediately from Theorem 5.2, shows that Λ_p has a chi-squared asymptotic distribution.

Theorem 3.1. *Suppose Assumptions* 2.1, 2.2, 2.4*, and conditions* (3), (4)*, and* (5) *hold. Then,* $\Lambda_p \xrightarrow{d} \chi^2(p)$*, as* $N \to \infty$*, where* Λ_p *is defined by* (6)*, and* $\chi^2(p)$ *is a chi-squared random variable with p degrees of freedom.*

We therefore propose the following test statistic when the covariances are equal

$$\hat{\Lambda}_p = N\left(1 + \frac{N}{M}\right)^{-1}(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}^*)^{\mathrm{T}}(\hat{\boldsymbol{\Sigma}}_p)^{-1}(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}^*). \tag{9}$$

We use the empirical diagonal approximation to the matrix $\mathbf{\Sigma}_p$ given by

$$\hat{\mathbf{\Sigma}}_p = \hat{\sigma}^2 \begin{bmatrix} \hat{\lambda}_1^{-1} & 0 & \cdots & 0 \\ 0 & \hat{\lambda}_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \hat{\lambda}_p^{-1} \end{bmatrix};$$

where $\hat{\sigma}$ are residual standard deviations from the estimated regression models. Here the estimates $\hat{\sigma}, \hat{\lambda}_1, \ldots, \hat{\lambda}_p$ are all computed with respect to the pooled sample.

In many applications, the covariance kernels $c(s, t)$ and $c^*(s, t)$ are not necessarily equal. Since the two kernels can have potentially different eigenfunctions we instead consider an arbitrary basis $\{w_i\}$ of $L^2([0, 1])$. The kernels ψ and ψ^* are expanded as

$$\psi(s) = \sum_{i=1}^{\infty} \mu_i w_i(s), \qquad \psi^*(s) = \sum_{j=1}^{\infty} \mu_j^* w_j(s), \tag{10}$$

and so

$$Y_i = \sum_{k=1}^{\infty} \mu_k \langle X_i, w_k \rangle + \varepsilon_i, \qquad Y_j^* = \sum_{k=1}^{\infty} \mu_k^* \left\langle X_j^*, w_k \right\rangle + \varepsilon_j^*.$$

Truncating both sums at p, the response variables can again be expressed as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\varepsilon}', \qquad \mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\mu}^* + \boldsymbol{\varepsilon}'^*,$$

with all terms analogously defined with respect to our new basis. While this appears similar to our prior calculations, we are expanding with respect to an arbitrary basis which means that $\mathbf{X}$ and $\boldsymbol{\varepsilon}'$ are now potentially correlated. The least squares estimators take the same form

$$\hat{\boldsymbol{\mu}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}, \qquad \hat{\boldsymbol{\mu}}^* = (\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\mathbf{Y}^*.$$

Thus we can once again compare $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\mu}}^*$ to test the null hypothesis. To analyze the asymptotic behaviour of these estimates we consider the relation

$$\hat{\boldsymbol{\mu}} = \boldsymbol{\mu} + (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}'.$$

The vector $\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}'$ can be expressed as

$$\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}' = \mathbf{A} + \mathbf{B} + N\mathbf{m},$$

where

$$\mathbf{A} = \begin{pmatrix} \displaystyle\sum_{i=1}^{N} \varepsilon_i \langle X_i, w_1 \rangle \\ \vdots \\ \displaystyle\sum_{i=p}^{N} \varepsilon_i \left\langle X_i, w_p \right\rangle \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \displaystyle\sum_{i=1}^{N} \sum_{k=p+1}^{\infty} \mu_k \left(\langle X_i, w_1 \rangle \langle X_i, w_k \rangle - E[\langle X_1, w_1 \rangle \langle X_1, w_k \rangle] \right) \\ \vdots \\ \displaystyle\sum_{i=1}^{N} \sum_{k=p+1}^{\infty} \mu_k \left(\left\langle X_i, w_p \right\rangle \langle X_i, w_k \rangle - E[\left\langle X_1, w_p \right\rangle \langle X_1, w_k \rangle] \right) \end{pmatrix},$$

have mean zero and are uncorrelated since the error terms are independent of the explanatory functions. The term $\mathbf{m}$ represents the bias introduced by using an arbitrary basis which is given by

$$\mathbf{m} = \begin{pmatrix} \sum_{k=p+1}^{\infty} \mu_k E[\langle X_1, w_1 \rangle \langle X_1, w_k \rangle] \\ \vdots \\ \sum_{k=p+1}^{\infty} \mu_k E[\langle X_1, w_p \rangle \langle X_1, w_k \rangle] \end{pmatrix}.$$

This yields the form

$$\hat{\boldsymbol{\mu}} = \boldsymbol{\mu} + (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{A} + (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{B} + N(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{m}.$$

Clearly $\mathbf{A}$ and $\mathbf{B}$ are sums of i.i.d. random vectors with means zero and finite covariance matrices due to Assumptions 2.1 and 2.2. Thus by the multivariate central limit theorem $N^{-1/2}(\mathbf{A}\ \mathbf{B})^{\mathrm{T}}$ is asymptotically normal. We have by the strong law of large numbers that

$$N^{-1} \sum_{i=1}^{N} \langle X_i, w_j \rangle \langle X_i, w_k \rangle \overset{a.s.}{\to} E \langle X_1, w_j \rangle \langle X_1, w_k \rangle,$$

for $j = 1, \ldots, p$ and $k = 1, \ldots, p$, or in matrix notation

$$N^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X} \overset{a.s.}{\to} \boldsymbol{\Sigma}_1,$$

where the (j, k) entry of $\boldsymbol{\Sigma}_1$ is $E \langle X_1, w_j \rangle \langle X_1, w_k \rangle$. Thus by Slutsky's Lemma $N^{-1/2}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu} - N(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{m})$ is asymptotically normal. Since $\mathbf{A}$ has zero mean, we have that the (i, j) entry of its covariance matrix is given by

$$E \sum_{k=1}^{N} \varepsilon_k \langle w_i, X_k \rangle \varepsilon_k \langle w_j, X_k \rangle = N\sigma^2 E \langle w_i, X_1 \rangle \langle w_j, X_1 \rangle,$$

and therefore

$$\mathrm{cov}(\mathbf{A}) = N\sigma^2 \boldsymbol{\Sigma}_1.$$

Turning to $\mathbf{B}$, the (i, j) entry of its covariance matrix is given by

$$N \sum_{k=p+1}^{\infty} \sum_{r=p+1}^{\infty} \mu_k \mu_r E \left\{ (\langle X_1, w_i \rangle \langle X_1, w_k \rangle - E[\langle X_1, w_i \rangle \langle X_1, w_k \rangle]) \right.$$
$$\left. (\langle X_1, w_j \rangle \langle X_1, w_r \rangle - E[\langle X_1, w_j \rangle \langle X_1, w_r \rangle]) \right\}.$$

We will denote the covariance matrix of $\mathbf{B}$ as $N\boldsymbol{\Sigma}_2$. Combining everything, we have by Slutsky's Lemma

$$N^{1/2} \left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu} - N(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{m} \right) \overset{d}{\to} N(0, \mathbf{C}),$$

where $\mathbf{C} = \sigma^2 \boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}$.

An identical argument gives, for the second sample,

$$M^{1/2}\left(\hat{\boldsymbol{\mu}}^* - \boldsymbol{\mu}^* - M(\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{m}^*\right) \xrightarrow{d} N(0, \mathbf{C}^*),$$

with all terms analogously defined. Combined with assumption 5 we therefore conclude

$$N^{1/2}\left(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}^* - (N(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{m} - M(\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{m}^*)\right) \xrightarrow{d} N(0, \mathbf{C} + \zeta\mathbf{C}^*).$$

Neglecting the biases $\mathbf{m}$ and $\mathbf{m}^*$, we thus arrive at the test statistic

$$\hat{\Lambda}_p = N(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}^*)^{\mathrm{T}}\left(\hat{\sigma}^2\hat{\boldsymbol{\Sigma}}_1 + (N/M)\hat{\sigma}^{*2}\hat{\boldsymbol{\Sigma}}_1^*\right)^{-1}(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}^*), \tag{11}$$

where $\hat{\sigma}^2$ and $\hat{\sigma}^{*2}$ are residual standard deviations from the regression models for the first and second sample respectively. The matrix $\boldsymbol{\Sigma}_1$ is now estimated with

$$\hat{\boldsymbol{\Sigma}}_1 = N^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X},$$

with $\hat{\boldsymbol{\Sigma}}_1^*$ defined analogously.

The distributions of statistics (9) and (11) are approximated by the chi-squared distribution with p degrees of freedom. If p is large (in terms of the percentage of variance explained), then all neglected terms are close to 0.

3.2. Functional Responses

Turning to the case when the response variables are functions, we now assume that ε_i is a random element of $L^2([0, 1])$ and $\psi(t, s) \in L^2([0, 1] \times [0, 1])$ is fixed, but unknown. Our goal is to use dimension reduction techniques to make this problem approachable. Our method will consist of choosing a basis to project the Y_i onto, and then using the results developed in the scalar case.

We first focus on the case of equal variances defined by assumptions (3) and, in place of (4),

$$\mathrm{E}(\varepsilon_1(s)\varepsilon_1(t)) = \mathrm{E}(\varepsilon_1^*(s)\varepsilon_1^*(t)). \tag{12}$$

Consider an arbitrary orthonormal basis $\{u_i\}_{i=1}^{\infty}$ for $L^2([0, 1])$ (on which the Y_i are to be projected), and analogous basis $\{u_j^*\}_{j=1}^{\infty}$. Though all our results hold for an arbitrary choice for $\{u_i\}_{i=1}^{\infty}$, we will use in our applications the eigenfunctions of the covariance operator for the $\{Y_i\}$, with the $\{u_i^*\}$ defined analogously. Because $\{u_i\}$ and $\{v_i\}$ are both bases for $L^2([0, 1])$, it follows that we can construct a basis for $L^2([0, 1] \times [0, 1])$ using the bivariate functions $u_i(t)v_j(s)$ for $(t, s) \in [0, 1] \times [0, 1]$, $i = 1, \ldots, \infty$, and $j = 1, \ldots, \infty$. We therefore have the expansion $\psi(t, s) = \sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\mu_{kl}u_l(t)v_k(s)$, but we will work with the approximation

$$\hat{\psi}(t, s) = \sum_{k=1}^{p}\sum_{l=1}^{r}\hat{\mu}_{k,l}\hat{u}_l(t)\hat{v}_k(s),$$

where $1 \leq r < \infty$ and $1 \leq p < \infty$ are fixed.

Extending the notation introduced in the case of scalar responses, define the matrices

$$\mathbf{Y}(i, j) = \left\langle Y_i, u_j \right\rangle, \quad i = 1, \ldots, N,\ j = 1, \ldots, r,$$
$$\mathbf{X}(i, j) \quad = \left\langle X_i, v_j \right\rangle, \quad i = 1, \ldots, N,\ j = 1, \ldots, p,$$

$$\boldsymbol{\mu}(i,j) = \int_0^1 \int_0^1 \psi(t,s) v_i(s) u_j(t)\,\mathrm{d}s\,\mathrm{d}t, \quad i = 1, \ldots p, \ j = 1, \ldots r.$$

As in the scalar case, we combine any errors made by our approximations with the error of the model, so we also introduce the matrix

$$\boldsymbol{\varepsilon}'(i,j) = \left\langle \varepsilon_i, u_j \right\rangle + \sum_{k=p+1}^{\infty} \mathbf{X}(i,k)\boldsymbol{\mu}(k,j), \quad i = 1, \ldots, N, \ j = 1, \ldots, r.$$

Projecting the relation $Y_i = \Psi X_i + \varepsilon_i$ onto the u_j, we obtain

$$\left\langle Y_i, u_j \right\rangle = \left\langle \Psi X_i, u_j \right\rangle + \left\langle \varepsilon_i, u_j \right\rangle = \sum_{k=1}^{\infty} \left\langle X_i, v_k \right\rangle \left\langle \Psi v_k, u_j \right\rangle + \left\langle \varepsilon_i, u_j \right\rangle$$

which implies

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\varepsilon}'. \tag{13}$$

The corresponding least squares estimator $\hat{\boldsymbol{\mu}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$ consistently estimates the matrix $\boldsymbol{\mu}$. This follows immediately by applying Theorem 5.1 to each column of $\hat{\boldsymbol{\mu}}$. Asymptotic normality follows from Theorem 5.4.

Since $\boldsymbol{\mu}$ is now a matrix, the task of constructing a quadratic form leading to a test statistic is somewhat painful notationally. We start by writing $\boldsymbol{\mu}$ as a column vector of length pr:

$$\boldsymbol{\mu}_v^{\mathrm{T}} = \mathrm{vec}(\boldsymbol{\mu})^{\mathrm{T}} = (\boldsymbol{\mu}(1,1), \boldsymbol{\mu}(2,1), \ldots, \boldsymbol{\mu}(p,1), \boldsymbol{\mu}(1,2), \ldots, \boldsymbol{\mu}(p-1,r), \boldsymbol{\mu}(p,r)). \tag{14}$$

In words, $\boldsymbol{\mu}_v$ is constructed by placing the columns of $\boldsymbol{\mu}$ on top of one another. The covariance matrix for the error terms is given by

$$\boldsymbol{\Sigma}_\varepsilon(i,j) = \mathrm{cov}\left[\left\langle \varepsilon_1, u_i \right\rangle, \left\langle \varepsilon_1, u_j \right\rangle\right] = \mathrm{E}\left[\left\langle \varepsilon_1, u_i \right\rangle \left\langle \varepsilon_1, u_j \right\rangle\right], \quad 1 \le i, j \le r,$$

and the diagonal matrix containing the largest p eigenvalues of C is

$$\boldsymbol{\Gamma}(i,j) = \lambda_i \delta_{ij} \quad \text{for } 1 \le i, j \le p,$$

where δ_{ij} is Kronecker's delta.

With this notation in place, we consider the random variable

$$\Lambda_{pr} = N(1+\zeta)^{-1}(\hat{\boldsymbol{\mu}}_v - \hat{\boldsymbol{\mu}}_v^*)^{\mathrm{T}} \left(\boldsymbol{\Sigma}_\varepsilon \otimes \boldsymbol{\Gamma}^{-1} + \mathrm{E}\left[\boldsymbol{\Delta}_1 \otimes (\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta}_2\boldsymbol{\Gamma}^{-1})\right]\right)^{-1} (\hat{\boldsymbol{\mu}}_v - \hat{\boldsymbol{\mu}}_v^*),$$

where $\boldsymbol{\Delta}_1$, $\boldsymbol{\Delta}_2$ are defined in (20) and (21), respectively.

Assuming equal covariances, Theorem 5.4 implies that under H_0, $\Lambda_{pr} \xrightarrow{d} \chi^2(pr)$. An extension of the argument used in the proof of Theorem 3.1 yields that $\Lambda_{pr} \xrightarrow{P} \infty$, under H_A, as long as p and r are so large that $\boldsymbol{\mu} \ne \boldsymbol{\mu}^*$. That such a pair (p, r) exists, follows immediately from the fact that $\{v_i u_j\}$ form a basis in $L_2([0,1] \times [0,1])$.

A computable approximation to Λ_{pr} is

$$\hat{\Lambda}_{pr} = N\left(1 + \frac{N}{M}\right)^{-1} (\hat{\boldsymbol{\mu}}_v - \hat{\boldsymbol{\mu}}_v^*)^{\mathrm{T}} \left(\hat{\boldsymbol{\Sigma}}_\varepsilon \otimes \hat{\boldsymbol{\Gamma}}^{-1}\right)^{-1} (\hat{\boldsymbol{\mu}}_v - \hat{\boldsymbol{\mu}}_v^*), \tag{15}$$

where $\hat{\mathbf{\Sigma}}_\varepsilon$ is the pooled sample covariance matrix of the residuals and $\hat{\mathbf{\Gamma}} = \mathrm{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_p)$, with the $\hat{\lambda}_i$ being the eigenvalues of the empirical covariance operator of the pooled X_i and X_j^*.

We finally turn to the most complex case of different covariances for the explanatory functions. We now expand both the explanatory and response functions with respect to two arbitrary, potentially different, bases in $L^2[0, 1]$, $\{u_i\}$ and $\{w_j\}$, respectively:

$$\psi(t, s) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \mu_{ji} u_i(t) w_j(s), \qquad \psi^*(t, s) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \mu^*_{ji} u_i(t) w_j(s).$$

This leads to the relations $\mathbf{Y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\varepsilon}'$, $\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\mu}^* + \boldsymbol{\varepsilon}^{*\prime}$, with all terms analogously defined as in the equal variance case, but using the bases $\{u_i\}$ and $\{w_j\}$. Thus the least squares estimates are again $\hat{\boldsymbol{\mu}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$ and $\hat{\boldsymbol{\mu}}^* = (\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\mathbf{Y}^*$. ($\boldsymbol{\mu}, \boldsymbol{\mu}^*, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}}^*$ are now $p \times r$ matrices) Extending the argument developed in Section 3.1, we arrive at the test statistic

$$\hat{\Lambda}_{pr} = N\mathrm{vec}(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}^*)^{\mathrm{T}} \left(\hat{\mathbf{\Sigma}}_\varepsilon \otimes \hat{\mathbf{\Sigma}}_1^{-1} + \left(\frac{N}{M}\right) \hat{\mathbf{\Sigma}}_\varepsilon^* \otimes \hat{\mathbf{\Sigma}}_1^{-1*} \right)^{-1} \mathrm{vec}(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}^*), \tag{16}$$

where the residual covariance matrices $\mathbf{\Sigma}_\varepsilon$ and $\mathbf{\Sigma}_\varepsilon^*$ are computed for each sample separately. The estimate $\hat{\mathbf{\Sigma}}_1^{-1}$ is given by $N^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}$, and $\hat{\mathbf{\Sigma}}_1^{*-1}$ is equivalently defined.

The distribution of statistics (15) and (16) is approximated by the chi-square distribution with pr degrees of freedom. The values of p and r depend on the selected bases, examples of suitable bases are given in Section 4. Selection of p and r is discussed in Section 3.3.

If p and/or r are large, the normalized χ^2 distribution can be approximated by a normal distribution, as in Cardot et al. (2003) who studied a single scalar response model and tested $\psi = 0$. In our case, due to the complexity of the problem, the rigorous derivation of the normal convergence with $p = p_n$ depending on a sample size would be far more tedious, so it is not pursued. To perform a test, a finite p (and r) must be chosen no matter what approximation is used, and as illustrated in Section 4 large p (and r) do not necessarily lead to meaningful results.

3.3. Summary of the Testing Procedure

In order to apply the tests, we must first verify if a linear functional model approximates the dependence structure of the data reasonably well. This can be done using the techniques of Chiou & Müller (2007). The assumptions of independence and identical distribution can be verified using the test of Gabrys & Kokoszka (2007). The regressors and the responses must be centred, so that their sample means are zero.

Next, the values of p and r must be chosen. In applications in which the FPCs have a clear interpretation, these values can be chosen so that the action of the operators on specific subspaces spanned by the FPCs of interest is compared. In the absence of such an interpretation, several data driven approaches are available. When the covariances are approximately equal, typically p is chosen so large that $\sum_{k=1}^{p} \hat{\lambda}_k$ exceeds a required percentage of the variance of the X_i (defined as $(N+M)^{-1}(\sum_{i=1}^{N} \int X_i^2(t)\,\mathrm{d}t + \sum_{j=1}^{M} \int X_i^{*2}(t)\,\mathrm{d}t)$ for the centred functions). We choose r analogously for the response functions. When the covariances cannot be assumed equal then we propose, as one possibility, a pooling technique to choose p and r. Pooling the explanatory functions we have

$$(N+M)^{-1} \left(\sum_{i=1}^{N} X_i(s)X_i(t) + \sum_{j=1}^{M} X_j^*(s)X_j^*(t) \right) \stackrel{a.s.}{\to} \left(1 + \frac{1}{\zeta}\right)^{-1} c(s, t) + (1 + \zeta)c^*(s, t).$$

We propose taking the w_i to be the eigenfunctions of $(1 + 1/\zeta)^{-1}c(s, t) + (1 + \zeta)c^*(s, t)$ which is itself a covariance kernel. The u_i can be defined in an analogous manner using the response functions. Such a choice will allow smaller values of p (and r) to be taken so that any bias from neglected terms is minimal, but we can still expect reasonable power. The values p and r can be chosen as before, but now with respect to the pooled variance. All these steps can be implemented in the `R` package `fda`, and ready-made functions for the percentage of variance explained by FPCs are available. Other methods of choosing p (or r) are implemented in the MATLAB PACE package developed at the University of California at Davis.

It is often useful to compute the test for a wide range of values of p (and r) and check if a uniform pattern emerges. This approach is illustrated in Section 4.

Finally, we compute the test statistic $\hat{\Lambda}$, and reject H_0 if it exceeds the χ^2 density with DF degrees of freedom according to the following table:

Response	Covariances	$\hat{\Lambda}$	DF
Scalar	Equal	(9)	p
Scalar	Different	(11)	p
Functional	Equal	(15)	pr
Functional	Different	(16)	pr

The term "equal covariances" refers to assumptions (3), (4) in the scalar case, and (3), (12) in the functional case.

4. APPLICATION TO MEDFLY AND MAGNETOMETER DATA

In this section we illustrate the application of the test on two examples. The first example is motivated by the work presented in Carey et al. (2002), Chiou, Müller & Wang (2004), Müller & Stadtmüller (2005), Chiou & Müller (2007), and among others, and studies egg-laying curves of Mediterranean fruit flies (medflies). The second example is an application to the measurements of the magnetic field generated by near Earth space currents. Techniques of FDA are very promising in space physics, see Kokoszka et al. (2008) and Maslova et al. (in press).

Before turning to these examples, we note that simulations performed on generic models show that the procedures have empirical sizes are close to nominal, and the power increases with the "size" of the difference. The power is smaller if the explanatory functions do not have a common distribution, and/or are non-normal. Some tables and the `R` code are available at `http://www.stat.uchicago.edu/~mreimherr/research.html`.

Egg-laying curves of Mediterranean fruit flies: Müller & Stadtmüller (2005), Section 6, consider 534 egg-laying curves of medflies who lived at least 30 days. Each function is defined over an interval [0, 30], and its value on day $t \leq 30$ is the count of eggs laid by fly i on that day. The 534 flies can be classified into long-lived, that is, those who lived longer than 44 days, and short-lived, that is, those who died before the end of the 44th day after birth. In the sample, there are 256 short-lived, and 278 long-lived flies. This classification naturally defines two samples: *Sample 1:* the egg-laying curves $X_i(t)$, $0 < t \leq 30$, $i = 1, 2, \ldots, 256$ of the short-lived flies, and the corresponding total number of eggs Y_i. *Sample 2:* the egg-laying curves $X_j^*(t)$, $0 < t \leq 30$, $j = 1, 2, \ldots, 278$ of the long-lived flies, and the corresponding total number of eggs Y_j^*.

The average of the Y_j^* is obviously larger than that of the Y_i, but a question of interest is whether after adjusting for the means, the structure of the dependence of the Y_j^* on the curves $X_j^*(t)$ is different from the dependence of the Y_i on the curves $X_i(t)$. Additionally, we would like

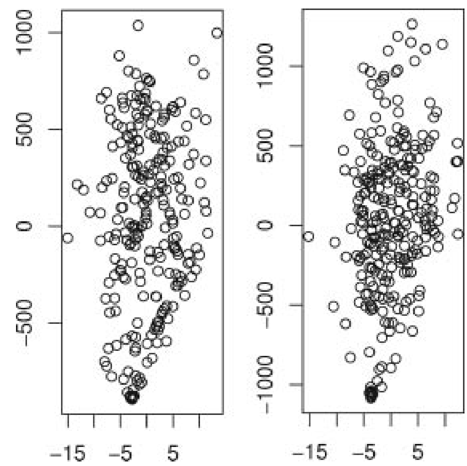

FIGURE 1: Diagnostic plots of responses versus scores of the second PC for medfly samples.

to know in which functional component the difference, if any, becomes significant. As the graphs and the analysis in Müller & Stadtmüller (2005) show, even distinguishing the two functional samples $X_i(t)$, $0 < t \le 30$, $i = 1, 2, \ldots, 256$ and $X_j^*(t)$, $0 < t \le 30$, $j = 1, 2, \ldots, 278$ is very difficult.

We thus consider two linear models:

$$Y_i - \bar{Y} = \int \psi(t)(X_i(t) - \bar{X}(t))\,\mathrm{d}t + \varepsilon_i, \quad i = 1, 2, \ldots, 256,$$

$$Y_j^* - \bar{Y}^* = \int \psi^*(t)X_j^*(t) - \bar{X}^*(t))\,\mathrm{d}t + \varepsilon_j^*, \quad j = 1, 2, \ldots, 278.$$

and test (without assuming equal variances) $H_0 : \psi = \psi^*$.

The above linear models describe the dependence structure of the data remarkably well. We applied the graphical test of Chiou & Müller (2007) in which the responses are graphed against the scores of the initial functional principal components. Figure 1 shows such graphs for the second principal component, the graphs for the other components also show nice elliptic shapes.

Table 1 shows the P-values for the five initial FPCs. The P-values for the remaining FPCs do not exceed half a percent. We see that there is a significant difference between the curves ψ and ψ^*, and that it occurs in the dependence on the second principal components of the egg-laying curves. Using only the first principal components ($p = p^* = 1$) is not enough to see the difference.

TABLE 1: Statistic (11) and the P-values for several values of $p = p^*$.

p	Λ_{pp^*}	P-Value
1	1.3323	0.2484
2	11.3411	0.0034
3	10.6097	0.0140
4	23.8950	0.0001
5	33.1144	0.0000

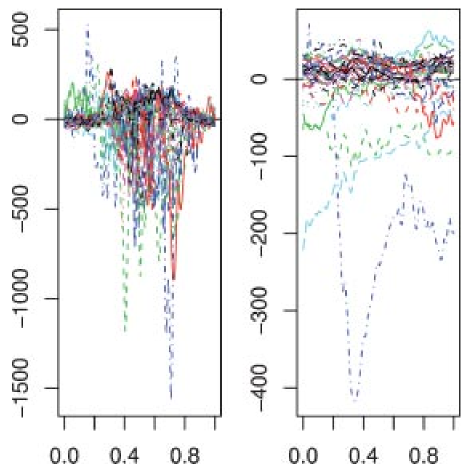

FIGURE 2: Observations for sample A: left panel CMO (*X*), right panel HON (*Y*). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Data from terrestrial magnetic observatories: We now apply our methodology to magnetometer data. A comprehensive case study is not our goal, but we would rather like to illustrate the steps outlined in Section 3.3 in a new, practically relevant setting.

A sample of 40 functional regressors and corresponding responses is shown in Figure 2. Each curve in Figure 2 shows one minute averages in a UT (Universal Time) day of the component of the magnetic field lying in the Earth's tangent plane pointing toward the magnetic North. We thus have 1,440 data points per curve, so a traditional multivariate analysis is not feasible. However, splitting the magnetogram data into days and treating the daily curves as functional observations is natural because of the daily rotation of the Earth. Moreover, for space physics researchers, it is the shape of these curves that conveys information, not the minute by minute values. These curves record magnetic fields generated by the magnetosphere–ionosphere (M–I) current system. The M–I currents are located in a region of space which is not sufficiently dense to allow operation of aircraft, but too dense for satellites, so in situ measurements are difficult to obtain. For this reason, and because of its inherent complexity, the M–I current system is only partially understood. An interested reader is referred to Kivelson & Russell (1997), especially Chapters 9, 10, 13, and 14. Here we utilize a simple and easy to explain setting, intended to illustrate our technique. The curves X_i reflect ionospheric magnetic activity in the polar region known as substorms, which is spectacularly manifested as northern lights (*aurora borealis*). The curves Y_i reflect magnetospheric activity in the magnetic equatorial region in the same UT day. We consider three samples: A, B, C. Each of them consists of about 40 pairs of curves. All measurements were recorded in 2001, the X_i at College (CMO), Alaska; the Y_i at Honolulu (HON), Hawaii. Sample A contains substorms which took place in January through March, B in April–June, C in July–September. Broader physical issues related to the dependence of the Y_i on the X_i in Kamide et al. (1998). Using a goodness-of-fit test of Chiou & Müller (2007), Kokoszka et al. (2008) verified that the fully functional linear model is a reasonable approximation to the dependence of the Y_i on the X_i. The identical distribution and lack of correlations of the observations were established in Gabrys & Kokoszka (2007) and Kokoszka et al. (2008). The data can be assumed to be approximately independent because the M-I system resets itself after each rotation of the Earth, and the effect of larger disturbances of solar origin decay within about 2 days.

Intuitively, we would expect rejections of the null for all three pairs: A–B, B–C, and A–C, as the position of the axis of the Earth relative to the Sun shifts with each season, and substorms are influenced by the solar wind. This is indeed the case for tests in cases B–C and A–C, for

TABLE 2: P-values for testing the equality of regression operators in samples A and B.

p/r	1	2	3	4	5	6	7	8	9	10
1	0.344	0.608	0.231	0.280	0.349	0.380	0.372	0.391	0.351	0.257
2	0.147	0.259	0.274	0.416	0.565	0.422	0.373	0.345	0.339	0.310
3	0.204	0.378	0.399	0.621	0.762	0.592	0.582	0.621	0.654	0.478
4	0.120	0.305	0.299	0.567	0.716	0.619	0.654	0.307	0.315	0.158
5	0.440	0.668	0.555	0.741	0.861	0.730	0.792	0.515	0.453	0.223
6	0.582	0.891	0.798	0.793	0.883	0.554	0.567	0.605	0.218	0.106
7	0.689	0.962	0.950	0.911	0.954	0.749	0.792	0.783	0.566	0.427
8	0.965	0.968	0.972	0.952	0.958	0.815	0.755	0.582	0.432	0.257
9	0.981	0.804	0.962	0.980	0.972	0.821	0.837	0.753	0.722	0.456
10	0.727	0.585	0.903	0.973	0.986	0.972	0.973	0.941	0.935	0.626
11	0.911	0.880	0.991	0.999	0.999	0.998	0.998	0.994	0.995	0.990
12	0.856	0.860	0.989	0.997	0.959	0.962	0.940	0.930	0.845	0.889
13	0.667	0.856	0.982	0.988	0.939	0.950	0.889	0.845	0.784	0.844
14	0.395	0.457	0.798	0.418	0.314	0.445	0.240	0.240	0.201	0.282
15	0.398	0.481	0.847	0.414	0.321	0.456	0.276	0.255	0.170	0.113

which the P-values are very small: for B–C the largest P-value is 0.034, and for A–C 0.007 (for $p \leq 15$, $r \leq 10$). The results for testing samples A and B presented in Table 2 indicate the acceptance of H_0. In retrospect, this conclusion is supported by the observation, well-known in the space-physics community, that M–I disturbances tend to be weaker in summer months. Our test thus shows that it is reasonable to assume that the effect of substorms on low-latitude currents is approximately the same in first and second quarter of 2001, but changes in the third quarter (due to weaker substorms).

5. ASYMPTOTIC THEORY

In this section we collect the statements and proofs of the theorems referred to in Section 3, and continue to use the notation introduced in that section. These results do not follow from the existing multivariate theory because the regression errors are not independent and include projections on the "left over" FPCs $v_{p+1}, v_{p+2}, \ldots, u_{r+1}, u_{r+2}, \ldots$, etc. Theorems 5.2 and 5.4 are of particular interest, as they state the exact asymptotic distribution of the LSE's in a multivariate regression obtained by projecting a functional regression.

Since the following simple lemma is used repeatedly in the proofs, it is stated first for ease of reference. It essentially follows from the orthonormality of the eigenfunctions of the covariance operator.

Lemma 5.1. *Suppose X is a random element of $L^2([0, 1])$ satisfying*

$$EX(t) = 0 \quad \text{and} \quad E\int_0^1 |X(t)|^2 \, \mathrm{d}t < \infty.$$

Let λ_i, v_i, $i \geq 1$, be, respectively, the eigenvalues and the eigenfunctions of its covariance operator. Then

$$E\left[\langle v_i, X\rangle \left\langle v_j, X\right\rangle\right] = \lambda_i \delta_{ij},$$

where δ_{ij} is Kronecker's delta.

Theorem 5.1. *Suppose Assumptions* 2.1, 2.2, *and* 2.4 *hold. Then,*

$$\hat{\boldsymbol{\mu}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} \overset{a.s.}{\to} \boldsymbol{\mu}, \text{ as } N \to \infty,$$

where $\boldsymbol{\mu}^{\mathrm{T}} = (\mu_1, \ldots, \mu_p)$ and $\overset{a.s.}{\to}$ refers to almost sure convergence.

Proof. To analyze the behaviour of $\hat{\boldsymbol{\mu}}$, let us start by considering

$$(\mathbf{X}^{\mathrm{T}}\mathbf{X})(i, j) = \sum_{k=1}^{N} \langle v_i, X_k\rangle \left\langle v_j, X_k\right\rangle.$$

Since the X_i are i.i.d., by the strong law of large numbers

$$\frac{1}{N}(\mathbf{X}^{\mathrm{T}}\mathbf{X})(i, j) \overset{a.s.}{\to} \mathrm{E}\left(\langle v_i, X_1\rangle \left\langle v_j, X_1\right\rangle\right), \text{ as } N \to \infty.$$

From Lemma 5.1 we have that $\mathrm{E}(\langle v_i, X_1\rangle\langle v_j, X_1\rangle) = \lambda_i\delta_{ij}$. Therefore $N^{-1}(\mathbf{X}^{\mathrm{T}}\mathbf{X})$ converges almost surely to a $p \times p$ diagonal matrix whose diagonal entries are the eigenvalues of C.

Turning to $\mathbf{X}^{\mathrm{T}}\mathbf{Y}$, using (1), observe

$$\mathbf{X}^{\mathrm{T}}\mathbf{Y}(i) = \sum_{j=1}^{N} \left\langle v_i, X_j\right\rangle Y_j = \sum_{j=1}^{N} \left\langle v_i, X_j\right\rangle \sum_{k=1}^{p} \mu_k \left\langle v_k, X_j\right\rangle + \sum_{j=1}^{N} \varepsilon'_j \left\langle v_i, X_j\right\rangle.$$

Applying again the strong law of large numbers and Lemma 5.1 again, we obtain, as $N \to \infty$,

$$N^{-1}\sum_{j=1}^{N} \left\langle v_i, X_j\right\rangle \sum_{k=1}^{p} \mu_k \left\langle v_k, X_j\right\rangle \overset{a.s.}{\to} \mathrm{E}\sum_{k=1}^{p} \mu_k \langle v_i, X_1\rangle \langle v_k, X_1\rangle = \mu_i\lambda_i\delta_{ij}.$$

Lastly, we will show that, as $N \to \infty$, $N^{-1}\sum_{j=1}^{N} \varepsilon'_j \left\langle v_i, X_j\right\rangle \overset{a.s.}{\to} 0$. Recalling the definition of ε'_i, (1), we have

$$N^{-1}\sum_{j=1}^{N} \varepsilon'_j \left\langle v_i, X_j\right\rangle = N^{-1}\sum_{j=1}^{N} \varepsilon_j \left\langle v_i, X_j\right\rangle + N^{-1}\sum_{j=1}^{N}\sum_{k=p+1}^{\infty} \mu_k \left\langle v_k, X_j\right\rangle \left\langle v_i, X_j\right\rangle.$$

Since $\{\varepsilon_i\}$ and $\{X_i\}$ are independent, by the strong law of large numbers and Assumption 2.2

$$N^{-1}\sum_{j=1}^{N} \varepsilon_j \left\langle v_i, X_j\right\rangle \overset{a.s.}{\to} 0.$$

Similarly, using Lemma 5.1 and noting that $i \le p$, we get

$$N^{-1}\sum_{j=1}^{N}\sum_{k=p+1}^{\infty}\mu_k\left\langle v_k, X_j\right\rangle\left\langle v_i, X_j\right\rangle \stackrel{a.s.}{\to} \mathrm{E}\sum_{k=p+1}^{\infty}\mu_k\left\langle v_k, X_1\right\rangle\left\langle v_i, X_1\right\rangle = 0.$$

∎

Theorem 5.2. *Suppose Assumptions* 2.1, 2.2, *and* 2.4 *hold. Then, as* $N \to \infty$,

$$\sqrt{N}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \stackrel{d}{\to} N(0, \boldsymbol{\Sigma}_p),$$

where $N(0, \boldsymbol{\Sigma}_p)$ *is a multivariate normal random vector with mean 0 and covariance matrix* $\boldsymbol{\Sigma}_p$ *defined by* (7) *and* (8).

Proof. By the definition of $\hat{\boldsymbol{\mu}}$ (2),

$$\sqrt{N}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) = \sqrt{N}((\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} - \boldsymbol{\mu}).$$

Defining $\boldsymbol{\varepsilon}'^{\mathrm{T}} = (\varepsilon'_1, \ldots, \varepsilon'_N)$, the above reduces to

$$\sqrt{N}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) = (N^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}N^{-1/2}\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}'.$$

By Lemma 5.1, $N^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}$ converges almost surely to a diagonal matrix whose diagonal elements are the first p eigenvalues of the covariance operator of the $\{X_i\}$. Therefore we need only focus on the behaviour of $N^{-1/2}\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}'$ and use Slutsky's Theorem to obtain the claimed limiting distribution. Considering the i^{th} coordinate of $N^{-1/2}\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}'$ we have

$$(N^{-1/2}\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}')(i) = N^{-1/2}\sum_{j=1}^{N}\left\langle v_i, X_j\right\rangle\varepsilon'_j, \quad 1 \le i \le p. \tag{17}$$

By Assumption 2.2 the above is a summation of i.i.d. random variables. Since each coordinate of $N^{-1/2}\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}'$ is given by such a sum, Assumption 2.2 implies that $\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}'$ can be expressed as a sum of i.i.d. random vectors. We can apply the multivariate central limit theorem to obtain the claimed multivariate normal limiting distribution if we can show that each entry of the covariance matrix is finite. Therefore we spend the rest of the proof deriving the form for $\boldsymbol{\Sigma}_p$ and showing that its entries are finite. Using the definition of ε'_i, we obtain

$$N^{-1/2}\sum_{j=1}^{N}\left\langle v_i, X_j\right\rangle\varepsilon'_j = N^{-1/2}\left(\sum_{j=1}^{N}\left\langle v_i, X_j\right\rangle\varepsilon_j + \sum_{j=1}^{N}\left\langle v_i, X_j\right\rangle\sum_{k=p+1}^{\infty}\mu_k\left\langle v_k, X_j\right\rangle\right).$$

Because the $\{X_j\}$ are independent, both sums (with respect to j) are sums of independent and identically distributed random variables. Furthermore, since $\{\varepsilon_j\}$ are independent of all other terms, we also have that the two sums above are uncorrelated. Therefore it follows that

$$\begin{aligned}&\operatorname{var}\left(N^{-1/2}\left(\sum_{j=1}^{N}\left\langle v_i, X_j\right\rangle\varepsilon_j + \sum_{j=1}^{N}\left\langle v_i, X_j\right\rangle\sum_{k=p+1}^{\infty}\mu_k\left\langle v_k, X_j\right\rangle\right)\right)\\ &\quad = \operatorname{var}\left(\left\langle v_i, X_1\right\rangle\varepsilon_1\right) + \operatorname{var}\left(\left\langle v_i, X_1\right\rangle\sum_{k=p+1}^{\infty}\mu_k\left\langle v_k, X_1\right\rangle\right).\end{aligned} \tag{18}$$

Considering the first term of (18), we have by the independence of X_1 and ε_1 and Lemma 5.1 that

$$\mathrm{var}(\langle v_i, X_1\rangle\, \varepsilon_1) = \lambda_i \sigma^2 < \infty.$$

Turning to the second term of (18), we have by Lemma 5.1

$$\mathrm{var}\left[\langle v_i, X_1\rangle \sum_{k=p+1}^{\infty} \mu_k \langle v_k, X_1\rangle\right] = \mathrm{E}\left[\langle v_i, X_1\rangle \sum_{k=p+1}^{\infty} \mu_k \langle v_k, X_1\rangle\right]^2.$$

Applying the Cauchy–Schwarz inequality it follows that

$$\mathrm{E}\left[\langle v_i, X_1\rangle \sum_{k=p+1}^{\infty} \mu_k \langle v_k, X_1\rangle\right]^2 \le \left(\mathrm{E}\,[\langle v_i, X_1\rangle]^4\, \mathrm{E}\left[\sum_{k=p+1}^{\infty} \mu_k \langle v_k, X_1\rangle\right]^4\right)^{1/2}.$$

As a consequence of Assumption 2.1, we obtain that

$$\mathrm{E}\,[\langle v_i, X_1\rangle]^4 < \infty.$$

Using the Cauchy–Schwarz inequality again we have

$$\mathrm{E}\left[\sum_{k=p+1}^{\infty} \mu_k \langle v_k, X_1\rangle\right]^4 \le \mathrm{E}\left[\sum_{k=p+1}^{\infty} \mu_k^2 \sum_{s=p+1}^{\infty} \langle v_s, X_1\rangle^2\right]^2.$$

Therefore we can infer that

$$\mathrm{E}\left[\langle v_i, X_1\rangle \sum_{k=p+1}^{\infty} \mu_k \langle v_k, X_1\rangle\right]^2 \le \left(\sum_{k=p+1}^{\infty} \mu_k^2\right)\left(\mathrm{E}\,[\langle v_i, X_1\rangle]^4\, \mathrm{E}\left[\sum_{s=p+1}^{\infty} \langle v_s, X_1\rangle^2\right]^2\right)^{1/2}.$$

Using Assumption 2.2 and Bessel's Inequality we have that

$$\sum_{k=p+1}^{\infty} \mu_k^2 \le \|\psi\|^2 < \infty.$$

Similarly, using Assumption 2.1 and Bessel's Inequality we have that

$$\mathrm{E}\left(\sum_{s=p+1}^{\infty} \langle v_s, X_1\rangle^2\right)^2 \le \mathrm{E}\|X_1\|^4 < \infty.$$

Combining the above with Assumption 2.1 we conclude

$$\mathrm{E}\left[\langle v_i, X_1\rangle \sum_{k=p+1}^{\infty} \mu_k \langle v_k, X_1\rangle\right]^2 < \infty. \tag{19}$$

and it follows that the diagonal elements of $\mathbf{\Sigma}_p$ are given

$$\mathbf{\Sigma}_p(i,i) = \lambda_i^{-1}\sigma^2 + \lambda_i^{-2}\,\mathrm{var}\left(\langle v_i, X_1\rangle \sum_{k=p+1}^{\infty} \mu_k \langle v_k, X_1\rangle\right), \quad i = 1, \ldots, p.$$

Next we examine the joint behaviour of the coordinates. Combining (19) with the Cauchy–Schwarz inequality we have

$$\mathrm{cov}\left[(\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}')(i), (\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}')(j)\right] < \infty, \quad i = 1, \ldots, p \text{ and } j = 1, \ldots, p.$$

Therefore to finish the proof we need only derive the form for the off diagonal terms of $\mathbf{\Sigma}_p$. Using (17), Assumption 2.2, and Lemma 5.1, it is easy to verify that for $i \neq j$

$$\begin{aligned}
\mathrm{cov}\left[(\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}')(i), (\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}')(j)\right] &= \mathrm{E}((\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}')(i)(\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}')(j)) \\
&= \mathrm{E}\left(\sum_{q=1}^{N}\langle v_i, X_q\rangle \sum_{k=p+1}^{\infty} \mu_k \langle v_k, X_q\rangle \sum_{s=1}^{N}\langle v_j, X_s\rangle \sum_{k=p+1}^{\infty} \mu_k \langle v_k, X_s\rangle\right) \\
&= \sum_{q=1}^{N}\mathrm{E}\left(\langle v_i, X_q\rangle \langle v_j, X_q\rangle \left(\sum_{k=p+1}^{\infty} \mu_k \langle v_k, X_q\rangle\right)^2\right) \\
&= N\mathrm{E}\left(\langle v_i, X_1\rangle \langle v_j, X_1\rangle \left(\sum_{k=p+1}^{\infty} \mu_k \langle v_k, X_1\rangle\right)^2\right).
\end{aligned}$$

Therefore it follows that the off diagonal terms of $\mathbf{\Sigma}_p$ are given by

$$\mathbf{\Sigma}_p(i,j) = \lambda_i^{-1}\lambda_j^{-1}\mathrm{E}\left(\langle v_i, X_1\rangle \langle v_j, X_1\rangle \left(\sum_{k=p+1}^{\infty} \mu_k \langle v_k, X_1\rangle\right)^2\right), \quad i \neq j,$$

which concludes the proof. ∎

Theorem 5.3. *If the assumptions of Theorem* 3.1 *are satisfied and p is so large that* $\boldsymbol{\mu} \neq \boldsymbol{\mu}^*$, *then* $\Lambda_p \xrightarrow{P} \infty$, *as* $N \to \infty$.

Proof. We start by expanding Λ_p as

$$\begin{aligned}
\Lambda_p &= N(1+\zeta)^{-1}(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}^*)^{\mathrm{T}}\mathbf{\Sigma}_p^{-1}(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}^*) \\
&= N(1+\zeta)^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu} - \hat{\boldsymbol{\mu}}^* + \boldsymbol{\mu}^*)^{\mathrm{T}}\mathbf{\Sigma}_p^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu} - \hat{\boldsymbol{\mu}}^* + \boldsymbol{\mu}^*) \\
&\quad + N(1+\zeta)^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}^*)^{\mathrm{T}}\mathbf{\Sigma}_p^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}^*) \\
&\quad + 2N(1+\zeta)^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}^*)^{\mathrm{T}}\mathbf{\Sigma}_p^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu} - \hat{\boldsymbol{\mu}}^* + \boldsymbol{\mu}^*).
\end{aligned}$$

Therefore we need only consider each term above. From Theorem 3.1 it follows that

$$N(1+\zeta)^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu} - \hat{\boldsymbol{\mu}}^* + \boldsymbol{\mu}^*)^{\mathrm{T}}\mathbf{\Sigma}_p^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu} - \hat{\boldsymbol{\mu}}^* + \boldsymbol{\mu}^*) = O_P(1).$$

From Theorem 5.2 it follows that

$$2N(1+\zeta)^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}^*)^{\mathrm{T}}\boldsymbol{\Sigma}_p^{-1}(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}^*+\boldsymbol{\mu}^*)=O_P(\sqrt{N}).$$

The last term we need to consider is

$$N(1+\zeta)^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}^*)^{\mathrm{T}}\boldsymbol{\Sigma}_p^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}^*).$$

Since $\boldsymbol{\Sigma}_p^{-1}$ is positive definite it follows that

$$(\boldsymbol{\mu}-\boldsymbol{\mu}^*)^{\mathrm{T}}\boldsymbol{\Sigma}_p^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}^*)>0,$$

and we have

$$N(1+\zeta)^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}^*)^{\mathrm{T}}\boldsymbol{\Sigma}_p^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}^*)\to\infty.$$

Furthermore when we divide the above by $\sqrt{N}$ we get

$$N^{1/2}(1+\zeta)^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}^*)^{\mathrm{T}}\boldsymbol{\Sigma}_p^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}^*)\to\infty.$$

Therefore $N(1+\zeta)^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}^*)^{\mathrm{T}}\boldsymbol{\Sigma}_p^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}^*)$ dominates all the other terms in the limit and the theorem follows. ∎

Theorem 5.4. *Suppose that Assumptions* 2.1, 2.3, 2.4, (3), (5), *and* (12) *hold. Then for each fixed* $p \geq 1$ *and* $r \geq 1$, *we have*

$$N^{1/2}(\hat{\boldsymbol{\mu}}_v-\boldsymbol{\mu}_v)\xrightarrow{d}N\left(0,\boldsymbol{\Sigma}_\varepsilon\otimes\boldsymbol{\Gamma}^{-1}+\mathrm{E}\left[\boldsymbol{\Delta}_1\otimes(\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta}_2\boldsymbol{\Gamma}^{-1})\right]\right)$$

where $\mathbf{I}_r$ *is the* $r\times r$ *identity matrix, and*

$$\boldsymbol{\Delta}_1(j,t)=\left(\sum_{s=p+1}^{\infty}\mu_{sj}\langle v_s,X_1\rangle\right)\left(\sum_{x=p+1}^{\infty}\mu_{xt}\langle v_x,X_1\rangle\right),\tag{20}$$

and

$$\boldsymbol{\Delta}_2(i,q)=\langle v_i,X_1\rangle\left\langle v_q,X_1\right\rangle.\tag{21}$$

Proof. The asymptotic normality follows from an application of the multivariate CLT. The derivation of the exact form of the asymptotic variance involves lengthy technical manipulations, and is omitted to conserve space. ∎

ACKNOWLEDGEMENTS

We thank Inga Maslova for compiling the magnetometer data used in Section 4. Lajos Horváth and Matthew Reimherr were partially supported by NSF grant DMS-0604670, Piotr Kokoszka by NSF grants DMS-0413653 and DMS-0804165.

BIBLIOGRAPHY

T. Ando, S. Imoto & S. Miyano (2004). Functional data analysis of the dynamics of gene regulatory networks. In "*Knowledge Exploration in Life Science Informatics*," Lecture Notes in Computer Science, Vol. 3303, Springer, pp. 69–83.

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

I. Berkes, R. Gabrys, L. Horváth & P. Kokoszka (in press). Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society (B)*.

P. Besse, H. Cardot & D. Stephenson (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics*, 27, 673–687.

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

T. Cai & P. Hall (2006). Prediction in functional linear regression. *The Annals of Statistics*, 34, 2159–2179.

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

J. R. Carey, P. Liedo, L. Harshman, H. G. Müller, L. Partridge & J. L. Wang (2002). Life history response of Mediterranean fruit flies to dietary restrictions. *Aging Cell*, 1, 140–148.

J-M. Chiou & H-G. Müller (2007). Diagnostics for functional regression via residual processes. *Computational Statistics and Data Analysis*, 15, 4849–4863.

J-M. Chiou, H-G. Müller & J-L. Wang (2004). Functional response models. *Statistica Sinica*, 14, 675–693.

A. Cuevas, M. Febrero & R. Fraiman (2002). Linear functional regression: the case of fixed design and functional response. *The Canadian Journal of Statistics*, 30, 285–300.

J. Cupidon, D. S. Gilliam, R. Eubank & F. Ruymgaart (2007). The delta method for analytic functions of random operators with application to functional data. *Bernoulli*, 13, 1179–1194.

J. Damon & S. Guillas (2002). The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics*, 13,759–774.

R. L. Eubank & T. Hsing (2008). Canonical correlation for stochastic processes. *Stochastic Processes and their Applications*, 118, 1634–1661.

M. Febrero, P. Galeano & W. González-Manteiga (2008). Outlier detection in functional data by depth measures with application to identify abnormal NO_x levels. *Environmetrics*, 19, 331–345.

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

F. Ferraty & P. Vieu (2006). "*Nonparametric Functional Data Analysis: Theory and Practice*," Springer, New York.

F. Ferraty, P. Vieu & S. Viguier-Pla (2007). Factor-based comparison of groups of curves. *Computational Statistics & Data Analysis*, 51, 4903–4910.

R. Gabrys & P. Kokoszka (2007). Portmanteau test of independence for functional observations. *Journal of the American Statistical Association*, 102, 1338–1348.

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

G. He, H-G. Müller & J-L. Wang (2003). Functional canonical analysis for square integrable stochastic processes. *Journal of Multivariate Analysis*, 85(1), 54–77.

D. Hlubinka & L. Prchal (2007). Changes in atmospheric radiation from the statistical point of view. *Computational Statistics & Data Analysis*, 51, 4926–4941.

L. Horváth, M. Hušková & P. Kokoszka (in press). Testing the stability of the functional autoregressive process. *Journal of Multivariate Analysis*.

Y. Kamide, W. Baumjohann, W. D. Gonzalez, I. A. Daglis, M. Grande, J. A. Joselyn, R. L. McPherron, J. L. Phillips, E. G. D. Reeves, G. Rostoker, A. S. Sharma, H. J. Singer, B. T. Tsurutani & V. M. Vasyliunas (1998). Current understanding of magnetic storms: Storm–substorm relationships. *Journal of Geophysical Research*, 103, 17705–17728.

M. G. Kivelson & C. T. Russell (eds) (1997). "*Introduction to Space Physics*," Cambridge University Press, Cambridge, UK.

P. Kokoszka, I. Maslova, J. Sojka & L. Zhu (2008). Testing for lack of dependence in functional linear model. *Canadian Journal of Statistics*, 36, 207–222.

A. Laukaitis & A. Račkauskas (2002). Functional data analysis of payment systems. *Nonlinear Analysis: Modeling and Control*, 7, 53–68.

A. Laukaitis & A. Račkauskas (2005). Functional data analysis for clients segmentation tasks. *European Journal of Operational Research*, 163, 210–216.

Y. Li & T. Hsing (2007). On rates of convergence in functional linear regression. *Journal of Multivariate Analysis*, 98, 1782–1804.

N. Malfait & J. O. Ramsay (2003). The historical functional model. *Canadian Journal of Statistics*, 31, 115–128.

I. Maslova, P. Kokoszka, J. Sojka & L. Zhu (2009). Removal of nonconstant daily variation by means of wavelet and functional data analysis. *Journal of Geophysical Research*, 114, A03202.

J. S. Morris, C. Arroyo, B. A. Coull, L. M. Ryan, R. Herric & S. Gortmaker (2006). Using wavelet based functional mixed models to characterize population heterogeneity in accelerometer profiles: a case study. *Journal of the American Statistical Association*, 101, 1352–1364.

H-G. Müller & U. Stadtmüller (2005). Generalized functional linear models. *The Annals of Statistics*, 33, 774–805.

H-G. Müller, U. Stadtmüller & F. Yao (2006). Functional variance processes. *Journal of the American Statistical Association*, 101, 1007–1018.

D. Nerini & B. Ghattas (2007). Classifying densities using functional regression trees: Applications in oceanology. *Computational Statistics & Data Analysis*, 51, 4984–4993.

J. O. Ramsay & B. W. Silverman (2005). "*Functional Data Analysis*," Springer, New York.

F. Yao, H-G. Müller & J-L. Wang (2005). Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, 33, 2873–2903.

Received 2 September 2008
Accepted 25 May 2009