Supplementary materials for this article are available online. Please click the JASA link at http://pubs.amstat.org.

Second-Order Comparison of Gaussian Random Functions and the Geometry of DNA Minicircles

Victor M. PANARETOS, David KRAUS, and John H. MADDOCKS

Given two samples of continuous zero-mean iid Gaussian processes on [0, 1], we consider the problem of testing whether they share the same covariance structure. Our study is motivated by the problem of determining whether the mechanical properties of short strands of DNA are significantly affected by their base-pair sequence; though expected to be true, had so far not been observed in three-dimensional electron microscopy data. The testing problem is seen to involve aspects of ill-posed inverse problems and a test based on a Karhunen–Loève approximation of the Hilbert–Schmidt distance of the empirical covariance operators is proposed and investigated. When applied to a dataset of DNA minicircles obtained through the electron microscope, our test seems to suggest potential sequence effects on DNA shape. Supplemental material available online.

KEY WORDS: Covariance operator; DNA shape; Functional data analysis; Hilbert–Schmidt norm; Karhunen–Loève expansion; Regularization; Spectral truncation; Two-sample testing.

1. INTRODUCTION

The understanding of the mechanical properties of the DNA molecule constitutes a fundamental biophysical task, as important biological processes, such as the packing of DNA in the nucleus or the regulation of genes, can be affected by properties such as stiffness and shape (Vilar and Leibler 2003; Tolstorukov et al. 2005). The study of these properties can focus on different scales, and accordingly involves a variety of mathematical models and techniques. At a coarse-grained level, the behavior of short (of the order of 150 base pairs) strands of DNA is likened to that of a continuous elastic rod. By means of a reaction called *cyclization*, two ends of this elastic rod bend and twist and bind together to form a loop called a *DNA minicircle*. These three-dimensional cyclic structures are an excellent specimen for examining the elastic properties of DNA since a minicircle is in a naturally stressed state without the application of external forces. Furthermore, the short length of these strands will amplify the dependence of the mechanistic behavior on intrinsic factors such as the specific base pair sequence.

Such sequence-dependent shape characteristics are of special interest as they potentially reveal a dual purpose of the DNA base-pair sequence: in addition to holding the genetic code, the sequence may influence the geometric properties of the molecule. While in principle certain particular subsequences are expected to have a strong effect on the mechanical properties of DNA, empirical detection of this effect on stereological data acquired through the electron microscope has been elusive (Hagerman 1988; Amzallag et al. 2006). A specific example is that of a subsequence called the *TATA box*, which promotes gene transcription. It is thought that the mechanical properties of this subseqence are intimately related with its function, and that its presence in a DNA minicircle will enhance its flexibility. Nevertheless, exploratory comparisons between reconstructed minicircles from microscope images containing TATA boxes with reconstructed minicircles with no TATA box did not reveal any effects due to the presence of the sequence (Amzallag et al. 2006).

Motivated by the need of two-sample comparison of loops, as exemplified in DNA minicircle experiments, this article considers the problem of second-order comparison of two samples of random functions, within a functional data analysis framework. In particular, given realisations of n_1 and n_2 independent copies of two continuous zero mean Gaussian processes X and Y on a compact set, we consider the problem of testing the hypothesis $H_0 : \mathscr{R}_X = \mathscr{R}_Y$ against the alternative $H_A : \mathscr{R}_X \neq \mathscr{R}_Y$, where the covariance operators $\mathscr{R}_X, \mathscr{R}_Y$ are not necessarily stationary. The literature on hypothesis testing for functional data is mostly concentrated on tests pertaining to the mean function (Fan and Lin 1998), as encountered, for instance, in functional linear models (Cardot et al. 2003; Cuevas, Febrero, and Fraiman 2004; Shen and Faraway 2004) or functional change detection (Berkes et al. 2009). Hall and Van Keilegom (2007) studied the important issue of the effect that the data smoothing step may have on two-sample testing. Second-order tests for functional data analysis pertaining to serial correlation were also investigated (e.g., Gabrys and Kokoszka 2007; Horváth, Hušková, and Kokoszka 2010). Although the seeds of functional two-sample covariance tests can be found in Grenander (1981), the problem of second-order comparison of functional data has—interestingly—so far received relatively little attention. A related recent article by Benko, Härdle, and Kneip (2009) proposed two-sample bootstrap tests for specific aspects of the spectrum of functional data, such as the equality of a subset of the eigenfunctions, or—assuming that the eigenfunctions are shared—equality of a subset of eigenvalues.

In this article, we consider the difficulties associated with this testing problem, and it is seen that the extension of finite-dimensional procedures can lead to complications, as the infinite-dimensional version of the problem constitutes an ill-posed inverse problem. As an alternative solution, we propose a test based on the approximation of the Hilbert–Schmidt distance of the empirical covariance operators of the two samples of functions based on the Karhunen–Loève expansion. The asymptotic distribution of the test statistic is determined and its

Victor M. Panaretos is Assistant Professor (E-mail: *victor.panaretos@epfl.ch*), David Kraus is Postdoctoral Researcher, and John H. Maddocks is Professor, Section de Mathématiques, Ecole Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland. The authors thank the editor, associate editor, and two referees for providing detailed and constructive comments, and for their fruitful suggestions. The last author wishes to acknowledge support from FN grant 205320-112178.

© 2010 American Statistical Association
Journal of the American Statistical Association
June 2010, Vol. 105, No. 490, Theory and Methods
DOI: 10.1198/jasa.2010.tm09239

performance is investigated computationally. The application of our methodology to an electron microscope dataset of two groups of minicirles characterized by the presence or absence of a TATA box suggests the potential existence of significant differences in the two groups, which eluded previous analyses as these focused on the mean (the shape of the minicircle), whereas we detect the differences in the covariance structure (the flexibility/stiffness).

The article is organized as follows. The next section describes the three-dimensional functional dataset of DNA minicircles, from acquisition to registration, and includes a preliminary exploratory analysis. The first part of the third section then provides some functional data analysis background. Section 3.2 introduces our spectral test statistic and develops its asymptotic distribution, while Section 3.3 treats the problem of tuning the amount of regularization. In Section 4 the power and level of the test under various scenarios is investigated by means of simulation. Section 5 presents the results of a two-sample analysis of the DNA minicircles through the spectral test statistics, and the article concludes with a short discussion.

2. DNA MINICIRCLE DATA

The dataset of interest was reconstructed from electron micrographs imaged by Jan Bednar at the Laboratory of Ultrastructural Analysis of the University of Lausanne, Switzerland. A total of 99 DNA minicircles of 158 base-pair length were vitrified and imaged under two different angles, yielding two projected images of the same specimen, which were then used to reconstruct three-dimensional structural models (Jacob et al. 2006). The reconstructed data consist of 99 closed curves (DNA minicircles) in $\mathbb{R}^3$ of two types: both types have identical base pair sequences, except for a 14 base-pair window where 65 curves contain the *TATA sequence*, while the remaining 34 contain a different sequence, called a *CAP sequence*. Biophysical considerations suggest that the presence of a TATA box will have a significant effect on the geometry of the minicircle, and the goal is to compare these two groups to probe for such an effect.

In its reconstructed form, each curve is represented as a combination of periodic B-spline basis functions taking values in $\mathbb{R}^3$. To perform a functional data analysis of the minicircles it is required to register the data. Each curve has thus been centered and scaled, so that the center of mass is at zero and the length of the curve is one. The nature of the experimental setup in single-particle electron microscopy requires that the minicircles be imbedded unconstrained in the aqueous solution, so that the reconstructed curves are not aligned: the original (x, y, z)-coordinates for the different curves are not directly comparable as each curve was subjected to a random unobservable orthogonal transformation. It is thus necessary to align the curves. Landmark alignment methods (e.g., Gasser and Kneip 1995) are not applicable as the exact DNA sequence is not detectable from an electron micrograph. On the other hand, more flexible methods such as warping (e.g., Gervini and Gasser 2004; Tang and Müller 2008) are inappropriate since nonrigid alignment will alter the second-order properties that are of principal interest. As an alternative, we rigidly align curves by their intrinsic characteristics: each curve was individually aligned using the coordinate system induced by its *moments of inertia tensor* (e.g., Arnold 1989), which is described as follows. Consider an object in three dimensions described by a mass distribution μ—for example, for a DNA minicircle, μ will be the uniform measure supported on the curve. Suppose that the object is rotating around an axis, which without loss of generality, is given by $\text{span}(\mathbf{u}) := \{\lambda\mathbf{u} : \lambda \in \mathbb{R}\}$ for some $\mathbf{u} \in \mathbb{S}^2$. Let $r(\mathbf{u}, \mathbf{x}) := \|(\mathbf{I} - \mathbf{u}\mathbf{u}^\top)\mathbf{x}\|$ denote the distance of a point $\mathbf{x}$ from the subspace $\text{span}(\mathbf{u})$. The moment of inertia of the object around the axis $\mathbf{u}$ is given by

$$\mathscr{J}(\mathbf{u}) := \int_{\mathbb{R}^3} r^2(\mathbf{u}, \mathbf{x})\mu(d\mathbf{x}) = \int_{\mathbb{R}^3} \|(\mathbf{I} - \mathbf{u}\mathbf{u}^\top)\mathbf{x}\|^2 \mu(d\mathbf{x}).$$

Given a coordinate system defined by an orthonormal basis, say the canonical basis $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, we can use only these basis vectors to compactly represent the moment of inertia with respect to *any other axis passing by the origin*. Define the inertia matrix as

$$\mathbf{J} := \left\{ \int_{\mathbb{R}^3} \mathbf{x}^\top (\mathbf{e}_i^\top \mathbf{e}_j \mathbf{I} - \mathbf{e}_i \mathbf{e}_j^\top)\mathbf{x}\mu(d\mathbf{x}) \right\}_{i,j}.$$

Notice that the diagonal elements of the above matrix are the moments of inertia with respect to the axes of the coordinate system. The moment of inertia around any unit vector $\mathbf{u}$ can now be recovered as $\mathscr{J}(\mathbf{u}) = \mathbf{u}^\top \mathbf{J}\mathbf{u}$. Since the tensor is symmetric, it possesses real eigenvalues and orthonormal eigenvectors forming a basis, which admit the following interpretation: the first eigenvector, say $\mathbf{w}_1$, determines the axis (first principal axis of inertia, PAI1) around which the curve is most difficult to rotate, in the sense that the corresponding angular moment is maximized: $\mathbf{w}_1^\top \mathbf{J}\mathbf{w}_1 \geq \mathbf{u}^\top \mathbf{J}\mathbf{u}$ for any other $\mathbf{u} \in \mathbb{S}^2$. The projection on the plane orthogonal to $\mathbf{w}_1$ is "most spread" in this sense. The second eigenvector determines the axis within the first principal plane around which the projected curve is most difficult to rotate. That is, within the first principal plane, the projection on the line orthogonal to PAI2 is most spread. Hence, PAI3 carries the most spatial information, whereas PAI1 contains the smallest amount of information. Then, for each curve, the starting point was determined as the point where the projection on the first principal plane intersects the horizontal (PAI2) positive semi-axis and the orientation was chosen as counterclockwise in this plane (i.e., at the beginning the PAI3 coordinate increases from zero and PAI2 is positive).

The projections onto the principal axes of the minicircle curves are depicted in Figures 1 and 2. The data appear to be well aligned, and seem to be elliptical on average within the principal plane of inertia. Deviations from this principal plane, on the other hand, seem to be lacking systematic structure. The effectiveness of this alignment method is of crucial importance, as we will not be able to otherwise proceed with the testing problem (procrustean alignment of the curves will require us to optimize a sum of squares criterion with respect to 99 orthogonal transformations).

A visual inspection reveals five curves (plotted with dashed lines) that appear to be "standing out" of the rest—outliers in a broad sense. Judging whether or not a curve (an infinite dimensional object) is an outlier or not can be far trickier than in the vector case. In particular, it can be that there are further "outlying curves" that do not appear to stick out of the crowd, but are nevertheless intrinsically different from the rest. For this reason, we pursue a robust analysis for the mean curve

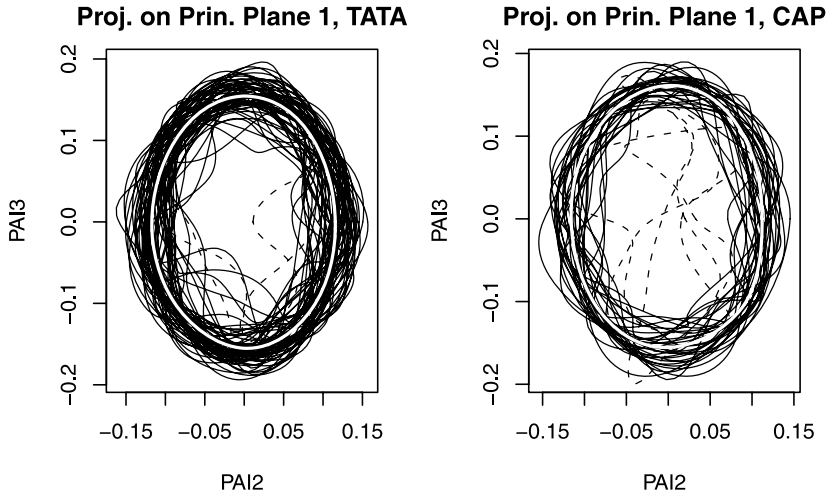

Figure 1. Projection of DNA curves on the first principal plane. Five removed outlying observations plotted in dashed lines. The mean curves (in white) are computed without outlying observations.

using a *functional median* introduced in Gervini (2008). The idea is simple: an iterative robust procedure will assign weights to each curve, and we can then detect outlying curves by looking at small weights. The method confirms our visual intuition, and reveals no further outliers. The outlying observations are removed, and after this preprocessing stage we are left with 94 aligned smooth curves.

3. METHODS

3.1 Background: FDA and Karhunen–Loève Expansions

We adopt a functional data analysis perspective (Ramsay and Silverman 2005; Ferraty and Vieu 2006) and model each curve as the realization of a stochastic process indexed by the closed interval $[0, 1]$ and taking values in $\mathbb{R}^3$ (but every-

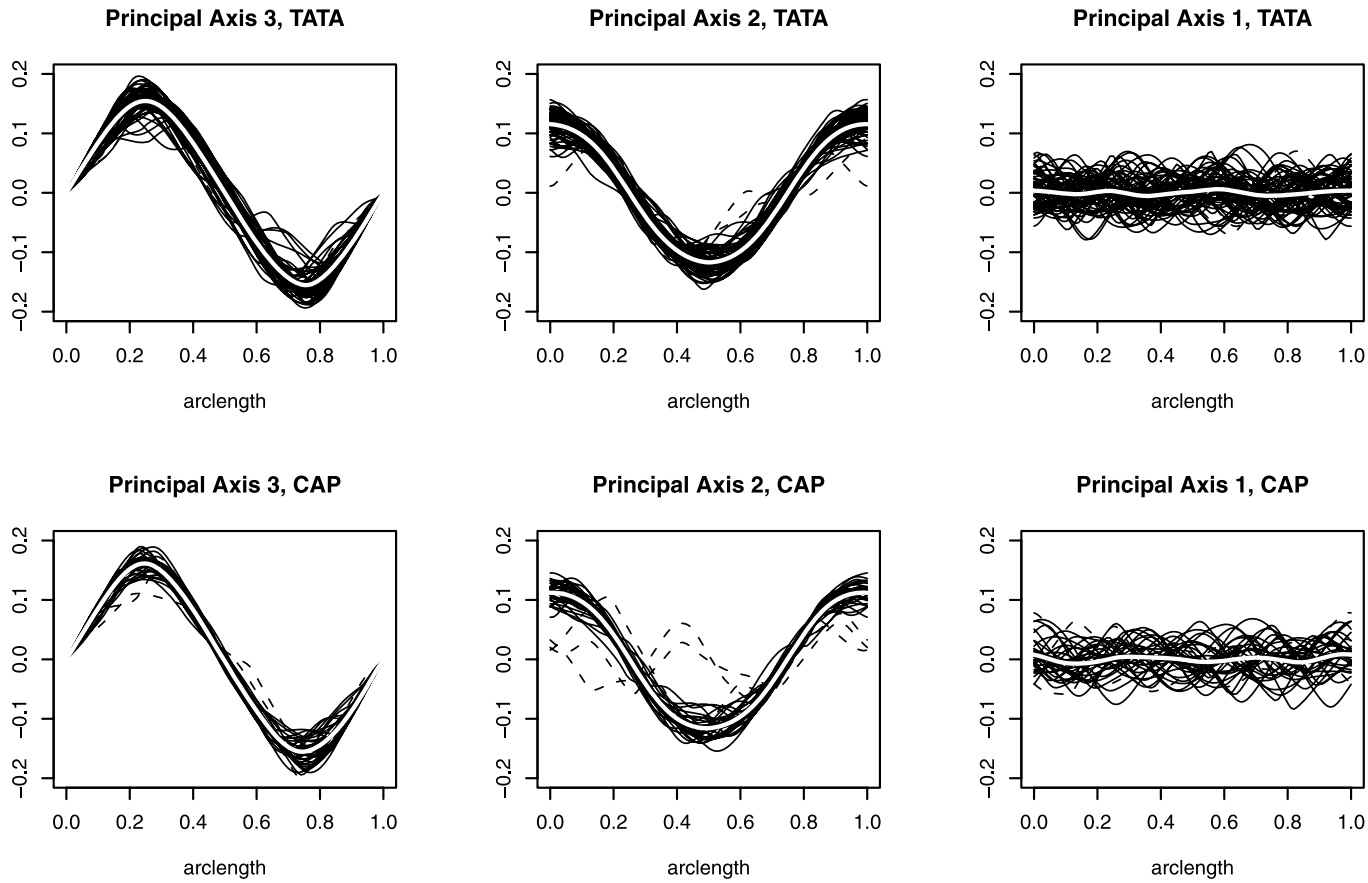

Figure 2. Coordinates of DNA curves on the principal axes of inertia. Five removed outlying observations plotted with dashed lines. Mean curves (in white) are computed without outlying observations.

thing readily extends to the case of $\mathbb{R}^d$). In particular, we assume that we have two independent collections $\mathbf{X}_1, \ldots, \mathbf{X}_{n_1}$ and $\mathbf{Y}_1, \ldots, \mathbf{Y}_{n_2}$, of iid Gaussian processes on $[0, 1]$, considered as random elements of the Hilbert space $\mathcal{L}^2[0, 1]$ of coordinate-wise square-integrable $\mathbb{R}^3$-valued functions with the inner product $\langle \mathbf{f}, \mathbf{g} \rangle = \int_0^1 \mathbf{f}(t)^\top \mathbf{g}(t)\, dt$. Here, $\mathbf{f}(t)^\top$ represents the transpose of the vector-valued function $\mathbf{f}(t) \in \mathbb{R}^3$. Assuming, without loss of generality, that the mean functions are zero, the processes are characterized by their respective covariance kernels $\mathbf{R}_X(s, t) = \operatorname{cov}(\mathbf{X}_i(s), \mathbf{X}_i(t)) = \mathbb{E}\{\mathbf{X}_i(s)\mathbf{X}_i^\top(t)\}$, and $\mathbf{R}_Y(s, t)$, respectively. Associated with the covariance kernel is the covariance operator $\mathscr{R}_X : \mathcal{L}^2[0, 1] \to \mathcal{L}^2[0, 1]$ defined as $\mathscr{R}_X(\mathbf{f})(t) = \operatorname{cov}(\langle \mathbf{X}_i, \mathbf{f} \rangle, \mathbf{X}_i(t)) = \int_0^1 \mathbf{R}_X(t, s)\mathbf{f}(s)\, ds$. Throughout the article, we will be assuming $\mathbf{R}_X$ to be continuous, so that $\mathscr{R}_X$ is bounded and the X process is continuous (resp. the Y process).

Inference for iid collections of infinite-dimensional random elements is often carried out in practice by an "optimal" reduction to a finite-dimensional setting, using finitely many appropriately chosen contrasts in a *functional principal component analysis* (e.g., Ramsay and Silverman 2002, 2005; Hall and Hosseini-Nasab 2006; also see Dauxois, Pousse, and Romain 1982 for distributional asymptotics). This procedure exploits the Karhunen–Loève theorem (e.g., Adler 1990), which allows for a representation of the process by a stochastic Fourier series with respect to the orthonormal eigenfunctions $\{\boldsymbol{\varphi}_X^{(j)}\}_{j=1}^\infty$ of the operator $\mathscr{R}_X$,

$$\mathbf{X}_i(t) = \sum_{j=1}^{\infty} \sqrt{\lambda_X^{(j)}}\, \xi_{ij} \boldsymbol{\varphi}_X^{(j)}(t),$$

where $\{\lambda_X^{(j)}\}_{j=1}^\infty$ is the nonincreasing sequence of corresponding eigenvalues and $\{\xi_{ij}\}$ is an iid array of standard Gaussian random variables. Convergence of the series is in mean square, uniformly in $t \in [0, 1]$.

Thus, in a practical setting, the empirical covariance kernel may be used to "optimally" reduce infinite-dimensional inferential problems to multivariate ones. Letting $\widehat{\mathbf{R}}_X$ stand for the empirical covariance kernel, $\widehat{\mathbf{R}}_X(s, t) := \frac{1}{n_1} \sum_{i=1}^{n_1} (\mathbf{X}_i(s) - \overline{\mathbf{X}}(s))(\mathbf{X}_i(t) - \overline{\mathbf{X}}(t))^\top$, we denote its eigenvalues (or *principal scores*) by $\{\widehat{\lambda}_X^{k,n_1}\}_{k=1}^{n_1}$ and its eigenfunctions (or *principal components*) by $\{\widehat{\boldsymbol{\varphi}}_X^{k,n_1}\}_{k=1}^{n_1}$. The finite-dimensional reduction is then achieved by retaining a finite number of *principal components* $\{\langle \mathbf{X}_i - \overline{\mathbf{X}}, \widehat{\boldsymbol{\varphi}}_X^{k,n_1} \rangle\}_{k=1}^K$ in lieu of each $\mathbf{X}_i$. These are zero mean and uncorrelated random variables, with corresponding sample variances $\widehat{\lambda}_X^{k,n_1}$. Similarly, for the second sample, the analogous quantities are $\mathbf{R}_Y, \mathscr{R}_Y, \lambda_Y^{(j)}, \boldsymbol{\varphi}_Y^{(j)}$ (and their empirical "hat" counterparts). The dimension reduction afforded by the Karhunen–Loève expansion is the tool we will next employ to construct our test.

3.2 Second-Order Comparison of Gaussian Processes

Let $\{\mathbf{X}_i\}_{i=1}^{n_1}$ and $\{\mathbf{Y}_i\}_{i=1}^{n_2}$ constitute two iid random samples of Gaussian processes indexed by the interval $[0, 1]$ and taking values in $\mathbb{R}^3$ (or indeed $\mathbb{R}^d$). As mentioned in the previous section, these are regarded as random elements of the Hilbert space $\mathcal{L}^2[0, 1]$ of square-integrable $\mathbb{R}^3$-valued functions (where integration is to be understood coordinate-wise). Assuming that the covariance operators $\mathscr{R}_X$ and $\mathscr{R}_Y$ associated with the processes are continuous, we wish to test the hypothesis pair

$$\begin{cases} H_0 : \mathscr{R}_X = \mathscr{R}_Y, \\ H_A : \mathscr{R}_X \neq \mathscr{R}_Y. \end{cases} \tag{1}$$

A natural first approach to developing a test for the hypothesis pair in Equation (1) is to attempt to extend tests developed for the finite-dimensional version of the problem, which was extensively studied. The majority of test statistics for the equality of covariance matrices of Gaussian vectors are based on the determinant, trace, or maximum/minimum eigenvalues of matrices such as: $\mathbf{S}_1\mathbf{S}_2\mathbf{S}^{-1}$, $\mathbf{S}_1\mathbf{S}_2^{-1}$, $\mathbf{S}_2(\mathbf{S}_1 + \mathbf{S}_2)^{-1}$ (Roy 1953; Pillai 1955; Kiefer and Schwartz 1965; Giri 1968); here, $\mathbf{S}_1$ and $\mathbf{S}_2$ are the empirical covariance matrices corresponding to each sample, and $\mathbf{S}$ is the pooled empirical covariance matrix. Evidently, such tests cannot immediately be carried over to the case of Gaussian processes: inversion of an empirical covariance operator will be required, which transforms the construction of the test statistic into an ill-posed inverse problem.

The operator $\widehat{\mathscr{R}}_X^{n_1}$ (resp. $\widehat{\mathscr{R}}_Y^{n_2}$) will be of rank at most n_1 (resp. n_2) as its image is the subspace spanned by $\{\mathbf{X}_i\}_{i=1}^{n_1}$ (resp. $\{\mathbf{Y}_i\}_{i=1}^{n_2}$). Therefore, we cannot talk of its inverse, except if we restrict the operator on $\operatorname{span}\{\mathbf{X}_i\}_{i=1}^{n_1}$ (resp. $\operatorname{span}\{\mathbf{Y}_i\}_{i=1}^{n_1}$), but the two spans will *not* coincide in general and the two empirical operators will *not* be diagonalized by the same basis. Furthermore, since the processes are assumed to be second order, the operators $\mathscr{R}_X$ and $\mathscr{R}_Y$ are necessarily bounded (in fact compact), and it must be the case that $\lambda_X^{(k)}, \lambda_Y^{(k)} \overset{k\to\infty}{\longrightarrow} 0$, the rate of convergence depending on the degree of smoothness of the Gaussian processes (the smoother the process, the faster the rate). Thus, for any finite n_1 and n_2, however large, a test statistic employing an "inverse" of $\widehat{\mathscr{R}}_X$ composed with $\widehat{\mathscr{R}}_Y$ will be unstable to perturbations of the Y-data.

In the infinite-dimesional case, we propose the use of a test statistic based on the norm of the difference of the two empirical covariance operators. Recall that for trace-class operators, one may define the *Hilbert–Schmidt norm*. Consider an integral operator $\mathscr{R} : \mathbf{f} \mapsto \int_0^1 \mathbf{R}(\cdot, s)\mathbf{f}(s)\, ds$ such that $\int_0^1 \int_0^1 \operatorname{trace}\{\mathbf{R}(s, t)^\top \mathbf{R}(s, t)\}\, ds\, dt < \infty$. The Hilbert–Schmidt norm of the operator $\mathscr{R}$ is defined as

$$\|\mathscr{R}\|_{\mathrm{HS}} := \sqrt{\int_0^1 \int_0^1 \operatorname{trace}\{\mathbf{R}(s, t)^\top \mathbf{R}(s, t)\}\, ds\, dt}.$$

Assuming that the covariance operators in question are Hilbert–Schmidt, a test may be based on the squared Hilbert–Schmidt distance $\|\widehat{\mathscr{R}}_X^N - \widehat{\mathscr{R}}_Y^N\|_{\mathrm{HS}}^2$ of their empirical counterparts. Of course, the sampling distribution of this latter quantity will depend on the unknown covariance operators even asymptotically. To be able to "normalize" the test statistic, we employ a very useful property of the Hilbert–Schmidt norm: for any orthonormal system $\{\mathbf{e}_i\}_{i=1}^\infty$ of $\mathcal{L}^2[0, 1]$, we have

$$\|\mathscr{R}\|_{\mathrm{HS}}^2 = \sum_{i=1}^{\infty} \|\mathscr{R}\mathbf{e}_i\|_{\mathcal{L}^2}^2. \tag{2}$$

Therefore, we may use a basis to obtain a countable expression for $\|\widehat{\mathscr{R}}_X^N - \widehat{\mathscr{R}}_Y^N\|_{\mathrm{HS}}^2$. In practice, one will need to truncate a series such as the above to obtain an "optimal" finite-dimensional

reduction, that is, the choice of contrasts $\{\mathbf{e}_i\}$ should be such that the truncated version of Equation (2) retains the bulk of the norm.

For each of the two empirical operators, the optimal contrasts will coincide with their eigenfunctions, as dictated by the Karhunen–Loève expansion, but to use the relation in Equation (2) we need to use a common basis. As a compromise, we thus choose the eigenfunctions $\{\widehat{\boldsymbol{\varphi}}_{XY}^{k,N}\}$ corresponding to the empirical covariance operator of the pooled sample of $N = n_1 + n_2$ curves and base our test on

$$\sum_{k=1}^{K} \|(\widehat{\mathscr{R}}_X^N - \widehat{\mathscr{R}}_Y^N)\widehat{\boldsymbol{\varphi}}_{XY}^{k,N}\|_{\mathcal{L}^2}^2,$$

which by Parseval's theorem, may be further approximated by

$$\sum_{i=1}^{K}\sum_{j=1}^{K} \langle(\widehat{\mathscr{R}}_X^N - \widehat{\mathscr{R}}_Y^N)\widehat{\boldsymbol{\varphi}}_{XY}^{i,N}, \widehat{\boldsymbol{\varphi}}_{XY}^{j,N}\rangle^2. \tag{3}$$

With this quantity in mind, the following theorem, whose proof may be found in the Appendix, provides the basis for our test:

Theorem 1. Let $\{\mathbf{X}_n\}_{n=1}^{n_1}$ and $\{\mathbf{Y}_n\}_{n=1}^{n_2}$ be two collections of zero mean iid continuous Gaussian random functions indexed by the interval $[0, 1]$ and taking values in $\mathbb{R}^d$, possessing covariance operators $\mathscr{R}_X$ and $\mathscr{R}_Y$ with distinct eigenvalues. Let $\widehat{\mathscr{R}}_X^{n_1}$ and $\widehat{\mathscr{R}}_Y^{n_2}$ denote the empirical covariance operators based on $\{\mathbf{X}_n\}_{n=1}^{n_1}$ and $\{\mathbf{Y}_n\}_{n=1}^{n_2}$. For $N = n_1 + n_2$, let $\widehat{\mathscr{R}}_{XY}^N$ denote the empirical covariance operator of the pooled collection, and $\{\widehat{\boldsymbol{\varphi}}_{XY}^{k,N}\}_{k=1}^N$ the corresponding eigenfunctions. Finally, let $\widehat{\lambda}_{X,XY}^{k,n_1}$, $\widehat{\lambda}_{Y,XY}^{k,n_2}$ denote the empirical variance of the kth Fourier coefficient of $\{\mathbf{X}_n\}_{n=1}^{n_1}$ and $\{\mathbf{Y}_n\}_{n=1}^{n_2}$, respectively, with respect to the eigenfunctions $\{\widehat{\boldsymbol{\varphi}}_{XY}^{n,K}\}_{n=1}^N$. Assuming that $\mathbb{E}[\|\mathbf{X}_1\|_{L^2}^4] < \infty$, $\mathbb{E}[\|\mathbf{Y}_1\|_{L^2}^4] < \infty$, and $n_1/N \to \theta \in (0, 1)$ as $N = n_1 + n_2 \to \infty$, it follows that, under the hypothesis $H_0 : \mathscr{R}_X = \mathscr{R}_Y$,

$$\begin{aligned} T_N(K) \;:=\; & \frac{n_1 n_2}{2N}\sum_{i=1}^{K}\sum_{j=1}^{K}\big(\langle(\widehat{\mathscr{R}}_X^{n_1} - \widehat{\mathscr{R}}_Y^{n_2})\widehat{\boldsymbol{\varphi}}_{XY}^{i,N}, \widehat{\boldsymbol{\varphi}}_{XY}^{j,N}\rangle^2\big) \\ & \Big/\left(\left(\frac{n_1}{N}\widehat{\lambda}_{X,XY}^{i,n_1} + \frac{n_2}{N}\widehat{\lambda}_{Y,XY}^{i,n_2}\right)\right. \\ & \left.\times\left(\frac{n_1}{N}\widehat{\lambda}_{X,XY}^{j,n_1} + \frac{n_2}{N}\widehat{\lambda}_{Y,XY}^{j,n_2}\right)\right) \\ \xrightarrow{w}\; & \chi^2_{K(K+1)/2} \end{aligned}$$

as $N \to \infty$, for any finite $K \le \operatorname{rank}(\mathscr{R}_X) = \operatorname{rank}(\mathscr{R}_Y) \le \infty$.

Under the alternative hypothesis, the test statistic will converge to a sum of $K(K+1)/2$ dependent shifted chi square random variables.

Our proposed test procedure is thus to *reject* the hypothesis $H_0 : \mathscr{R}_X = \mathscr{R}_Y$ at level α, whenever the test statistic exceeds the corresponding critical value,

$$T_N(K) \ge \chi^2_{K(K+1)/2,1-\alpha}.$$

Of course, conducting the test requires the selection of a spectral truncation level, K. This choice must be made judiciously, as it has a direct bearing on the power of the test:

1. Conservative choices of K [i.e., choosing $K \ll \operatorname{rank}(\mathscr{R}_X) \wedge \operatorname{rank}(\mathscr{R}_Y)$] may result in Type II error due to differences in the higher frequency covariance structure, especially in situations where the two covariances share the same eigenfunctions, but have different eigenvalues at higher frequencies.
2. Greedy choices of K [choosing $K > \operatorname{rank}(\mathscr{R}_X) \wedge \operatorname{rank}(\mathscr{R}_Y)$] will inflate the variance of the test statistic since an element of ill-posedness will enter when dividing with the empirical eigenvalues of higher order terms.

In the latter sense, the test can also be thought of as an $\mathcal{L}^2$-regularized test. These aspects are further considered quantitatively in Section 4. It should be noted that the problem of choosing K is directly analogous to the choice of a cutoff point in principal component analysis and the choice of a bandwidth in a nonparametric problem; thus we deal with it using empirical eigenvalue scree-plots as well as penalized goodness-of-fit criteria (see Sections 3.3 and 5.1).

A more user-friendly expression for the test statistic T can be given if we introduce some additional notation. Let $\widehat{\lambda}_{X,XY}^{ij,N} := \langle\widehat{\mathscr{R}}_X^{n_1}\widehat{\boldsymbol{\varphi}}_{XY}^{i,N}, \widehat{\boldsymbol{\varphi}}_{XY}^{j,N}\rangle = n_1^{-1}\sum_i \langle\mathbf{X}_i - \overline{\mathbf{X}}, \widehat{\boldsymbol{\varphi}}_{XY}^{i,N}\rangle\langle\mathbf{X}_i - \overline{\mathbf{X}}, \widehat{\boldsymbol{\varphi}}_{XY}^{j,N}\rangle$ be the empirical covariance of the ith and jth Fourier coefficients of the X-curves, with respect to the basis $\{\widehat{\boldsymbol{\varphi}}_{XY}^{k,N}\}_{k\ge1}$ (resp. $\widehat{\lambda}_{Y,XY}^{ij,N}$). For simplicity, we also write $\widehat{\lambda}_{X,XY}^{ji,N} \equiv \widehat{\lambda}_{X,XY}^{ij,N}$ (resp. $\widehat{\lambda}_{Y,XY}^{ji,N}$). Then we may re-express the test statistic as

$$\begin{aligned} T_N(K) := \frac{n_1 n_2}{2N}\sum_{i=1}^{K}\sum_{j=1}^{K}\big((\widehat{\lambda}_{X,XY}^{ij,N} - \widehat{\lambda}_{Y,XY}^{ij,N})^2\big) \\ \Big/\left(\left(\frac{n_1}{N}\widehat{\lambda}_{X,XY}^{i,n_1} + \frac{n_2}{N}\widehat{\lambda}_{Y,XY}^{i,n_2}\right)\right. \\ \left.\times\left(\frac{n_1}{N}\widehat{\lambda}_{X,XY}^{j,n_1} + \frac{n_2}{N}\widehat{\lambda}_{Y,XY}^{j,n_2}\right)\right). \end{aligned}$$

If for some reason, we a priori know the eigenfunctions of $\mathscr{R}_X$ and $\mathscr{R}_Y$ to be equal, then the following test statistic may be used instead of T:

$$T_1 = \sum_{k=1}^{K}\frac{n_1 n_2}{N}\frac{(\widehat{\lambda}_{X,XY}^{k,N} - \widehat{\lambda}_{Y,XY}^{k,N})^2}{2((n_1/N)\widehat{\lambda}_X^{k,N} + (n_2/N)\widehat{\lambda}_Y^{k,N})^2}.$$

The motivation for this statistic is that when the eigenfunctions coincide, then

$$\sum_{k=1}^{K}\|(\widehat{\mathscr{R}}_X^{n_1} - \widehat{\mathscr{R}}_Y^{n_2})\widehat{\boldsymbol{\varphi}}_{XY}^{k,N}\|_{\mathcal{L}^2}^2 \approx \sum_{k=1}^{K}(\widehat{\lambda}_{X,XY}^{k,N} - \widehat{\lambda}_{Y,XY}^{k,N})^2.$$

It follows as an immediate corollary to Theorem 1 that, under H_0, the statistic T_1 is asymptotically chi-square distributed with K degrees of freedom [assuming $n_1/N \to \theta \in (0, 1)$]. One may also wish to consider modified versions of the test statistics T and T_1, obtained via suitable variance-stabilizing transformations. In the case of the test statistic T, we apply a log transformation to the diagonal terms of the sum in Equation (3), and Fisher's z-transformation to the off-diagonal terms to obtain a

test statistic with the same asymptotic distribution as T (an immediate corollary to Theorem 1),

$$T^* = \sum_{k=1}^{K} \frac{n_1 n_2}{N} \frac{(\log \widehat{\lambda}_{X,XY}^{k,N} - \log \widehat{\lambda}_{Y,XY}^{k,N})^2}{2} + \sum_{1 \le j < k \le K} \frac{n_1 n_2}{N} \left(\frac{1}{2} \log \frac{\sqrt{\widehat{\lambda}_{XY}^{j,N} \widehat{\lambda}_{XY}^{k,N}} + \widehat{\lambda}_{X,XY}^{jk,N}}{\sqrt{\widehat{\lambda}_{XY}^{j,N} \widehat{\lambda}_{XY}^{k,N}} - \widehat{\lambda}_{X,XY}^{jk,N}} - \frac{1}{2} \log \frac{\sqrt{\widehat{\lambda}_{XY}^{j,N} \widehat{\lambda}_{XY}^{k,N}} + \widehat{\lambda}_{Y,XY}^{jk,N}}{\sqrt{\widehat{\lambda}_{XY}^{j,N} \widehat{\lambda}_{XY}^{k,N}} - \widehat{\lambda}_{Y,XY}^{jk,N}} \right)^2.$$

A variance-stabilized alternative to T_1 may also be similarly constructed by retaining only the first component of T^* (the diagonal terms), yielding

$$T_1^* = \sum_{j=1}^{K} \frac{n_1 n_2}{N} \frac{(\log \widehat{\lambda}_{X,XY}^{j,N} - \log \widehat{\lambda}_{Y,XY}^{j,N})^2}{2}.$$

The latter statistic is approximately χ^2-distributed with K degrees of freedom. Simulations conducted in Section 4 seem to suggest that the modified tests achieve a level closer to the nominal level, and consequently, may provide higher power.

In the infinite rank case, one might wish to let K to grow along with N, allowing for the comparison of progressively finer and finer differences (located at the extreme tails of the operator spectra) as sample size increases. As noted previously, any such attempt will necessarily lead to instabilities: due to the fast decay of the eigenvalues, we are attempting to compare extremely small quantities, based on the empirical tails of the spectra, which are highly unstable. This instability will manifest itself through the very large integrated mean squared errors involved when estimating higher order eigenfunctions, whose available bounds grow for fixed N depending inversely on the rate of decay of the spectrum (see also Bosq 2000, lemma 4.3); the ill-posedness is especially severe for smooth processes. Controlling the rate of growth of K with respect to both N and the rate of decay of the true eigenvalues will thus be necessary—decreasing the amount of regularization requires an increase in sample size, depending also on the spectral decay properties. Modifying the test statistic to obtain a central limit theorem as $K_N \to \infty$ will require a very slow rate of growth of K_N with respect to N since:

1. Although the truncation level grows as K_N, the number of summands in the test statistic grows like K_N^2.
2. While these K_N^2 summation terms do become independent as N grows (allowing for a CLT phenomenon), no *mixing concept* applies. In effect this means that one has to look at the convergence in distribution to independence of a random vector of increasing dimension ($= K_N^2$). For any fixed dimension the required weak convergence will be at a rate of $N^{-1/2}$—therefore K_N must grow slow enough to allow the $N^{-1/2}$ rate to compensate for the K_N^2 rate of increase of the dimension.
3. The required global convergence to independence is regulated by the convergence of the empirical eigenfunctions to the true ones; this in turn depends on the spacings between the true eigenvalues. For K components, the rate of convergence of the Kth empirical eigenfunction decays like $N^{-1/2} \max\{(\lambda_{K-1} - \lambda_K)^{-1}, (\lambda_K - \lambda_{K+1})^{-1}\}$. Therefore, when we let K_N grow, it has to be at a rate slow enough to annihilate the blow-up of the inverse spacing of order K_N.

The study of these intricacies is rather technical, and further development is contained in the supplement.

3.3 On the Selection of Truncation Level

By analogy to finite-dimensional principal component analysis (PCA), the choice of a truncation parameter K can be made on the basis of scree plots and cumulative variance plots. A visual inspection of the scree plots can be employed to identify inflection points, which combined with the information provided by the cumulative variance plots, can suggest an appropriate truncation level K for use in testing. Note that the decrease of the scores $\widehat{\lambda}_{X,XY}^{k,N}$ and $\widehat{\lambda}_{Y,XY}^{k,N}$ is not monotone, since the basis $\{\widehat{\boldsymbol{\varphi}}_{XY}^{k,N}\}$ does not correspond to the eigenbasis of either of the two groups of curves. Therefore, a little more care needs to be taken, although the basic idea still holds.

The truncation of the Hilbert–Schmidt norm expansion effectively induces smoothing upon the curves, and can be regarded as a choice of a *regularization tuning parameter*. Consequently, potentially more automatic criteria can be based on tuning the amount of smoothing so as to minimize a penalized goodness-of-fit error. Concentrating on the X-curves, a natural definition of goodness-of-fit error is,

$$\mathrm{PE}_X(K) := \sum_{n=1}^{n_1} \left\| \sum_{k=1}^{K} \langle \mathbf{X}_n^*, \widehat{\boldsymbol{\varphi}}_{XY}^{k,N} \rangle \widehat{\boldsymbol{\varphi}}_{XY}^{k,N} - \mathbf{X}_n^* \right\|_{\mathcal{L}^2}^2 = \sum_{n=1}^{n_1} \|\widetilde{\mathbf{X}}_n(K) - \mathbf{X}_n^*\|_{\mathcal{L}^2}^2,$$

where $\mathbf{X}_i^*$ is the ith mean-corrected curve. Of course, the above criterion is nonincreasing in K since it accounts only for the fit, and there is no penalty for the "complexity" of $\widetilde{\mathbf{X}}_n(K)$. Such a penalty is often based on the norm of the image of $\widetilde{\mathbf{X}}_n(K)$ through a suitably chosen differential operator (in the spirit of Ramsay and Silverman 2005, section 5.3.3). The choice of penalty reflects the qualitative specification of what "parsimonious" is in a given context. In the present scenario, a sample of curves is available, and so the penalty can be made to be data-dependent, by penalizing deviations from the average smoothness properties of the observed curves. These smoothness properties are naturally reflected by the norm of the *reproducing kernel Hilbert space* (RKHS) generated by the empirical covariance operator of the X-sample, $\widehat{\mathscr{R}}_X$, yielding the penalized

fit criterion,

$$\mathrm{PFC}_X(K) = \underbrace{\sum_{n=1}^{n_1} \|\widetilde{\mathbf{X}}_n(K) - \mathbf{X}_n^*\|^2_{\mathcal{L}^2}}_{\mathrm{GOF}_X(K)} + \underbrace{\frac{2\sum_{j=1}^{N}\widehat{\lambda}_{XY}^{j,N}}{n_1}\sum_{n=1}^{n_1}\sum_{j=1}^{n_1}\frac{1}{\widehat{\lambda}_X^{j,N}}\langle\widetilde{\mathbf{X}}_n(K), \widehat{\boldsymbol{\varphi}}_X^{j,N}\rangle^2}_{\mathrm{PEN}_X(K)}. \quad (4)$$

When the null hypothesis is true, we expect to have $\widehat{\boldsymbol{\varphi}}_X^{j,N} \approx \widehat{\boldsymbol{\varphi}}_{XY}^{j,N}$; this essentially reduces $\mathrm{PFC}_X(K)$ to the Gaussian pseudo-likelihood-based Akaike information criterion (AIC) employed by Yao, Müller, and Wang (2005a) (see also Yao, Müller, and Wang 2005b). The analogous quantity $\mathrm{PFC}_Y(K)$ can similarly be defined for the Y-curves. Since the sample size for the two groups are not equal, the natural choice of K is then given by minimizing the sum of goodness-of-fit terms [$\mathrm{GOF}_X(K)$ and $\mathrm{GOF}_Y(K)$] plus the convex combination of the smoothness penalties [$\mathrm{PEN}_X(K)$ and $\mathrm{PEN}_Y(K)$]:

$$\arg\min_K \Big\{\mathrm{GOF}_X(K) + \mathrm{GOF}_Y(K) + \frac{n_1}{N}\mathrm{PEN}_X(K) + \frac{n_2}{N}\mathrm{PEN}_Y(K)\Big\}.$$

In practice, the number of terms taken in the sum comprising the penalty may be less than n_i, to avoid dividing by terms that are numerically zero. A variant of this selection criterion can be based on the leave-one-out cross-validated prediction error, where one whole curve is left out at a time (Rice and Silverman 1991). The performance of the selection criterion is investigated in simulations presented in the next section.

4. A SIMULATION STUDY

To assess the behavior of the proposed tests under the null hypothesis and under various alternatives we carry out a number of simulations. We consider one situation with equal covariance functions (simulation scenario A) and several alternative configurations (scenarios B–I). The two test statistics T and T^* introduced in the previous section are considered under various choices of K, the truncation level, and for the automatic selection K^* given by the penalized fit criterion. The number of observations in each sample is 50. The tests are replicated 5000 times under H_0 and 1000 times under H_A, respectively, at the 5% nominal level of significance using the asymptotic χ^2 approximation.

In the first eight scenarios, the Gaussian processes in both samples are of the form

$$\sum_{j=1}^{3}\xi_j\sqrt{2}\sin(2\pi j(t+\delta_j)) + \sum_{j=1}^{3}\zeta_j\sqrt{2}\cos(2\pi j(t+\eta_j)), \qquad t\in[0,1],$$

where the coefficients ξ_j, ζ_j are independent Gaussian random variables with mean zero and $\mathrm{var}(\xi_j) = v_j$, $\mathrm{var}(\zeta_j) = w_j$ (the variance terms where chosen so as to induce "elbow" effects as one expects to see in practice). Various values of v_j, w_j, δ_j, η_j used in A–H are reported together with the corresponding results in Table 1 (the shift parameters δ_j, η_j are reported only for F, the only case where they are nonzero). The last scenario deals with rough processes (infinitely many components).

Results for scenario A show that the true level for all variants of the test is close to the nominal level, provided the number of

Table 1. Empirical rejection probabilities on the nominal level 5%, sample size $n_1 = n_2 = 50$, number of replications 5000 for A, 1000 for B–I. Here, $\mathbf{u}^X = (\mathbf{v}^X, \mathbf{w}^X)$ (resp. $\mathbf{u}^Y$) and K^* is the automatic truncation choice given by the penalised fit criterion

			K				
	Parameters	Test	1	2	3	4	K^*
A	$\mathbf{u}^X = (12, 7, 0.5, 9, 5, 0.3)$	T	0.045	0.049	0.044	0.044	0.047
	$\mathbf{u}^Y = (12, 7, 0.5, 9, 5, 0.3)$	T^*	0.051	0.056	0.057	0.056	0.059
B	$\mathbf{u}^X = (14, 7, 0.5, 6, 5, 0.3)$	T	0.422	0.264	0.185	0.150	0.148
	$\mathbf{u}^Y = (8, 7, 0.5, 6, 5, 0.3)$	T^*	0.443	0.315	0.223	0.174	0.175
C	$\mathbf{u}^X = (15, 10, 0.5, 4, 3, 0.3)$	T	0.186	0.331	0.218	0.169	0.167
	$\mathbf{u}^Y = (11, 6, 0.5, 4, 3, 0.3)$	T^*	0.201	0.366	0.269	0.207	0.208
D	$\mathbf{u}^X = (12, 7, 0.5, 9, 3, 0.3)$	T	0.040	0.204	0.836	0.973	0.962
	$\mathbf{u}^Y = (12, 7, 0.5, 2, 5, 0.3)$	T^*	0.047	0.221	0.848	0.984	0.980
E	$\mathbf{u}^X = (12, 7, 0.5, 9, 3, 0.3)$	T	0.047	0.246	0.644	0.964	0.962
	$\mathbf{u}^Y = (12, 7, 0.5, 3, 9, 0.3)$	T^*	0.055	0.267	0.686	0.976	0.975
F	$\mathbf{u}^X = \mathbf{u}^Y = (12, 7, 4, 0.5, 0.3, 0.1)$	T	0.257	0.693	0.909	1.000	1.000
	$\boldsymbol{\delta}^X = (0.15, 0.15, 0.15)$	T^*	0.273	0.706	0.916	1.000	1.000
G	$\mathbf{u}^X = (12, 7, 0.5, 8, 6, 0.3)$	T	0.042	0.040	0.054	1.000	1.000
	$\mathbf{u}^Y = (12, 7, 0.5, 8, 0, 0.3)$	T^*	0.047	0.048	0.068	1.000	1.000
H	$\mathbf{u}^X = (12, 7, 0.5, 9, 5, 0.3)$	T	0.044	0.140	0.500	1.000	1.000
	$\mathbf{u}^Y = (12, 7, 0.5, 0, 5, 0.3)$	T^*	0.049	0.154	0.520	1.000	1.000
I	Brownian motion versus	T	0.719	0.608	0.483	0.377	0.493
	Ornstein–Uhlenbeck process	T^*	0.731	0.644	0.532	0.443	0.546

components K does not exceed the effective complexity of the covariance operator (which is 4 in this case). The slight conservatism of T is removed by variance stabilizing transformations used in T^*. Indeed, the stabilized statistics seem to be preferable because they also provide slightly higher power (as is seen in the remaining simulations).

Under scenario B, both covariance operators are of effective complexity 4 and possess the same sequence of eigenfunctions (the same set with the same order), but the sequences of eigenvalues differ (the largest eigenvalue is different). Not surprisingly, the power decreases as K increases because there is no difference in the components other than in the first one, so adding them increases the degrees of freedom without any significant contribution to the test statistic. Configuration C is similar to B, but with the two largest eigenvalues being different. The highest power is achieved with $K = 2$, as expected. When compared to the next few scenarios, where there are differences associated with the eigenfunctions also, the power in B and C is clearly lower. This is due to the fact that the test statistic takes the comparison of the eigenfunctions—where there are no differences—into account, and thus is not as powerful in detecting differences that lie only on the eigenvalues (the diagonal form of the tests T_1 and T_1^* will be more powerful in this case).

In scenario D, the effective complexity of the operators is the same in Equation (4), the operators have the same set of eigenfunctions (in different order) and different sequences of eigenvalues. The difference of the covariance operators is not detected by tests with one component because the largest eigenvalue and the corresponding eigenfunction are the same in both samples. When the choice of K is close to the true effective complexity, the power of the tests is very high (this includes the automatic choice). The same is true for the next four scenarios as well.

Under scenario E, both operators (of effective rank 4) have the same sequence of eigenvalues, and the same set of eigenfunctions, but the latter are permuted to correspond to different eigenvalues. This scenario illustrates a situation where the diagonal form of the test statistics (T_1 and T_1^*) will be inapplicable. It is interesting to make the comparison with scenario D, where the sets of eigenfunctions are the same for both samples as well. In D the sequences of eigenvalues differ also, hence more information is on the diagonal.

Scenario F differs from the previous configurations in that the sets of eigenfunctions are completely different (sines versus shifted sines). The eigenvalues are the same, and the effective operator rank is 3 in both cases.

In the next configuration, scenario G, the first three eigenvalues and eigenfunctions are the same in both samples. The covariance operators have different effective ranks: 4 in the first sample, 3 in the second sample. Therefore, it is not surprising that the departure from H_0 is not detected by tests with less than 4 components while it is clearly detected by four-component tests. Note that with the automatic choice K^*, the alternative is always detected.

Configuration H is again a situation with different effective ranks of operators (4 versus 3) but unlike the previous situation, only the first eigenfunction and eigenvalue coincide in both samples. The next two eigenvalues are different and the corresponding eigenfunctions differ as well. Thus, as of $K = 2$, the tests start detecting the alternative, with highest power for $K = 4$.

Under scenario I, curves in both samples come from distributions with covariance operators with infinite rank, namely the standard Brownian motion $W(t)$ and the Ornstein–Uhlenbeck process $U(t)$ satisfying $dU(t) = -\theta U(t)\,dt + dW(t)$ with $\theta = 1$. The covariance operators of the two processes differ in all components. The major portion of the difference is captured by tests with one component, then the power slowly decays.

A general observation when focusing on the behavior of the tests when the number of components K was selected using the selection criterion introduced in the previous section is that the power and level are comparable with those when employing the true effective rank. Under scenario A, the selection criterion chose $K = 4$ in 96.3% of simulations and $K = 5$ in 3.7% of simulations. Doing the same for the alternative configurations, it turned out that the power is similar to the power of tests with fixed values of K close to the values most frequently selected by the selection criterion. Hence this automatic dimension reduction technique appears to be useful in practice.

It should be mentioned that the role of the selection criterion is to probe the effective complexity of the data and not the complexity of the difference between the two samples. The selection rule is not related to the null hypothesis or the alternative and does not reflect validity or invalidity of either of them. This explains the reliability of the post-selection test. Note that a completely different approach can be based on the selection of the "most different" components (the most likely alternative) using a criterion involving the test statistic in the spirit of data-driven smooth tests (e.g., Ledwina 1994).

5. ANALYSIS OF DNA MINICIRCLES

5.1 Finite-Dimensional Approximation

Figure 3 shows the empirical variance of the scores with respect to the basis $\{\boldsymbol{\varphi}_{XY}^{k,N}\}$ separately for the TATA and CAP groups ($\widehat{\lambda}_{X,XY}^{k,N}$ and $\widehat{\lambda}_{Y,XY}^{k,N}$, respectively, in the notation used previously) as well as for the pooled sample ($\widehat{\lambda}_{XY}^{j,N}$). The plots also display cumulative proportions of the total variance explained by the corresponding components. Separate plots are constructed for the analysis carried out marginally on each principal axis and jointly on the principal plane.

When inspecting the marginal plots for the projections on each axis of inertia, we observe that four or at most five principal components should constitute an adequate choice. When looking at the marginal plot for the projection onto the principal plane of inertia, it seems that setting $K = 6$ or $K = 7$ is more than adequate (accounting for at least 85% and 90% of the variance, respectively, and with a clear "elbow" effect).

The reason for placing special emphasis on the principal plane is that, as one can observe from Figure 2, the DNA minicircle curves tend to be planar on average, and the more interesting signal is not to be found in the deviations from the planar aspect of the structure, but within the planar structure itself (see the discussion at the end of the next section). The penalized prediction error criterion introduced in Section 3.3 yields $K = 7$ components in the principal plane.

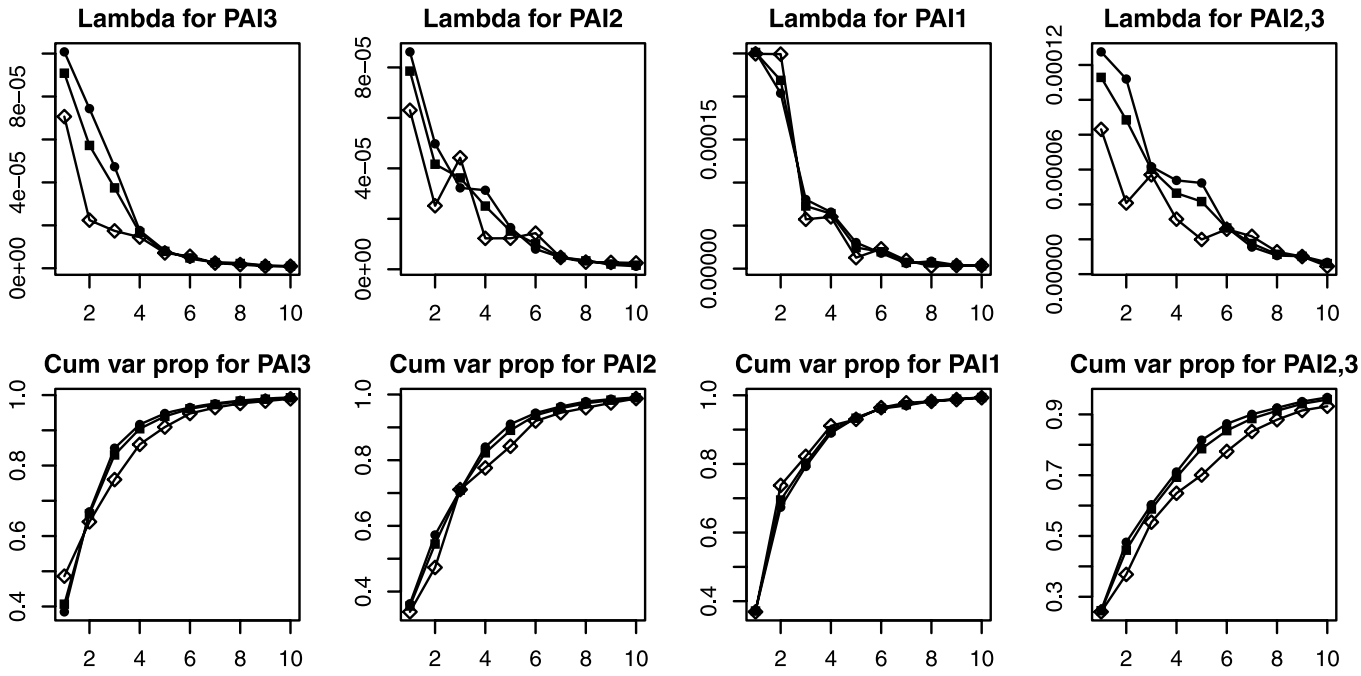

Figure 3. Empirical variances (scree plot) and cumulative proportions of variance explained by components for the TATA (circles) and CAP (diamonds) group and for both groups together (squares).

5.2 First-Order Inference

As was mentioned in the Introduction, a previous exploratory analysis of the data (Amzallag et al. 2006) that used clustering of the minicircles with respect to a Procrustean metric did not reveal any observable differences between the geometry of the two groups. The clustering distance used (a mean-square-based pairwise Procrustean distance) induces clustering with respect to the mean shape of the minicircles, which can be seen to be essentially identical between the two groups (Figure 2). To probe this finding more formally, we test the hypothesis of equal mean curves versus a general alternative, based on a variant of the test proposed by Berkes et al. (2009). We reject the hypothesis of equal mean curves when the value of the statistic

$$\sum_{j=1}^{K} \frac{n_1 n_2}{N} \frac{(\langle \overline{\mathbf{X}}, \widehat{\boldsymbol{\varphi}}_{XY}^{j,N} \rangle - \langle \overline{\mathbf{Y}}, \widehat{\boldsymbol{\varphi}}_{XY}^{j,N} \rangle)^2}{\widehat{\lambda}_{XY}^{j,N}}$$

is large compared to a χ_K^2 distribution (the approximation employs results in Dauxois, Pousse, and Romain 1982). The results of this comparison are displayed in Table 2. The corresponding values of the test statistic are insignificant and one cannot reject the null hypothesis; indeed, the results of the test do not vary much with K.

Table 2. p-values for comparison of mean functions in the TATA and CAP group for various truncation levels K, for the full three-dimensional curves, and their projections onto the prinipal plane of inertia

K	PAI1, 2, 3	PAI2, 3
1	0.40	0.64
2	0.68	0.69
3	0.85	0.64
4	0.60	0.55
5	0.34	0.58
6	0.46	0.61

As discussed in the previous section, it seems, in fact, that the interesting "signal" of the minicircles is effectively planar (see Figure 2). It is, therefore, interesting to test the hypothesis that the mean function of the PAI1 coordinate is zero—for this will suggest that our analysis should concentrate on the prinicipal inertia plane (the projection of the Gaussian processes on this plane is obviously a Gaussian process). To this aim, we use the one-sample version of the test statistic used for mean comparison (which in the one-sample situation, is in fact an approximate likelihood-ratio statistic; Grenander 1981). For the TATA group the p-value of the test with $K = 4$ components is 0.29. For the CAP curves the p-value is 0.30 (also using four components). Hence the tests show no significant systematic deviation of the curves from the first principal plane, and their three-dimensional nature seems to only be due to random variation around a planar mean shape. For this reason, in the next section we concentrate on the comparison of the curves projected onto the principal plane of inertia.

5.3 Second-Order Inference

As the first-order comparison of the two minicircle groups did not reveal any significant differences, we turn our attention to the detection of second-order differences. Indeed, since the scientific hypothesis is that one type of curve (TATA) is more flexible, it may be intuitively expected that a detectable difference will lie in the covariance structure rather than the mean structure.

We test the hypothesis that both groups of curves share the same covariance operator by employing the test statistic T^*. The results are summarized in Table 3. Marginal tests on each

Table 3. p-values for the comparison of covariance functions in the TATA and CAP group on different principal inertia axes using the test statistic T^* under various truncation levels K

	p-value			
K	PAI3	PAI2	PAI1	PAI2, 3
1	0.252	0.313	0.976	0.167
2	0.001	0.118	0.823	0.005
3	0.000	0.087	0.782	0.025
4	0.001	0.022	0.886	0.051
5	0.001	0.053	0.555	0.009
6	0.010	0.087	0.327	0.005
7	0.019	0.098	0.360	0.023
8	0.046	0.173	0.148	0.094

inertia axis show that the covariance functions of the projections onto PAI3 seem significantly different for the two groups (with either the empirical selection $K = 4$ or the automatic choice $K = 5$). Differences of projections onto PAI2 appear marginally insignificant depending on the choice of K (the empirical choice is $K = 5$ and the automatic choice is $K = 7$). No significant difference is observed for PAI1, indicating that random deviations from the first principal plane may have the same covariance structure in the two groups (which is in keeping with our previous finding that the deviations from the principal plane can be thought to be residual). Since the curves appear to be planar on average, it is the covariance of their planar components where most structure is to be found. Indeed, when our test is carried out for the projection of the curves onto the principal plane of inertia using $K = 6$ (empirical) or $K = 7$ (automatic), it *rejects the null hypothesis of no flexibility differences, at the 1% and 3% significance levels, respectively*. In fact, the test based on T_1^* gives even more significant results, yielding a p-value that is numerically zero.

In the frequency domain, these differences can already be seen in the scree plots (Figure 3), where the TATA curves are seen to be more flexible in the sense that the variances of their Fourier coefficients are more inflated when compared to the CAP curves. Since the covariance kernels associated with the two operators under comparison are matrix-valued functions, there is no easy way to visualize the detected differences in the time domain. Figure 4 contains surface and contour plots of the empirical covariance kernels restricted to the third principal axis—the axis where the most significant differences were detected. The plot reveals differences both in terms of the norm as well as in terms of the structure.

6. CONCLUDING REMARKS

Motivated by the problem of comparison of groups of DNA minicircles, we introduce and study a testing procedure for two sample-comparison of Gaussian processes with respect to their covariance structure.

The proposed test function is based on an approximation of the Hilbert–Schmidt distance between the empirical covariance operators of the two groups, by means of the Karhunen–Loève representation of the pooled sample. The approximation

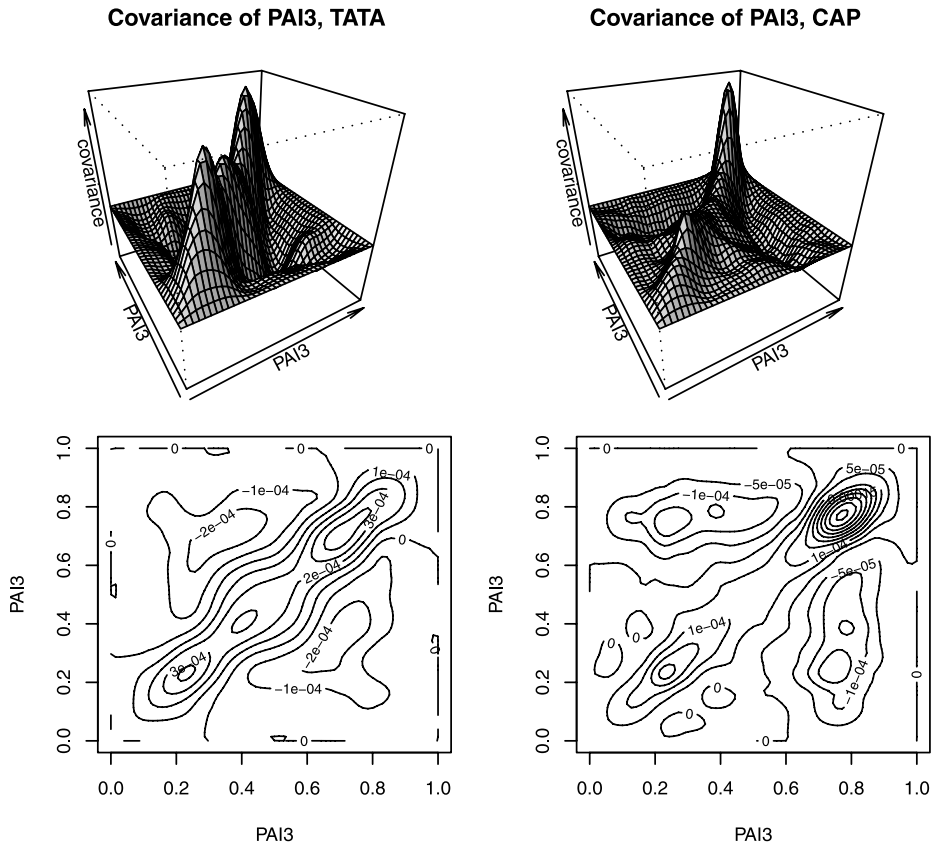

Figure 4. Surface and contour plots of the empirical covariance kernels corresponding to the TATA and CAP projections onto the third axis of inertia.

was seen to admit a *regularization* interpretation, the problem of testing presenting aspects of ill-posedness. The asymptotic distribution of the test function was established, and variance-stabilized variants with similar asymptotic properties were proposed. Finite-sample simulations under the null and various alternatives were used to investigate the performance of the proposed test. It should be noted that the results obtained readily extend to random functions defined over arbitrary compact Euclidean domains, and taking values in Euclidean spaces of arbitrary dimension (i.e., random fields).

The test was then carried out for a sample of 94 DNA minicircles of two different types. One type is believed to possess higher flexibility than the other, but this eluded empirical confirmation via electron microscopy. Our test rejected the hypothesis that the curves share the same covariance structure on their principal plane of inertia (the signals are essentially planar), providing support for the potential existence of differences between the geometry of the two groups. Interestingly, the difference was detected in the second-order characteristics, whereas previous analyses focused on first-order characteristics.

An important aspect of our testing procedure, as is the case with any spectral truncation regularization procedure, is the choice of truncation level K for the series representation of the Hilbert–Schmidt norm. A careless choice of truncation can affect the power of the test procedure. Our proposed approach for the choice of K was through visual inspection of functional PCA scree plots, combined with penalized prediction error minimization. Interesting further work will be to investigate LASSO-type component selection. Yet a further approach will be to consider *adaptive* modifications of the proposed tests that will automatically choose the level K based on the data; for example, tests based on statistics of the form $\max_K(T_N(K) - \beta K \log N)$, for some tuning parameter $\beta > 0$.

The asymptotic approximations for the distributions of the test statistics investigated hold for Gaussian processes. Departures from this assumption will affect the limiting law of the statistics. In simulations we observed that the test derived under the Gaussian assumption used in a non-Gaussian case becomes conservative when the scores have lighter tails than the normal distribution and anticonservative in the opposite case. Our tests are based on sums of squares of components which are asymptotically normal independent variables. When the data are not Gaussian, these components have asymptotically a multivariate normal distribution with unknown covariance structure. The limiting covariance matrix can be estimated and a chi-square test statistic can be based on the corresponding quadratic form (see also Horváth, Hušková, and Kokoszka 2010 for a similar approach in a different context). Some simulations showed that the convergence to the limiting distribution might be slow and one has to use only a small value of K, especially for the off-diagonal test.

Of course, testing whether a process is Gaussian is a research project in itself, but informal qq-plots constructed for the Karhunen–Loève coefficients of the minicircle data did not reveal any noteworthy departures from normality. For the benefit of the doubt, however, we also employed permutation tests based on our test statistics, with similar results—but with slightly more inflated p-values (Panaretos and Kraus 2009).

APPENDIX

Proof of Theorem 1

Introduce the notation $\mathscr{X}_i\mathbf{f} := \langle \mathbf{X}_i, \mathbf{f}\rangle \mathbf{X}_i$ and $\mathscr{Y}_i\mathbf{f} = \langle \mathbf{Y}_i, \mathbf{f}\rangle \mathbf{Y}_i$, so that $\widehat{\mathscr{R}}_X^n = n^{-1}\sum_i \mathscr{X}_i$ and $\widehat{\mathscr{R}}_Y^n = n^{-1}\sum_i \mathscr{Y}_i$. These are viewed as random elements of the Hilbert space of Hilbert–Schmidt operators acting on $\mathcal{L}^2[0,1]$. Under the hypothesis $H_0 : \mathscr{R}_X = \mathscr{R}_Y$, the collections $\{\mathscr{X}_i\}$ and $\{\mathscr{Y}_i\}$ are iid random operators with mean $\mathscr{R}_X = \mathscr{R}_Y$ and common covariance $\mathfrak{S} := \mathbb{E}[\mathscr{X}_i \otimes \mathscr{X}_i] - \mathscr{R}_X \otimes \mathscr{R}_X = \mathbb{E}[\mathscr{Y}_i \otimes \mathscr{Y}_i] - \mathscr{R}_Y \otimes \mathscr{R}_Y$, where $\otimes$ denotes the tensor product, $(u \otimes v)w = \langle v, w\rangle_{\mathcal{H}} u$ for any elements v, w, u of a Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$. In addition, our moment assumptions imply that $\mathbb{E}\|\mathscr{X}_i\|_{\mathrm{HS}}^2 < \infty$. We may, therefore, apply the Hilbert space central limit theorem (e.g., Bosq 2000, theorem 2.7) to conclude that

$$\sqrt{n_1}(\widehat{\mathscr{R}}_X^{n_1} - \mathscr{R}_X) \xrightarrow{w} \mathscr{Z}_1 \qquad \text{and}$$

$$\sqrt{n_2}(\widehat{\mathscr{R}}_Y^{n_2} - \mathscr{R}_Y) \xrightarrow{w} \mathscr{Z}_2 \quad \text{as } n_1, n_2 \to \infty,$$

where $\mathscr{Z}_1$ and $\mathscr{Z}_2$ are independent Gaussian random operators with mean 0 and covariance operator $\mathfrak{S}$. Now, given i, j, consider the sequence of random variables

$$W_N^{i,j} = \big\langle \sqrt{n_1 n_2/N}(\widehat{\mathscr{R}}_X^{n_1} - \widehat{\mathscr{R}}_Y^{n_2})\,\mathrm{sgn}[\langle \widehat{\boldsymbol{\varphi}}_{XY}^{i,N}, \boldsymbol{\varphi}_i\rangle]\widehat{\boldsymbol{\varphi}}_{XY}^{i,N}, \mathrm{sgn}[\langle \widehat{\boldsymbol{\varphi}}_{XY}^{j,N}, \boldsymbol{\varphi}_j\rangle]\widehat{\boldsymbol{\varphi}}_{XY}^{j,N}\big\rangle.$$

On the one hand, the strong law in Hilbert space implies that $\|\widehat{\mathscr{R}}_{XY}^N - \mathscr{R}_X\|_{\mathrm{HS}} \xrightarrow{\text{a.s.}} 0$ under the hypothesis H_0. Consequently, convergence also occurs with probability 1 in the strong operator topology, so that by Bosq (2000, lemma 4.3)

$$\big\|\,\mathrm{sgn}[\langle \widehat{\boldsymbol{\varphi}}_{XY}^{k,N}, \boldsymbol{\varphi}_k\rangle]\widehat{\boldsymbol{\varphi}}_{XY}^{k,N} - \widetilde{\boldsymbol{\varphi}}_k\big\|_{\mathcal{L}^2} \xrightarrow{\text{a.s.}} 0 \qquad \forall k \geq 1. \tag{A.1}$$

On the other hand, as $N \to \infty$ with $n_1/N \to \theta \in (0,1)$ we will have

$$\sqrt{\frac{n_2}{N}}\sqrt{n_1}\widehat{\mathscr{R}}_X^{n_1} - \sqrt{\frac{n_1}{N}}\sqrt{n_2}\widehat{\mathscr{R}}_Y^{n_2} \xrightarrow{w} \sqrt{1-\theta}\,\mathscr{Z}_1 - \sqrt{\theta}\,\mathscr{Z}_2 = \mathscr{Z}, \tag{A.2}$$

with $\mathscr{Z}$ a zero-mean Gaussian random operator with covariance $\mathfrak{S}$. Combining Equations (A.1) and (A.2) with the Hilbert space Slutsky lemma establishes that, for all $i, j \in \{1, \ldots, K\}$,

$$W_N^{i,j} \xrightarrow{w} \langle \mathscr{Z}\boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j\rangle.$$

For the next step, we note that $\mathscr{Z}$, being a Gaussian process itself, also admits a Karhunen–Loève decomposition, with respect to the eigenfunctions of $\mathfrak{S}$. These eigenfunctions can be retrieved directly from the definition of $\mathfrak{S}$ and the Karhunen–Loève expansion of the typical X process, $\mathbf{X} = \sum_i \sqrt{\lambda_i}\xi_i\boldsymbol{\varphi}_i$. Defining the operator $\Phi_{ij}\mathbf{f} := \langle \boldsymbol{\varphi}_i, \mathbf{f}\rangle\boldsymbol{\varphi}_j$, we immediately see that $\mathscr{X} = \sum_{i,j}\sqrt{\lambda_i\lambda_j}\xi_i\xi_j\Phi_{ij}$ and $\mathscr{R}_X = \sum_j \lambda_j\Phi_{jj}$. Hence, upon recalling that the $\{\xi_i\}$ are an iid standard Gaussian array we may write

$$\begin{aligned}
\mathfrak{S} &= \mathbb{E}[\mathscr{X} \otimes \mathscr{X}] - \mathscr{R}_X \otimes \mathscr{R}_X \\
&= \sum_{i,j,q,p}\sqrt{\lambda_i\lambda_j\lambda_p\lambda_q}\,\mathbb{E}[\xi_i\xi_j\xi_p\xi_q]\Phi_{ij} \otimes \Phi_{qp} - \sum_{i,j}\lambda_i\lambda_j\Phi_{ii} \otimes \Phi_{jj} \\
&= \sum_{i\neq j}\lambda_i\lambda_j\Phi_{ii} \otimes \Phi_{jj} + \sum_{i\neq j}\lambda_i\lambda_j\Phi_{ij} \otimes \Phi_{ji} + \sum_{i\neq j}\lambda_i\lambda_j\Phi_{ij} \otimes \Phi_{ij} \\
&\quad + \sum_i 3\lambda_i^2\Phi_{ii} \otimes \Phi_{ii} - \sum_i \lambda_i^2\Phi_{ii} \otimes \Phi_{ii} - \sum_{i\neq j}\lambda_i\lambda_j\Phi_{ii} \otimes \Phi_{jj} \\
&= 2\sum_i \lambda_i^2\Phi_{ii} \otimes \Phi_{ii} + \sum_{i\neq j}\lambda_i\lambda_j(\Phi_{ij} \otimes \Phi_{ji} + \Phi_{ij} \otimes \Phi_{ij}),
\end{aligned}$$

since $\mathbb{E}[\xi_i\xi_j\xi_p\xi_q]$ is 1 whenever pairs of indices are equal but not all indices are totally coincident, 3 when all indices are equal, and zero

otherwise. Regrouping the summation by adding the terms that are symmetric with respect to their indices, we further obtain

$$\begin{aligned}\mathfrak{S} &= 2\sum_i \lambda_i^2 \Phi_{ii}\otimes\Phi_{ii} \\ &\quad + \sum_{i<j}\lambda_i\lambda_j(\Phi_{ij}\otimes\Phi_{ji}+\Phi_{ij}\otimes\Phi_{ij}+\Phi_{ji}\otimes\Phi_{ij}+\Phi_{ji}\otimes\Phi_{ji}) \\ &= 2\sum_i \lambda_i^2 \Phi_{ii}\otimes\Phi_{ii} \\ &\quad + \sum_{i<j}\lambda_i\lambda_j\{\Phi_{ij}\otimes(\Phi_{ij}+\Phi_{ji})+\Phi_{ji}\otimes(\Phi_{ij}+\Phi_{ji})\} \\ &= \sum_i(\sqrt{2}\lambda_i)^2\Phi_{ii}\otimes\Phi_{ii}+\sum_{i<j}\lambda_i\lambda_j(\Phi_{ij}+\Phi_{ji})\otimes(\Phi_{ij}+\Phi_{ji}).\end{aligned}$$

It is straightforward to verify that $\{\Phi_{ij}+\Phi_{ji}\}_{i<j}\cup\{\Phi_{ii}\}_{i\geq 1}$ constitutes a complete orthogonal system of operators for the Hilbert space of Hilbert–Schmidt operators acting on $\mathcal{L}^2[0,1]$. We may, therefore, represent $\mathscr{Z}$ in a Karhunen–Loève expansion as

$$\mathscr{Z} = \sqrt{2}\sum_i \lambda_i\zeta_{ii}\Phi_{ii}+\sum_{i<j}\lambda_i^{1/2}\lambda_j^{1/2}\zeta_{ij}(\Phi_{ij}+\Phi_{ji})$$

for $\{\zeta_{ij}\}_{i,j=1}^{\infty}$ an iid array of standard Gaussian variables. Consequently, we may express the Gaussian process $\mathscr{Z}\boldsymbol{\varphi}_k$ as

$$\begin{aligned}\mathscr{Z}\boldsymbol{\varphi}_k &= \sqrt{2}\sum_{i=1}^{\infty}\lambda_i\zeta_{ii}\langle\boldsymbol{\varphi}_i,\boldsymbol{\varphi}_k\rangle\boldsymbol{\varphi}_i \\ &\quad + \sum_{i<j}\lambda_i^{1/2}\lambda_j^{1/2}\zeta_{ij}(\langle\boldsymbol{\varphi}_i,\boldsymbol{\varphi}_k\rangle\boldsymbol{\varphi}_j+\langle\boldsymbol{\varphi}_j,\boldsymbol{\varphi}_k\rangle\boldsymbol{\varphi}_i) \\ &= \sqrt{2}\lambda_k\zeta_{kk}\boldsymbol{\varphi}_k+\sum_{i<j}\lambda_i^{1/2}\lambda_j^{1/2}\zeta_{ij}\langle\boldsymbol{\varphi}_i,\boldsymbol{\varphi}_k\rangle\boldsymbol{\varphi}_j \\ &\quad + \sum_{i<j}\lambda_i^{1/2}\lambda_j^{1/2}\zeta_{ij}\langle\boldsymbol{\varphi}_j,\boldsymbol{\varphi}_k\rangle\boldsymbol{\varphi}_i \\ &= \sqrt{2}\lambda_k\zeta_{kk}\boldsymbol{\varphi}_k+\sum_{k<j}\lambda_k^{1/2}\lambda_j^{1/2}\zeta_{kj}\boldsymbol{\varphi}_j+\sum_{i<k}\lambda_i^{1/2}\lambda_k^{1/2}\zeta_{ik}\boldsymbol{\varphi}_i,\end{aligned}$$

where we used the fact that $\{\boldsymbol{\varphi}_i\}$ is an orthonormal system. It follows that for arbitrary $k,n\in\{1,\ldots,K\}$, the random variable $\langle\mathscr{Z}\boldsymbol{\varphi}_k,\boldsymbol{\varphi}_n\rangle$ admits the representation

$$\begin{aligned}\langle\mathscr{Z}\boldsymbol{\varphi}_k,\boldsymbol{\varphi}_n\rangle &= \sqrt{2}\lambda_k\zeta_{kk}\langle\boldsymbol{\varphi}_k,\boldsymbol{\varphi}_n\rangle+\sum_{k<j}\lambda_k^{1/2}\lambda_j^{1/2}\zeta_{kj}\langle\boldsymbol{\varphi}_j,\boldsymbol{\varphi}_n\rangle \\ &\quad + \sum_{i<k}\lambda_i^{1/2}\lambda_k^{1/2}\zeta_{ik}\langle\boldsymbol{\varphi}_i,\boldsymbol{\varphi}_n\rangle \\ &= \begin{cases}\sqrt{2}\lambda_k\zeta_{kk} & \text{if } k=n\\ \lambda_k^{1/2}\lambda_n^{1/2}\zeta_{kn} & \text{if } k<n\\ \lambda_k^{1/2}\lambda_n^{1/2}\zeta_{nk} & \text{if } k>n.\end{cases}\end{aligned}$$

It follows that $\langle\mathscr{Z}\boldsymbol{\varphi}_k,\boldsymbol{\varphi}_k\rangle\overset{\text{iid}}{\sim}\mathcal{N}(0,2\lambda_k^2)$ independently of $\langle\mathscr{Z}\boldsymbol{\varphi}_m,\boldsymbol{\varphi}_n\rangle\overset{\text{iid}}{\sim}\mathcal{N}(0,\lambda_m\lambda_n)$, $m\neq n$. Consequently, we have

$$\frac{1}{2}\frac{\langle\mathscr{Z}\boldsymbol{\varphi}_k,\boldsymbol{\varphi}_k\rangle^2}{\lambda_k^2}\overset{\text{iid}}{\sim}\chi_1^2,$$

independently of

$$\frac{1}{2}\frac{\langle\mathscr{Z}\boldsymbol{\varphi}_m,\boldsymbol{\varphi}_n\rangle^2+\langle\mathscr{Z}\boldsymbol{\varphi}_n,\boldsymbol{\varphi}_m\rangle^2}{\lambda_m\lambda_n}=\frac{\langle\mathscr{Z}\boldsymbol{\varphi}_m,\boldsymbol{\varphi}_n\rangle^2}{\lambda_m\lambda_n}\sim\chi_1^2.$$

The continuous mapping theorem now implies that

$$\begin{aligned}\frac{1}{2}\frac{(W_N^{ij})^2+(W_N^{ji})^2}{\lambda_i\lambda_j} &= \frac{n_1n_2}{2N}\sum_{i=1}^K\sum_{j=1}^K\frac{\langle(\widehat{\mathscr{R}}_X^{n_1}-\widehat{\mathscr{R}}_Y^{n_2})\widehat{\boldsymbol{\varphi}}_{XY}^{i,N},\widehat{\boldsymbol{\varphi}}_{XY}^{j,N}\rangle^2}{\lambda_i\lambda_j} \\ &\xrightarrow{w}\chi^2_{K(K+1)/2}.\end{aligned}$$

To complete the proof, we note that

$$\frac{n_1}{N}\widehat{\lambda}_{X,XY}^{k,n_1}+\frac{n_2}{N}\widehat{\lambda}_{Y,XY}^{k,n_2}\xrightarrow{p}\theta\lambda_k+(1-\theta)\lambda_k=\lambda_k\quad\forall k\in\{1,\ldots,K\},$$

so that the result follows from the application of Slutsky's lemma.

SUPPLEMENTAL MATERIALS

Additional plots and tables and detailed study: Additional plots and tables are available in a supplementary file. In addition, the supplementary file contains a more detailed study of the problem of comparing the complete spectrum, extending the discussion in the last part of Section 3.2. (Supplement.pdf)

[Received April 2009. Revised December 2009.]

REFERENCES

Adler, R. J. (1990), *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes. Lecture Notes and Monographs Series*, Hayward: Institute of Mathematical Statistics. [673]

Amzallag, A., Vaillant, C., Jacob, M., Unser, M., Bednar, M., Kahn, J. D., Dubochet, J., Stasiak, A., and Maddocks, J. H. (2006), "3D Reconstruction and Comparison of Shapes of DNA Minicircles Observed by Cryo-Electron Microscopy," *Nucleic Acids Research*, 34 (18), e125. [670,678]

Arnold, V. I. (1989), *Mathematical Methods of Classical Mechanics*, New York: Springer. [671]

Benko, M., Härdle, W., and Kneip, A. (2009), "Common Functional Principal Components," *The Annals of Statistics*, 37, 1–34. [670]

Berkes, I., Gabrys, R., Horváth, L., and Kokoszka, P. (2009), "Detecting Changes in the Mean of Functional Observations," *Journal of the Royal Statistical Society, Ser. B*, 71, 927–946. [670,678]

Bosq, D. (2000), *Linear Processes in Function Spaces*, New York: Springer. [675,680]

Cardot, H., Ferraty, F., Mas, A., and Sarda, P. (2003), "Testing Hypotheses in the Functional Linear Model," *Scandinavian Journal of Statistics*, 30 (1), 241–255. [670]

Cuevas, A., Febrero, M., and Fraiman, R. (2004), "An ANOVA Test for Functional Data," *Computational Statistics and Data Analysis*, 47, 111–122. [670]

Dauxois, J., Pousse, A., and Romain, Y. (1982), "Asymptotic Theory for the Principal Component Analysis of a Random Vector Function: Some Applications to Statistical Inference," *Journal of Multivariate Analysis*, 12, 136–154. [673,678]

Fan, J., and Lin, S.-K. (1998), "Tests of Significance When the Data Are Curves," *Journal of the American Statistical Association*, 93, 1007–1021. [670]

Ferraty, F., and Vieu, P. (2006), *Nonparametric Functional Data Analysis*, New York: Springer. [672]

Gabrys, R., and Kokoszka, P. (2007), "Portmanteau Test of Independence for Functional Observations," *Journal of the American Statistical Association*, 102, 1338–1348. [670]

Gasser, T., and Kneip, A. (1995), "Searching for Structure in Curve Samples," *Journal of the American Statistical Association*, 90, 1179–1188. [671]

Gervini, D. (2008), "Robust Functional Estimation Using the Median and Spherical Principal Components," *Biometrika*, 95 (3), 587–600. [672]

Gervini, D., and Gasser, T. (2004), "Self-Modelling Warping Functions," *Journal of the Royal Statistical Society, Ser. B*, 66 (4), 959–971. [671]

Giri, N. (1968), "On Tests of the Equality of Two Covariance Matrices," *The Annals of Mathematical Statistics*, 39, 275–277. [673]

Grenander, U. (1981), *Abstract Inference*, New York: Wiley. [670,678]

Hagerman, P. J. (1988), "Flexibility of DNA," *Annual Review Biophysics and Biophysical Chemistry*, 17, 265–286. [670]

Hall, P., and Hosseini-Nassab, M. (2006), "On Properties of Functional Principal Components Analysis," *Journal of the Royal Statistical Society, Ser. B*, 68 (1), 109–126. [673]

Hall, P., and Van Keilegom, I. (2007), "Two Sample Tests in Functional Data Analysis Starting From Discrete Data," *Statistica Sinica*, 17, 1511–1531. [670]

Horváth, L., Hušková, M., and Kokoszka, P. (2010), "Testing the Stability of the Functional Autoregressive Process," *Journal of Multivariate Analysis*, 101 (2), 352–367. [670,680]

Jacob, M., Blu, T., Vaillaint, C., Maddocks, J. H., and Unser, M. (2006), "3-D Shape Estimation of DNA Molecules From Stereo Cryo-Electron Micrographs Using a Projection Steerable Snake," *IEEE Transations on Image Processing*, 15 (1), 214–227. [671]

Kiefer, J., and Schwartz, R. (1965), "Admissible Bayes Character of T^2-Test, R^2-Test, and Other Fully Invariant Tests for Classical Multivariate Normal Problems," *The Annals of Mathematical Statistics*, 36, 747–770. [673]

Ledwina, T. (1994), "Data-Driven Version of Neyman's Smooth Test of Fit," *Journal of the American Statistical Association*, 89, 1000–1005. [677]

Panaretos, V. M., and Kraus, D. (2009), "Second Order Comparison of Gaussian Processes With Applications to DNA Shape Analysis," Technical Report 01-09, Chair of Mathematical Statistics, EPFL. [680]

Pillai, K. C. S. (1955), "Some New Test Criteria in Multivariate Analysis," *The Annals of Mathematical Statistics*, 26, 117–121. [673]

Ramsay, J. O., and Silverman, B. W. (2002), *Applied Functional Data Analysis: Methods and Case Studies*, New York: Springer. [673]

——— (2005), *Functional Data Analysis*, New York: Springer. [672,673,675]

Rice, J., and Silverman, B. W. (1991), "Estimating the Mean and Covariance Structure Nonparametrically When the Data Are Curves," *Journal of the Royal Statistical Society, Ser. B*, 53, 233–243. [676]

Roy, S. N. (1953), "On a Heuristic Method of Test Construction and Its Use in Multivariate Analysis," *The Annals of Mathematical Statistics*, 24, 220–238. [673]

Shen, Q., and Faraway, J. (2004), "An *F* Test for Linear Models With Functional Responses," *Statistica Sinica*, 14, 1239–1257. [670]

Tang, R., and Müller, H. G. (2008), "Pairwise Curve Synchronization for Functional Data," *Biometrika*, 95 (4), 875–889. [671]

Tolstorukov, M. Y., Virnik, K. M., Adhya, S., and Zhurkin, V. B. (2005), "A-Tract Clusters May Facilitate DNA Packaging in Bacterial Nucleoid," *Nucleic Acids Research*, 33 (12), 3907–3918. [670]

Vilar, J. M. G., and Leibler, S. (2003), "DNA Looping and Physical Constraints on Transcription Regulation," *Journal of Molecular Biology*, 331 (5), 981–989. [670]

Yao, F., Müller, H. G., and Wang, J. L. (2005a), "Functional Data Analysis of Sparse Longitudinal Data," *Journal of the American Statistical Association*, 100, 577–590. [676]

——— (2005b), "Functional Linear Regression Analysis for Longitudinal Data," *The Annals of Statistics*, 33, 2873–2903. [676]