Biometrika (2014), **xx**, x, *pp*. 1–18

Printed in Great Britain

Fréchet Integration and Adaptive Metric Selection for Interpretable Covariances of Multivariate Functional Data

BY ALEXANDER PETERSEN
Department of Statistics, University of California, Davis, California 95616, U.S.A.
alxpetersen@ucdavis.edu

HANS-GEORG MÜLLER
Department of Statistics, University of California, Davis, California 95616, U.S.A.
hgmueller@ucdavis.edu

SUMMARY

For multivariate functional data recorded for a sample of subjects on a common domain, one is often interested in the covariance between pairs of the component functions, extending the notion of a covariance matrix for multivariate data to the functional case. A straightforward approach is to integrate the pointwise covariance matrices over the functional time domain. We generalize this approach by defining the Fréchet integral, which depends on the metric chosen for the space of covariance matrices, and demonstrate that the ordinary integration is a special case when the Frobenius metric is used. As the space of covariance matrices is nonlinear, we propose a class of power metrics as alternatives to the Frobenius metric. For any such power metric, the calculation of Fréchet integrals is equivalent to transforming the covariance matrices with the chosen power, applying the classical Riemann integral to the transformed matrices, and finally applying the inverse transformation to return to the original scale. We also propose data-adaptive metric selection with respect to a user-specified target criterion, for example fastest decline of the eigenvalues, establish consistency of the proposed procedures and demonstrate their effectiveness in a simulation. The proposed functional covariance approach through Fréchet integration is illustrated in a comparison of connectivity between brain voxels for normal subjects and Alzheimer's patients based on fMRI data.

Some key words: Box–Cox Transformation, Covariance Matrix, Connectivity, Dependency, Fréchet Mean, Functional Correlation, Functional Data Analysis, functional Magnetic Resonance Imaging.

1. INTRODUCTION

Multivariate functional data are frequently encountered in practice. While several aspects of multivariate functional data have been studied, including various versions of principal component expansions (Zhou et al., 2008a; Berrendero et al., 2011; Serban et al., 2013; Chiou et al., 2014) and measures of depth (Claeskens et al., 2014), there are many open questions regarding the statistical modeling and analysis of such data. A prominent example of multivariate functional data occurs in neuroimaging, where fMRI scans result in signals that are measured over a time domain for a large number of voxels in the brain, for each individual in a sample. Measuring the association of the components of such multivariate functional data is of specific interest in many applications, for example to quantify functional connectivity in the brain. Functional connectivity between brain regions is thought to be associated with the overall functioning of the brain and is

used to quantify its aging and pathology (Friston, 2011; Lee et al., 2013). Specifically, patterns of connectivity seen in subjects with Alzheimer's disease may differ from those observed in cognitively normal subjects (Sheline & Raichle, 2013).

For many traditional applications of multivariate techniques, covariance matrices are a fundamental tool to quantify the dependence of the multivariate components. Functional theory for cross-covariance operators was developed by Baker (1974) and Gualtierotti (1979). Instead of cross-covariance operators, which are infinite-dimensional objects, our targets are covariance matrices that reflect the key features of dependency of the multivariate components. Important applications of covariance matrices in standard multivariate analysis include principal component analysis, which provides an optimal means of linear dimensionality reduction, and partial correlation, which is central for the construction of Gaussian graphical models. In the latter, precision matrices are used to compute partial correlations and to infer conditional independence between component pairs. The extension of these techniques to multivariate functional data is not straightforward since there is no obvious notion of a scalar covariance between two random functions. We address this issue here by proposing a simple approach to defining scalar measures of functional covariance.

Various notions of functional covariance and correlation have been explored. A first concept was functional canonical correlation (Leurgans et al., 1993; He et al., 2003, 2004) which, however, suffers from instability due to the inverse nature of canonical correlation. While this does not impede its use in low-dimensional multivariate settings, it causes problems for higher-dimensional and especially functional data. In such settings, regularization must be used to obtain a viable inverse of the auto-covariance operators, in generalization of the multivariate case where canonical correlation relies on a reliable estimate of the inverse covariance matrices of the random vectors between which canonical correlation is to be defined. When regularizing, the problem with the inverse is reflected by the fact that the canonical correlation is highly sensitive to the choice of the regularization parameter.

Alternative approaches aiming to avoid the inverse problem include dynamic correlation (Dubin & Müller, 2005), where one applies the interpretation of correlation as the cosine of an angle to the functional case by using the Hilbert space inner product (Chiou & Li, 2007), and singular correlation (Yang et al., 2011), where the inverse problem is circumvented by a singular value decomposition of the functional covariance surfaces. Dynamic correlation has been extended to a notion of functional partial correlation (Opgen-Rhein & Strimmer, 2006) with the goal of constructing networks based on Gaussian graphical models in genomics. Nonparametric versions such as Spearman correlation have also been proposed (Valencia et al., 2014).

The methodology presented in this paper aims at three central goals: first, to extend and complement the existing approaches for quantifying functional covariance by extending pointwise covariance to the multivariate functional case through generalized integration of covariance matrices over the time domain. This generalized integration is dependent on the metric chosen for the space of covariance matrices, which determines the underlying geometry. Second, we demonstrate the effects of different metrics in statistical applications and propose a method of data-adaptive metric selection. Various such metrics and their statistical implications have been recently discussed (Dryden et al., 2009; Pigoli et al., 2014). It is then a natural next step to discuss the data-adaptive selection of a specific metric from a pool of candidate metrics. Thirdly, we showcase the importance of functional covariance for the analysis of fMRI brain imaging data, where resting state scans are available for a sample of subjects. Specifically, we will show that our methods are well-suited to explore differences in neural connectivity between cognitively normal subjects and subjects who have been diagnosed with Alzheimer's disease.

2. Covariance Integration for Multivariate Functional Data

2·1. *Preliminaries*

Consider a multivariate stochastic process $X = \{X_1(t), \ldots, X_p(t) : t \in \mathcal{T}\}$ on a common domain $\mathcal{T}$, where we assume that $\mathcal{T} = [0, 1]$ and that X has the pointwise covariance matrices

$$\{\Sigma(t)\}_{kl} = \operatorname{cov}\{X_k(t), X_l(t)\} \quad (0 \leq t \leq 1). \tag{1}$$

The goal is to utilize this collection of covariance matrices to define an overall functional $p \times p$ covariance matrix S, where the (k, l)th element provides a scalar measure of the covariance between component functions X_k and X_l. This problem does not have a unique solution. A straightforward idea is to integrate the pointwise covariances for each component, obtaining $S_{kl} = \int_0^1 \{\Sigma(t)\}_{kl}\, dt$. In practice, the integral would be approximated by taking an average of the estimated pointwise covariances over a grid of support points at which measurements of processes X are available, as integrals can be defined as a limit of Riemann sums, $\int_0^1 f(x)\, dx \approx N^{-1} \sum_{i=1}^N f\{(i-1)/N\}$ for a continuous function f, if one has N equidistant support points.

This averaging aspect opens a connection to the work of Dryden et al. (2009) and Pigoli et al. (2014). Working in a different context, the authors propose to average a sample of covariance matrices by computing Fréchet means for metrics other than the Frobenius distance. They show that various such metrics improve upon the Frobenius distance in a variety of applications. Intuitively, since the Frobenius distance is induced by a Hilbert inner product on the space of all square matrices, its use for the nonlinear and more structured subset of covariance matrices may be suboptimal. In our simulation and data analysis in Section 4, we find that the use of other metrics often yields more powerful analyses than the Frobenius distance. Our starting point is to demonstrate how the usual Riemann integral of a multivariate function can be generalized to a Fréchet integral that depends on the specific metric on the space of covariance matrices.

2·2. *Fréchet Integrals*

For a metric space $(\mathcal{Q}, d)$, an interval $\mathcal{T} \subset \mathbb{R}$ and a function $q : \mathcal{T} \to \mathcal{Q}$, we define the Fréchet integral of q as an element $\omega^* \in \mathcal{Q}$ that satisfies

$$\omega^* = \operatorname*{arginf}_{\omega \in \mathcal{Q}} \int_{\mathcal{T}} d^2\{q(t), \omega\}\, \mathrm{d}t. \tag{2}$$

The existence and uniqueness of the Fréchet integral are not guaranteed in general, although existence is implied if $\mathcal{Q}$ is compact. We use this term to emphasize the interpretation of Fréchet integrals as an extension of the commonly used Fréchet mean (Fréchet, 1948) to the case of a continuous index, which is time in our case. If q were to be observed on a finite lattice of fixed dimension, the usual Fréchet mean would be applicable and, for the case of constructing a mean covariance matrix from a sample of covariance matrices, regular Fréchet means have been studied for various metrics and settings (Le, 2001; Ando et al., 2004; Arsigny et al., 2007).

Consider the special case where $\mathcal{Q}$ is a convex subset of $\mathcal{R}^m$ or, more generally, of an m-dimensional vector space. If $\int_{\mathcal{T}} q(t)\, \mathrm{d}t$ exists, it lies in $\mathcal{Q}$ by convexity. If one chooses the distance or metric d as the Euclidean distance d_{E}, it holds that

$$\int_{\mathcal{T}} d_{\mathrm{E}}^2 \left\{ q(t), \int_{\mathcal{T}} q(s)\, \mathrm{d}s \right\} \mathrm{d}t < \int_{\mathcal{T}} d_{\mathrm{E}}^2\{q(t), \omega\}\, \mathrm{d}t$$

for all $\omega \in \mathcal{R}^m$ with $\omega \neq \int_{\mathcal{T}} q(t)\, \mathrm{d}t$, so that $\omega^* = \int_{\mathcal{T}} q(t)\, \mathrm{d}t$. Hence, the usual Riemann integral is a special case of the Fréchet integral when $d = d_{\mathrm{E}}$.

2·3. *Fréchet Integration of Covariance Matrices*

Let $\mathcal{P}$ denote the space of $p \times p$ symmetric positive definite matrices and d a generic metric on $\mathcal{P}$. Using the matrix-valued function $\Sigma(t)$ defined in (1), a functional $p \times p$ covariance matrix can then be obtained by means of the Fréchet integral in (2) by taking $\mathcal{Q} = \mathcal{P}$ and $q(t) = \Sigma(t)$. For the metric d, we consider a general class of power metrics, motivated by the square root and logarithmic metrics utilized in Dryden et al. (2009). These metrics can be viewed in the context of the class of Box–Cox matrix transformations H_α and their inverses

$$H_\alpha(\Sigma) = \begin{cases} \alpha^{-1}(\Sigma^\alpha - I), & \alpha > 0, \\ \log(\Sigma), & \alpha = 0, \end{cases} \quad H_\alpha^{-1}(A) = \begin{cases} (\alpha A + I)^{1/\alpha}, & \alpha > 0, \\ \exp(A), & \alpha = 0. \end{cases} \tag{3}$$

The domain of H_α is $\mathcal{P}$ for all α, while the domain of H_α^{-1} is the space of symmetric matrices for $\alpha = 0$ and, for $\alpha > 0$, it is the space of symmetric matrices with eigenvalues greater than $-\alpha^{-1}$. Let d_{F} be the Frobenius metric, defined as $d_{\mathrm{F}}(A, B)^2 = \operatorname{tr}\left\{(A - B)^2\right\}$ for A and B two matrices of the same dimensions. Corresponding to the above class of transformations is the class of power metrics, indexed by $\alpha \geq 0$, with members

$$d_\alpha(\Gamma_1, \Gamma_2) = d_{\mathrm{F}}\left\{H_\alpha(\Gamma_1), H_\alpha(\Gamma_2)\right\} \quad (\Gamma_1, \Gamma_2 \in \mathcal{P}). \tag{4}$$

The particular cases $\alpha = 0$ and $\alpha = 1/2$ correspond to the logarithmic and square root metrics in Dryden et al. (2009), while $\alpha = 1$ is the ordinary Frobenius metric.

Assuming $R = \int_0^1 H_\alpha\{\Sigma(t)\}\, dt$ exists and $\Sigma(t) \in \mathcal{P}$ $(0 \leq t \leq 1)$, for a given value α we can compute the Fréchet integral

$$S_\alpha = \operatorname*{arginf}_{\Gamma \in \mathcal{P}} \int_0^1 d_\alpha\left\{\Sigma(t), \Gamma\right\}^2 \, \mathrm{d}t \tag{5}$$

analytically. If $\alpha > 0$, the eigenvalues of $H_\alpha\{\Sigma(t)\}$ are all above $-\alpha^{-1}$ which implies that the same is true for R. Hence, $H_\alpha^{-1}(R)$ is well-defined for any $\alpha \geq 0$. Properties of the Frobenius metric imply that

$$\int_0^1 d_\alpha\left\{\Sigma(t), \Gamma\right\}^2 \, \mathrm{d}t = \int_0^1 d_\alpha\left\{\Sigma(t), H_\alpha^{-1}(R)\right\}^2 \, \mathrm{d}t + d_\alpha\left\{\Gamma, H_\alpha^{-1}(R)\right\}^2,$$

whence

$$S_\alpha = H_\alpha^{-1}(R) = H_\alpha^{-1}\left[\int_0^1 H_\alpha\{\Sigma(t)\}\, \mathrm{d}t\right]. \tag{6}$$

2·4. *Adaptive Metric Selection*

The proposed functional covariance matrix through Fréchet integration depends on the chosen geometry as represented by the transformation parameter α. For practical utility of this approach, a method for the selection of α is needed. For this we propose a criterion-based approach, where the user specifies a target criterion M that measures the quality of the metric induced by the parameter α. The metric can then be selected as

$$\alpha^* = \operatorname*{argmax}_{\alpha \in \mathcal{A}} \; M(\alpha), \tag{7}$$

where $\mathcal{A}$ is an interval of the form $[0, T]$, for some $T > 0$.

As a concrete example, consider the common use of the covariance matrix for dimensionality reduction through principal component analysis. Then a desirable feature for a covariance matrix is that its eigenvalues decay rapidly, so that the first few eigenvectors provide a good approximation. One suitable criterion function which measures this decay rate is motivated by the

cumulative scree plot. Let $C \in \mathcal{P}$ and define $\lambda_{C,j}$ to be the jth largest eigenvalue of C. For values of k between 1 and p, one plots the pairs $(k,\ \sum_{j=1}^{k} \lambda_{C,j} / \sum_{j=1}^{p} \lambda_{C,j})$, reflecting the fraction of variance explained by the first k components. If the eigenvalues decay rapidly, these cumulative sums will increase quickly for small k, with only small increases for large k. The speed of the increase in the cumulative sums can then be measured by

$$m_{\mathrm{FVE}}(C) = \sum_{k=1}^{p} \left(\frac{\sum_{j=1}^{k} \lambda_{C,j}}{\sum_{j=1}^{p} \lambda_{C,j}} \right) = \frac{\sum_{j=1}^{p} (p - j + 1) \lambda_{C,j}}{\sum_{j=1}^{p} \lambda_{C,j}}. \tag{8}$$

The criterion function can be set as $M(\alpha) = m_{\mathrm{FVE}}(S_\alpha)$. As more weight is given to larger eigenvalues, functional covariance matrices S_α with faster increasing cumulative sums of eigenvalues will lead to larger values of $M(\alpha)$, which are then harnessed to choose a best metric.

A second option is motivated by observing that the undesirable extreme is that all of the eigenvalues are equal. Let $\nu, \lambda_C \in \mathcal{R}^p$ be the vector of all ones and the vector of eigenvalues of $C \in \mathcal{P}$, respectively. The metric can be chosen so as to maximize the discrepancy between ν and λ_C. While accounting for scaling differences, the geodesic distance between ν and λ_C is

$$m_{\mathrm{G}}(C) = \arccos \left(\frac{\langle \lambda_C, \nu \rangle_{\mathrm{E}}}{\|\lambda_C\|_{\mathrm{E}} p^{1/2}} \right), \tag{9}$$

where $\langle \cdot, \cdot \rangle_{\mathrm{E}}$ and $\|\cdot\|_{\mathrm{E}}$ are the standard Euclidean inner product and norm, respectively. This criterion measures the angle between these vectors, so that large values of $m_{\mathrm{G}}(C)$ indicate large discrepancies between λ_C and ν. Hence, another criterion of interest is $M(\alpha) = m_{\mathrm{G}}(S_\alpha)$.

Beyond dimensionality reduction, criterion-based metric selection can be used quite generally. For example, in Section 4·2 we demonstrate how metric selection can be utilized to great advantage in a two sample setting to distinguish two populations. Since many such criterion functions are unobservable, we address the estimation of M in the next section.

2·5. *Estimation*

The main quantity requiring estimation is the functional covariance matrix S_α in (6). As a first step, cross-covariance estimators $\hat{\Sigma}(t)$ are computed; such estimators exist for functional data observed on dense or sparse grids, with and without noise (Yang et al., 2011). The final functional covariance estimator is

$$\hat{S}_\alpha = H_\alpha^{-1} \left[\int_0^1 H_\alpha \{ \hat{\Sigma}(t) \} \, \mathrm{d}t \right]. \tag{10}$$

Convergence of $\hat{S}_\alpha$ to the target S_α naturally depends on the convergence properties of the plug-in estimators $\hat{\Sigma}_{kl}(t)$ as well as properties of the Box–Cox transformations H_α and will be addressed in Section 3. As an illustrative example, assume one observes a sample of n realizations $(X_{i1}, \ldots, X_{ip})$ $(i = 1, \ldots, n)$ of the $p-$dimensional process $(X_1, \ldots, X_p)$ across the entire continuum $\mathcal{T}$. The empirical cross-sectional covariance matrices for each fixed time t are

$$\hat{\Sigma}_{kl}(t) = \frac{1}{n} \sum_{i=1}^{n} \{ X_{ik}(t) - \hat{\mu}_k(t) \} \{ X_{il}(t) - \hat{\mu}_l(t) \} \quad (k, l = 1, \ldots, p), \tag{11}$$

where $\hat{\mu}_k(t)$ are the sample means of $X_{ik}(t)$.

Additionally, the implementation of data-adaptive metric selection as outlined in Section 2·4 requires estimation if the criterion function M is not observable. This is notably the case for the criterion functions based on the fraction of variance explained quantity in (8) and geodesic

distance in (9). Given an estimator $\hat{M}$ of M, the obvious estimator of the best metric α^* in (7) is

$$\hat{\alpha}^* = \underset{\alpha \in \mathcal{A}}{\operatorname{argmax}} \ \hat{M}(\alpha). \tag{12}$$

For the specific criterion functions suggested for dimensionality reduction in Section 2·4, a simple estimator is obtained by substituting $\hat{S}_\alpha$ for S_α, i.e. $\hat{M}(\alpha) = m_{\mathrm{FVE}}(\hat{S}_\alpha)$ for the fraction of variance explained criterion or $\hat{M}(\alpha) = m_{\mathrm{G}}(\hat{S}_\alpha)$ for the geodesic criterion.

3. THEORY

In Section 2·5, the plug-in estimator in (10) was proposed to estimate S_α in (6). In this section, we study the convergence of this estimator by examining the random quantity $d_{\mathrm{F}}(\hat{S}_\alpha, S_\alpha)$ for a fixed α. The main question is to what extent the metric affects the rate of convergence. In addition, we investigate the effect of metric selection. Specifically, we consider α^* and $\hat{\alpha}^*$ in (7) and (12) for a generic criterion function M and the convergence of $\hat{S}_{\hat{\alpha}^*}$, the estimated covariance matrix with data-adaptive metric selection, to the target S_{α^*}.

A key preliminary finding is Proposition 1 in Appendix 1. This proposition demonstrates that d_{F} and d_α are equivalent metrics on local neighborhoods. Utilizing this relationship between the metrics leads to the rate of convergence of our estimator $\hat{S}_\alpha$ under the following assumptions.

Assumption 1. For all t, the pointwise covariance matrices in (1) satisfy $\Sigma(t) \in \mathcal{P}$ and the number of distinct eigenvalues of $\Sigma(t)$ is p. Also, the function $\Sigma : [0,1] \to \mathcal{P}$ is continuous in the metric d_{F}.

Assumption 2. The estimators $\hat{\Sigma}_{kl}(t)$ $(k, l = 1, \ldots, p)$ of $\Sigma_{kl}(t) = \operatorname{cov}\{X_k(t), X_l(t)\}$ satisfy

$$\sup_{0 \le t \le 1} \left| \hat{\Sigma}_{kl}(t) - \Sigma_{kl}(t) \right| = o_p(1), \quad d_{L^2}\left(\hat{\Sigma}_{kl}, \Sigma_{kl}\right) = O_p(r_n),$$

where $r_n \to 0$ as $n \to \infty$ and $d_{L^2}(f,g)^2 = \int_{[0,1]} \{f(t) - g(t)\}^2 \, \mathrm{d}t$.

Assumption 1 implies that the integral $R = \int_0^1 H_\alpha\{\Sigma(t)\} \, \mathrm{d}t$ exists, so that (6) is the solution of (5). It also ensures nice behavior of the eigenvalues of $\Sigma(t)$ as t varies. Assumption 2 is the key factor for the convergence rates in Theorems 1 and 2 below. The rate r_n will depend on whether the functional data are observed with or without noise and on a sparse or dense grid. For the scenario of fully observed functional data as in Section 2·5, with cross-covariance estimator as in (11), Lemma 1 in Appendix 3 shows that we may take $r_n = n^{-1/2}$.

THEOREM 1. *For a fixed value α, suppose Assumptions* 1 *and* 2 *hold. Then the functional covariance estimator (*10*) satisfies $d_{\mathrm{F}}(\hat{S}_\alpha, S_\alpha) = O_p(r_n)$.*

Hence, the proposed methodology preserves the L^2 convergence rate of the cross-covariance estimators.

To provide theoretical support for data-adaptive metric selection as described in Section 2·4, we next study the convergence of $d_{\mathrm{F}}(\hat{S}_{\hat{\alpha}^*}, S_{\alpha^*})$. Let $\mathcal{A} = [0, T]$ be as in Section 2·4, for an arbitrary T. The main idea is to use the inequality

$$d_{\mathrm{F}}(\hat{S}_{\hat{\alpha}^*}, S_{\alpha^*}) \le d_{\mathrm{F}}(S_{\alpha^*}, S_{\hat{\alpha}^*}) + \sup_{\alpha \in \mathcal{A}} d_{\mathrm{F}}(\hat{S}_\alpha, S_\alpha). \tag{13}$$

The first term on the right hand side can be controlled as described in Proposition 2 in Appendix 1, by examining the effects of the transformation H_α as α varies, and making use of the following additional assumptions.

Assumption 3. For any $\alpha \in \mathcal{A}$, the matrix S_α has p distinct eigenvalues.

Assumption 4. For a sequence $s_n \to 0$, the estimator $\hat{M}$ satisfies $\sup_{\alpha\in\mathcal{A}} \left|M(\alpha) - \hat{M}(\alpha)\right| = O_p(s_n)$.

Assumption 5. The target criterion function M has the following properties: the maximizer α^* in (7) is unique and, for any open $G \subset \mathcal{A}$ containing α^*, $\sup_{\alpha\in G^c} M(\alpha) < M(\alpha^*)$. Additionally there exist η, b, $\beta > 0$ such that $|\alpha^* - \alpha| < \eta$ implies $M(\alpha^*) - M(\alpha) \geq b|\alpha - \alpha^*|^\beta$.

Assumption 3 controls the spacings of the eigenvalues of S_α as α varies. Indeed, perusing the proof of Theorem 1 and combining Assumption 3 with the results of Proposition 2 yields $\sup_{\alpha\in\mathcal{A}} d_{\mathrm{F}}(S_\alpha, \hat{S}_\alpha) = O_p(r_n)$. Thus, the second term on the right hand side of (13) is $O_p(r_n)$, while Assumptions 4 and 5 determine the rate of the second term. In the case of a general criterion M, the parameter β is difficult to specify, but it can be estimated numerically from the data by using the estimator $\hat{M}$ as a proxy for M. However, if M is known to be twice continuously differentiable, Taylor's theorem implies that $\beta = 2$ is permissible. For the specific examples of the fraction of variance explained and geodesic criteria, based on (8) and (9), Proposition 3 in Appendix 1 demonstrates that we may take $s_n = r_n$ in Assumption 4 and $\beta = 2$ in Assumption 5. Our main result for metric selection is

THEOREM 2. *Under Assumptions* 1–5, $d_{\mathrm{F}}(S_{\alpha^*}, \hat{S}_{\hat{\alpha}^*}) = O_p\left(r_n + s_n^{1/\beta}\right)$.

This result provides the rate of convergence of the estimator with adaptive metric selection. The rate of convergence depends on the precise nature of the criterion function M, its estimator $\hat{M}$ and also on the L^2 convergence of the plug-in estimators.

Although we have presented our results so far in the framework of the Box–Cox transformation family for symmetric positive definite matrices, these results and the underlying approach and theory are not limited to this class of transformations. While the Box–Cox class is an immediately applicable transformation family, other families of invertible transformations, indexed by a single parameter α, may also be of interest, and the restriction to this class in the previous descriptions and discussions was merely due to ease of presentation. In a more general approach, we only need to require that the transformation family under consideration satisfies certain structural assumptions that are laid out at the beginning of Appendix 2. Specifically, under these assumptions on the transformation family, Theorems 1 and 2 will remain valid. Indeed, the proofs of the main results in Appendix 2 cover this more general case.

Selecting a specific transformation family leads to an associated family of metrics, similar to (4), for which the Fréchet integral of $\Sigma(t)$ can then be defined. As a general example of a family of transformations, consider a parameterized family of monotone univariate functions $h_\alpha : \mathcal{R}^+ \to \mathcal{R}$. This includes the Box–Cox family as a special case. Another special case is the exponential transformation family $h_\alpha(x) = e^{\alpha x}$. Once the family h_α has been specified, for a diagonal $p \times p$ matrix D, define $H_\alpha(D) = \mathrm{diag}\{h_\alpha(D_{11}), \ldots, h_\alpha(D_{pp})\}$. The transformation can then be extended to any $A \in \mathcal{P}$ by writing $A = UDU^T$, with D diagonal, and defining $H_\alpha(A) = UH_\alpha(D)U^T$.

4. SIMULATIONS AND APPLICATION TO RESTING STATE BRAIN CONNECTIVITY

4·1. *Simulations*

A simulation study was conducted to demonstrate the benefits of data-adaptive metric selection, as outlined in Section 2·4. Specifically, we implemented metric selection for dimensionality reduction by utilizing the fraction of variance explained criterion m_{FVE} in (8) and compared the

resulting eigenvalue decay rate with that resulting from the Frobenius metric that corresponds to the choice $\alpha = 1$ in (3)–(6). We also based data-adaptive metric selection on the geodesic criterion m_{G} in (9). The results were virtually indistinguishable from those using the fraction of variance explained criterion. Hence, for simplicity, we present results only for the latter.

Multivariate functional data of dimension $p = 5$ were generated from a multivariate stochastic process $X = \{X_1(t), \ldots, X_5(t) : 0 \leq t \leq 1\}$. The individual component functions $X_1, \ldots, X_5$ were generated according to the model

$$X_j(t) = \mu_j(t) + \sum_{k=1}^{10} \xi_{jk}(t)\psi_k(t) \quad (0 \leq t \leq 1;\ j = 1, \ldots 5), \tag{14}$$

where μ_j is the Gaussian density function with mean and variance j and the ψ_k are Fourier basis functions $\psi_k(t) = \sin(\pi k t)$, k even, and $\psi_k(t) = \cos\{\pi(k-1)t\}$, k odd.

The random elements in the simulation are the coefficients ξ_{jk}. Each of the five component functions of the multivariate stochastic process X has 10 such coefficients, thus requiring the generation of 50 random numbers for each realization of X. This was done by sampling from a $\mathcal{N}_{50}(0, C)$ distribution, with C constructed by forming a 50×50 diagonal matrix D with diagonal entries $D_{jj} = 1.03e^{-j/1.03}$ and computing $C = VDV^T$, for an orthonormal matrix V. The dependence between distinct coefficients ensures that the component functions $X_1, \ldots, X_5$ are indeed correlated. For each of 1000 simulation runs, a random sample of coefficients of size $n = 200$ was generated from the $\mathcal{N}_{50}(0, C)$ distribution, yielding a sample of size n of 5-dimensional functional data by inserting these coefficients into (14).

For $\mathcal{A} = [0, 4]$ and $\alpha \in \mathcal{A}$, denote by $\hat{S}_{\alpha l}$ the estimated functional covariance matrix of X using the data from the lth simulation run. The data-adaptive metric selection was then performed by utilizing the criterion function m_{FVE} in (8) and computing

$$\hat{\alpha}_l^* = \underset{\alpha \in \mathcal{A}}{\operatorname{argmax}}\ m_{\mathrm{FVE}}\left(\hat{S}_{\alpha l}\right), \tag{15}$$

as outlined in Section 2·4. The functional covariance that resulted from Fréchet integration using $\alpha = \hat{\alpha}_l^*$ was then compared with the covariance generated by the default Frobenius metric, for which $\alpha = 1$, by examining the decay of their respective eigenvalues, as quantified by the fraction of variance explained by the first k eigenvalues, $k = 1, \ldots, 5$; see (8).

Figure 1 gives the cumulative scree plots obtained by averaging over all simulations. It indicates that the data-adaptive metric selection technique produces a functional covariance matrix with eigenvalues that decay more rapidly than that produced by the Frobenius metric, so covariance matrices produced by data-adaptive metric selection can be approximated with greater accuracy by a small number of eigenvalues/eigenvectors.

To compare the chosen values $\hat{\alpha}_l^*$ in (15) with the theoretical maximizer $\alpha^* = \operatorname{argmax}_{\alpha \in \mathcal{A}}\, m_{\mathrm{FVE}}(S_\alpha)$, we computed the latter numerically. On a grid of 401 equally spaced points on $[0, 4]$ the theoretical maximizer was found to be $\alpha^* = 0.04$. This compares well with the simulation results where, for 1000 simulations, the estimates $\hat{\alpha}_l^*$ had a mean of 0.042 and a variance of 1.69×10^{-4}, with estimates ranging between 0.01 and 0.09.

4·2. *Functional Connectivity in the Resting State Brain*

In recent years, the problem of identifying functional connectivity between brain voxels or regions has received a great deal of attention, especially for resting state fMRI (Allen et al., 2014; Ferraty et al., 2013; Sheline & Raichle, 2013). Subjects are asked to relax while undergoing a fMRI brain scan, where blood-oxygen-level dependent signals are recorded and then processed to yield voxel-specific time courses of signal strength. The connectivity between brain regions

is usually quantified by a measure of correlation between the corresponding time courses of an fMRI scan. Then the primary source for the analysis of functional connectivity in the brain is a covariance matrix. Each time course of a subject's scan provides a functional datum and the multivariate aspect arises as multiple voxels or regions are considered. The estimation of covariances of these multivariate data for groups of subjects that are characterized by covariates such as disease status is of great interest.

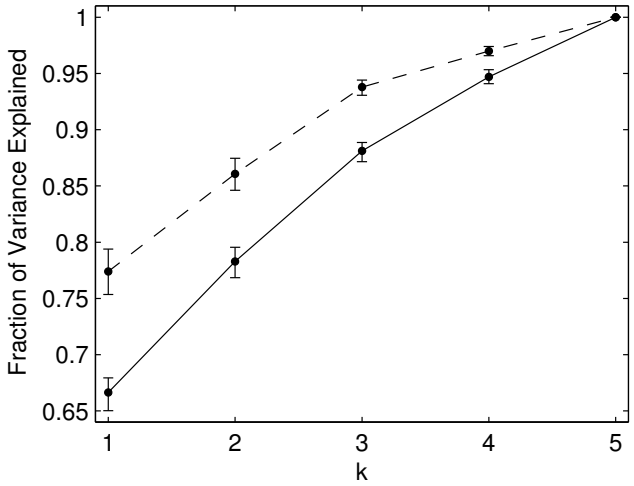


Fig. 1: Average cumulative scree plots for estimated functional covariance matrices of 5-dimensional functional data generated by the model in (14), for 1000 simulation runs. The solid line corresponds to the default choice $\alpha = 1$ while the dashed line corresponds to the data-adaptive metric selection using the criterion in (8). Vertical bars indicate the range of the cumulative scree plots across all simulations.

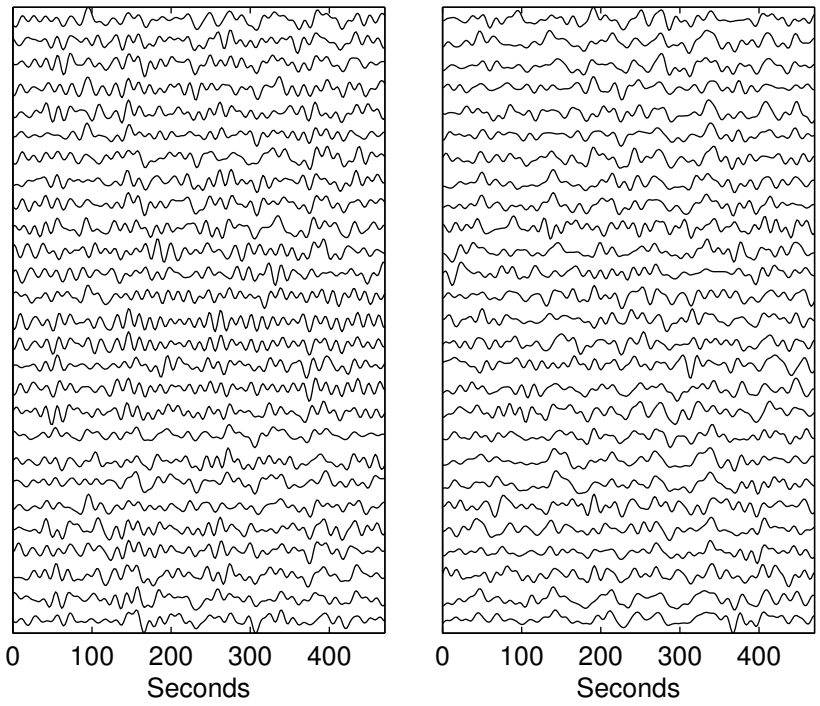
The data to be analyzed come from a study of 239 elderly individuals, each of whom underwent an fMRI scan at the UC Davis Imaging Research Center. Each subject received a clinical evaluation which resulted in a diagnosis of cognitive state. Of the 239 patients, 173 were clinically diagnosed as normal and 66 as Alzheimer's. Summary statistics of age within the two groups are in Table 1, indicating only minor differences in age distribution between normal and Alzheimer's subjects. This is an important consideration, as connectivity is known to decline naturally with age.

Table 1: Summary statistics of distribution of age within normal and Alzheimer's groups

Group	Mean	SD	Minimum	Maximum
Normal	77.01	6.62	64	94
Alzheimer's	78.74	9.04	51	93

SD, standard deviation.

Preprocessing of the recorded blood oxygenation-level-dependent signals was implemented by adopting the standard procedures of slice-timing correction, head motion correction and coregistration to the subject's MRI scan. Multiple linear regression was applied to the signal at each voxel to remove global linear trends and signal drift, along with two other global signals corresponding to cerebral spinal fluid and white matter. Finally, each signal was band-pass filtered to preserve frequency components between 0.01 and 0.08 Hz.



(a) Normal Subject (b) Alzheimer's Subject

Fig. 2: Time courses of fMRI signals, standardized to have the same vertical range, for 27 spatial locations within the posterior cingulate/precuneus region, where the courses are offset vertically. The left panel displays the signals of a cognitively normal subject and the right panel those of a subject diagnosed with Alzheimer's disease.

For the connectivity analysis, we compared normal and Alzheimer's subjects and restricted our attention to the posterior cingulate/precuneus region of the brain, which has been shown to be highly affected by Alzheimer's disease (Grady et al., 2001; Wang et al., 2006, 2007; Zhou et al., 2008b; Zhang et al., 2009; Bai et al., 2009). In particular, functional connectivity between voxels located in this region has been shown to be weaker for subjects with amnestic mild cognitive impairment, a common stage of transition between normal aging and Alzheimer's (Bai et al., 2008). For each subject, a common $3 \times 3 \times 3$ cube of voxels was identified and the corresponding fMRI signals extracted, thus giving multivariate functional data with $p = 27$. The coordinates of the center of this cube were taken from Table 3 of Buckner et al. (2009). Each of the 27 signals is a time course taken over the interval $[0, 470]$ seconds, with 236 measurements available at 2-second intervals. The fMRI time courses of a normal subject and one with Alzheimer's disease are shown in Figure 2.

Denoting by $\hat{S}_\alpha^N$ and $\hat{S}_\alpha^A$, as in (10), the estimated functional covariance matrices for the normal and Alzheimer's groups, respectively, one would expect a comparison of the corresponding correlation matrices, $\hat{C}_\alpha^N$ and $\hat{C}_\alpha^A$, to reflect differences between normal and Alzheimer's subjects. The estimated correlations for the default choice $\alpha = 1$ are shown in Figure 3. Both groups demonstrate strong functional connectivity within this region, as there are many large positive correlations. However, there is no visually observable difference between the two groups.

We find that by adaptively choosing the metric parameter α to maximize the distance,

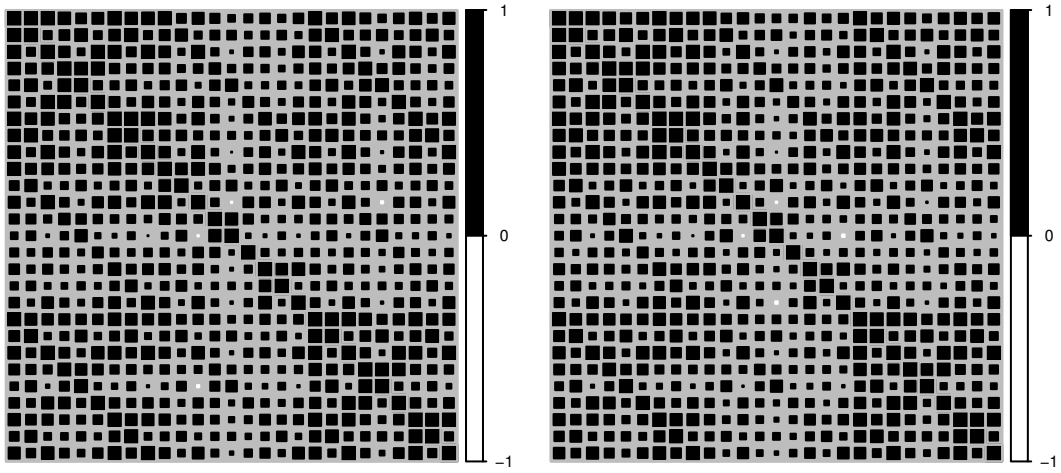
$$\hat{\alpha}^* = \underset{\alpha\in[0,4]}{\operatorname{argmax}}\, d_{\mathrm{F}}\left(\hat{C}_\alpha^N, \hat{C}_\alpha^A\right) = 2.8,$$

one obtains a more informative comparison between the groups. Specifically, using this value for α in (10), the correlations are seen to be sparser for the Alzheimer's group, in comparison to the normal subjects in the correlation plots in Figure 4. In order to visualize the distinction more clearly, the differences $\hat{C}_\alpha^N - \hat{C}_\alpha^A$ are depicted in Figure 5. The default value $\alpha = 1$ leads to the conclusion that, on average, functional connectivity within the posterior cingulate/precuneus region is similar between normal and Alzheimer's subjects, while the data-adaptive choice $\hat{\alpha}^* = 2.8$ reveals that Alzheimer's disease is associated with weaker connectivity in this brain region.

5. Discussion

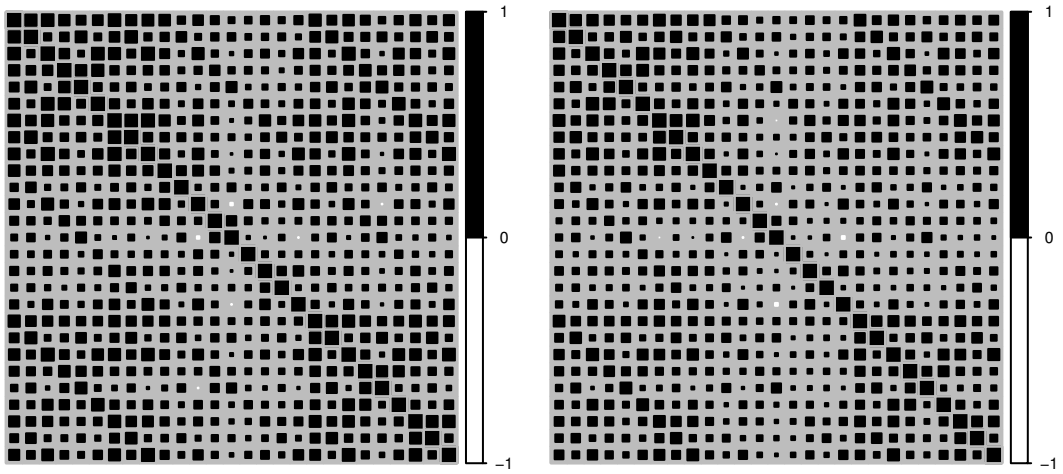
The proposed Fréchet integration of covariance matrices can also be applied in non-stationary multivariate time series analysis, when one aims at summary dependency measures. A main difference in the case of functional data is that, for each fixed time, the cross-sectional covariance matrix can be easily estimated from the data with $n^{1/2}$ convergence rate, while strong model assumptions are needed for consistency in nonlinear time series analysis.

For multivariate functional data, the transformation-based dependence measures are attractive because they are easy to compute, do not involve inverse operators as does canonical functional correlation, and can be interpreted as extensions of pointwise covariance measures, while other covariance measures such as dynamic correlation only focus on the functional aspects. For this reason, the proposed approach is expected to be useful for sparsely and densely sampled functional data, where consistent pointwise covariances can be obtained between all component pairs of the multivariate process.



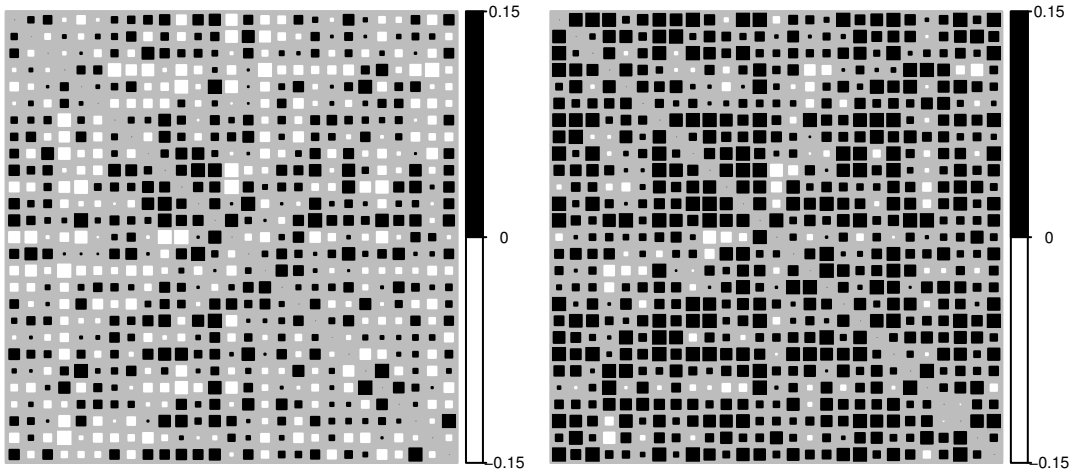
(a) Normal Group (b) Alzheimer's Group

Fig. 3: Estimated functional correlation matrices $\hat{C}_\alpha^N$ and $\hat{C}_\alpha^A$ for $\alpha = 1$. Positive (negative) correlations are shown in black (white), with larger (smaller) squares corresponding to correlations closer to ± 1 (0).



(a) Normal Group (b) Alzheimer's Group

Fig. 4: Estimated functional correlation matrices $\hat{C}_\alpha^N$ and $\hat{C}_\alpha^A$ for the adaptive choice $\alpha = \hat{\alpha}^* = 2.8$. Positive (negative) correlations are shown in black (white), with larger (smaller) squares corresponding to correlations closer to ± 1 (0).



(a) $\alpha = 1$ (b) $\alpha = \hat{\alpha}^* = 2.8$

Fig. 5: Differences between correlation matrices, $\hat{C}_\alpha^N - \hat{C}_\alpha^A$, for default (left) and adaptive (right) metric choices. Positive (negative differences) are indicated by black (white). Larger (smaller) squares correspond to values closer to 0.15 (-0.15).

Families of transformations other than the Box–Cox family that we have used in our simulations and the application could be of interest in some scenarios, and are also covered by our approach, as long as they form a family that is indexed by a single parameter α and satisfies

some basic regularity conditions. Our theory covers these more general families, which generate associated families of metrics.

The basic Fréchet integration approach can be applied to general metric spaces under regularity conditions, and is not limited to the space of covariance matrices. An example is time-indexed daily air pollution profiles where the time of peak pollution varies from day to day. If one is interested in a typical profile, then averaging the profiles over a range of calendar days, which is akin to construction a Riemann integral, will usually be inferior to a Fréchet integral or average when using a suitable metric. In this example, it might be sensible to adopt a Wasserstein-type distance (Villani, 2003) that takes into account time warping between profiles.

ACKNOWLEDGEMENTS

We wish to thank the Associate Editor and several reviewers for helpful comments. This research was supported by the National Science Foundation.

APPENDIX 1

Notation

Recall that d_{F} denotes the Frobenius metric, d_α is defined in (4), d_{L^2} denotes the standard L^2 metric for functions and $\|\cdot\|_{\mathrm{E}}$ denotes the Euclidean norm for vectors. Let h_α denote the Box–Cox transformations as in (3) for the univariate case $p = 1$. Then $\mathrm{d}h_\alpha/\mathrm{d}x = x^{\alpha-1}$ and $\mathrm{d}h_\alpha^{-1}/\mathrm{d}x = \{h_\alpha^{-1}(x)\}^{1-\alpha}$.

Propositions

PROPOSITION 1. *Let $A_j \in \mathcal{P}$, with eigenvalues $\lambda_{j1} \geq \cdots \geq \lambda_{jp} > 0$ $(j = 1, 2)$, and $\alpha \geq 0$. Suppose A_1 has k $(k = 2, \ldots, p)$ distinct eigenvalues $d_1 > \cdots > d_k > 0$, and set $\delta = \min_{l=1,\ldots,k-1}(d_l - d_{l+1})$ and $\pi_\alpha = \min_{l=1,\ldots,k-1}\{h_\alpha(d_l) - h_\alpha(d_{l+1})\}$. Set $C > 1$ such that $C^{-1} \leq d_k/2 \leq 2d_1 \leq C$ and define*

$$L_1 = p^{1/2}\left[C^{\max(\alpha,1)} + 2^{3/2}p\delta^{-1}\max\{-h_\alpha(C^{-1}), h_\alpha(C)\}\right], \quad L_2 = p^{1/2}\left\{C^{\max(\alpha,1)} + 2^{3/2}Cp\pi_\alpha^{-1}\right\}.$$

Then $d_{\mathrm{F}}(A_1, A_2) < d_k/2$ implies

$$d_\alpha(A_1, A_2) \leq L_1 d_{\mathrm{F}}(A_1, A_2)$$

and $d_\alpha(A_1, A_2) < \min\{h_\alpha(d_k) - h_\alpha(d_k/2), h_\alpha(2d_1) - h_\alpha(d_1)\}$ implies

$$d_{\mathrm{F}}(A_1, A_2) \leq L_2 d_\alpha(A_1, A_2).$$

Proof. Take the eigendecompositions $A_i = U_i\Lambda_i U_i^T$, with the diagonal elements of Λ_i being in decreasing order, and set $J_l = \{j : \lambda_{1j} = d_l\}$ $(l = 1, \ldots, k)$. Let $u_{i,j}$ denote the jth column of U_i. Then

$$\begin{aligned} d_\alpha(A_1, A_2) &\leq d_{\mathrm{F}}\{H_\alpha(\Lambda_1), H_\alpha(\Lambda_2)\} + d_{\mathrm{F}}\{U_1 H_\alpha(\Lambda_1)U_1^T, U_2 H_\alpha(\Lambda_1)U_2^T\} \\ &\leq \left[\sum_{j=1}^{p}\{h_\alpha(\lambda_{1j}) - h_\alpha(\lambda_{2j})\}^2\right]^{1/2} + \sum_{l=1}^{k}\left\{|h_\alpha(d_l)|\, d_{\mathrm{F}}\left(\sum_{j\in J_l} u_{1,j}u_{1,j}^T, \sum_{j\in J_l} u_{2j}u_{2,j}^T\right)\right\}. \end{aligned} \tag{A1}$$

Lemma 4.3 of Bosq (2000) implies that $\max_{j=1,\ldots,p}|\lambda_{1j} - \lambda_{2j}| \leq d_{\mathrm{F}}(A_1, A_2)$. By the mean value theorem, the summands of the first term in the last line of (A1) can be bounded by $\tau_j^{2(\alpha-1)} d_{\mathrm{F}}(A_1, A_2)^2$, where τ_j lies between λ_{1j} and λ_{2j}. Define $\Pi_l(v) = \sum_{j\in J_l}\langle u_{1,j}, v\rangle u_{1,j}$. Lemma 4.3 of Bosq (2000) implies that, for $j \in J_l$,

$$\|u_{2,j} - \Pi_l(u_{2,j})\|_{\mathrm{E}}^2 = 1 - \sum_{j'\in J_l}\langle u_{1,j'}, u_{2,j}\rangle^2 \leq 4\delta^{-2} d_{\mathrm{F}}(A_1, A_2)^2.$$

It follows that

$$d_{\mathrm{F}}\left(\sum_{j\in J_l} u_{1,j}u_{1,j}^T, \sum_{j\in J_l} u_{2,j}u_{2,j}^T\right)^2 = 2\sum_{j\in J_l}\|u_{2,j} - \Pi_l(u_{2,j})\|_{\mathrm{E}}^2 \leq 8p\delta^{-2}d_{\mathrm{F}}(A_1,A_2)^2.$$

Combining these bounds leads to

$$d_\alpha(A_1,A_2) \leq \left[\left\{\sum_{j=1}^p \tau_j^{2(\alpha-1)}\right\}^{1/2} + 2^{3/2}p^{1/2}\delta^{-1}\sum_{l=1}^k |h_\alpha(d_l)|\right] d_{\mathrm{F}}(A_1,A_2).$$

When $d_{\mathrm{F}}(A_1,A_2) < d_k/2$, Lemma 4.3 of Bosq (2000) implies $C_1^{-1} \leq \tau_j \leq C \ \ (j=1,\ldots,p)$. Hence, $\tau_j^{\alpha-1} \leq C^{\max(\alpha,1)}$. Furthermore, h_α is monotone for all $\alpha \geq 0$, so that $|h_\alpha(d_l)| \leq \max\{-h_\alpha(C^{-1}), h_\alpha(C)\}$. This gives the first inequality in the proposition, and the second can be obtained similarly. □

PROPOSITION 2. *Let A_1, A_2, π_α and C be as in Proposition* 1 *and let $0 \leq \alpha,\ \beta \leq T$, $\alpha \neq \beta$ and $T > 0$ arbitrary. There exists $L_3 = L_3(C)$ such that*

$$d_{\mathrm{F}}\{H_\alpha(A_1), H_\beta(A_1)\} \leq L_3|\alpha-\beta|.$$

Additionally, there exist $L_4 = L_4(C)$ and $L_5 = L_5(C,\alpha)$ such that, for $d_{\mathrm{F}}\{H_\alpha(A_1), H_\beta(A_2)\}$ small,

$$d_{\mathrm{F}}(A_1,A_2) \leq L_4|\alpha-\beta| + L_5 d_{\mathrm{F}}\{H_\alpha(A_1), H_\beta(A_2)\}.$$

Proof. For $\kappa,\ x \geq 0$, define $f(x,\kappa) = h_\kappa(x)$. Observe that $\partial f/\partial\kappa = \kappa^{-2}\left\{(\log x^\kappa - 1)x^\kappa + 1\right\}$ and set

$$\tilde{f}(x,\kappa) = \begin{cases} \frac{\partial f}{\partial \kappa}, & \kappa > 0, \\ \frac{1}{2}(\log x)^2, & \kappa = 0. \end{cases}$$

Then $\tilde{f}$ is continuous. Lemma 4.3 of Bosq (2000) implies that

$$d_{\mathrm{F}}\{H_\alpha(A_1), H_\beta(A_1)\}^2 = \sum_{j=1}^p \{h_\alpha(\lambda_{1j}) - h_\beta(\lambda_{1j})\}^2 \leq \left[\sum_{j=1}^p \left\{\tilde{f}(\lambda_{1j},\gamma_j)\right\}^2\right]^{1/2} |\alpha-\beta|,$$

for some values γ_j between α and β, so we may take

$$L_3 = p^{1/2} \sup_{C^{-1}\leq x\leq C} \sup_{0\leq\kappa\leq T} |\tilde{f}(x,\kappa)| < \infty.$$

Hence, the first inequality of the proposition holds since $C^{-1} \leq \lambda_{1j} \leq C\ (j=1,\ldots,p)$.

For the second inequality, take the eigendecompositions $A_j = U_j\Lambda_j U_j^T$ and set $g(x,\kappa) = h_\kappa^{-1}(x)$, for $\kappa \geq 0$. Then, by similar arguments as those in the proof of proposition 1,

$$d_{\mathrm{F}}(A_1,A_2) \leq \left(\sum_{j=1}^p [g\{h_\alpha(\lambda_{1j}),\alpha\} - g\{h_\beta(\lambda_{2j}),\beta\}]^2\right)^{1/2} + (2p)^{3/2}\pi_\alpha^{-1}C d_{\mathrm{F}}\{H_\alpha(A_1), H_\beta(A_2)\}.$$

The mean value theorem yields

$$|g\{h_\alpha(\lambda_{1j}),\alpha\} - g\{h_\beta(\lambda_{2j}),\beta\}| \leq \|\nabla g(\tau_j,\gamma_j)\|_{\mathrm{E}}\left\{|h_\alpha(\lambda_{1j}) - h_\beta(\lambda_{2j})| + |\alpha-\beta|\right\}$$

for some γ_j between α and β and τ_j between $h_\alpha(\lambda_{1j})$ and $h_\beta(\lambda_{2j})$. For $\kappa > 0$ and $x > -\kappa^{-1}$,

$$\frac{\partial g}{\partial x} = g(x,\kappa)^{1-\kappa}, \quad \frac{\partial g}{\partial \kappa} = \kappa^{-1}g(x,\kappa)\left\{\frac{x}{\kappa x+1} - \log g(x,\kappa)\right\},$$

so we can define

$$\tilde{g}(x,\kappa) = \begin{cases} \nabla g(x,\kappa), & \kappa > 0,\ x > -\kappa^{-1}, \\ -\frac{x^2 e^x}{2}, & \kappa = 0,\ x \in \mathcal{R}, \end{cases}$$

which is continuous on its domain. Hence

$$L_4 = p^{1/2} \sup_{C^{-1} \le y \le C} \sup_{0 \le \kappa \le T} \|\tilde{g}\{h_\kappa(y), \kappa\}\|_{\mathrm{E}}, \quad L_5 = L_4 + (2p)^{3/2} \pi_\alpha^{-1} C$$

are both finite. So, if $d_{\mathrm{F}}\{H_\alpha(A_1), H_\beta(A_2)\} < \lambda_{1p}/2$, the second inequality of the proposition holds. □

PROPOSITION 3. *Let m be either of the functions m_{FVE} or m_G in (8) and (9), set $\mathcal{A} = [0,T]$ for arbitrary $T > 0$ and suppose Assumptions* 1–3 *hold, as well as Assumptions* 6–8 *from Appendix 2. For $M(\alpha) = m(S_\alpha)$, we may take $\beta = 2$ in Assumption* 5*. Furthermore, with $\hat{M}(\alpha) = m(\hat{S}_\alpha)$, we may take $s_n = r_n$ in Assumption* 4*, where r_n is the rate in Assumption* 2*.*

Proof. Let $\lambda_{\alpha,j}$ and $\hat{\lambda}_{\alpha,j}$ be the jth largest eigenvalues of S_α and $\hat{S}_\alpha$, respectively, for $\alpha \in \mathcal{A}$. Using Assumption 1 along with the fact that eigenvalues and eigenvectors are infinitely differentiable with respect to the matrix elements (Magnus, 1985), one finds that $\mathrm{d}^2\lambda_{\alpha,j}/(\mathrm{d}\alpha)^2$ exists and is continuous on the interior of $\mathcal{A}$. Using the chain rule, one then finds that m_{FVE} and m_{G} are both twice continuously differentiable. By Assumption 5, $\mathrm{d}^2M/(\mathrm{d}\alpha)^2$ is negative at $\alpha = \alpha^*$, so that $\beta = 2$ follows from Taylor's theorem.

To show that we may take $s_n = r_n$, consider first the fraction of variance explained criterion. Set $c_\alpha = \sum_{j=1}^p \lambda_{\alpha,j}$ and $\hat{c}_\alpha = \sum_{j=1}^p \hat{\lambda}_{\alpha,j}$. Using Lemma 4.3 of Bosq (2000), we have $\sup_{\alpha\in\mathcal{A}} |c_\alpha - \hat{c}_\alpha| \le p \sup_{\alpha\in\mathcal{A}} d_{\mathrm{F}}(S_\alpha, \hat{S}_\alpha)$, hence

$$|M(\alpha) - \hat{M}(\alpha)| \le \frac{p(p+1)}{2c_\alpha \hat{c}_\alpha}(p\lambda_{\alpha,1} + c_\alpha) d_{\mathrm{F}}(S_\alpha, \hat{S}_\alpha).$$

As in the proof of Theorem 2 in Appendix 2, the assumptions imply that $\sup_{\alpha\in\mathcal{A}} d_{\mathrm{F}}(S_\alpha, \hat{S}_\alpha) = O_p(r_n)$. Since $\inf_{\alpha\in\mathcal{A}} c_\alpha > 0$ and $\sup_{\alpha\in\mathcal{A}} c_\alpha < \infty$ are guaranteed by Assumptions 1 and 3, we have $\sup_{\alpha\in\mathcal{A}} |M(\alpha) - \hat{M}(\alpha)| = O_p(r_n)$. Similar arguments can be used to obtain the result for $m = m_{\mathrm{G}}$, the geodesic criterion. □

APPENDIX 2

General Assumptions

We consider here general classes of transformations H_α of covariance matrices that are indexed by a parameter α, of which the Box–Cox transformation class is just one example. In addition to Assumptions 1–5, we require three assumptions regarding structural properties of the general transformation class H_α under which the theory of adaptive metric selection can be developed.

Assumption 6. The matrix $R = \int_0^1 H_\alpha\{\Sigma(t)\}\,\mathrm{d}t$ exists and $H_\alpha^{-1}(R)$ is well-defined. In other words, $S_\alpha = H_\alpha^{-1}(R)$, where S_α is defined in (5).

Assumption 7. For $A_1,\ A_2 \in \mathcal{P}$ and fixed α, there exist positive constants $\varepsilon_j = \varepsilon_j(A_1,\alpha)$ and $L_j = L_j(A_1,\alpha)$ $(j = 1,2)$ such that $d_{\mathrm{F}}(A_1,A_2) < \varepsilon_1$ implies $d_\alpha(A_1,A_2) \le L_1 d_{\mathrm{F}}(A_1,A_2)$ and $d_\alpha(A_1,A_2) < \varepsilon_2$ implies $d_{\mathrm{F}}(A_1,A_2) \le L_2 d_\alpha(A_1,A_2)$. The constants ε_j and L_j vary continuously in the first argument with respect to d_{F} and can be taken uniformly in the second argument over $\alpha \in \mathcal{A}$.

Assumption 8. For $A_1,\ A_2 \in \mathcal{P}$ and $\alpha,\ \beta \in \mathcal{A},\ \alpha \ne \beta$, there exists $L_3 = L_3(A_1,\alpha)$ such that $d_{\mathrm{F}}\{H_\alpha(A_1), H_\beta(A_1)\} \le L_3|\alpha - \beta|$. Furthermore, there exists $\varepsilon = \varepsilon(A_1,\alpha) > 0$ and $L_j = L_j(A_1,\alpha) > 0$ $(j = 4,5)$ such that $d_{\mathrm{F}}\{H_\alpha(A_1), H_\beta(A_2)\} < \varepsilon$ implies $d_{\mathrm{F}}(A_1,A_2) \le L_4|\alpha - \beta| + L_5 d_{\mathrm{F}}\{H_\alpha(A_1), H_\beta(A_2)\}$. The constants L_j $(j = 3,4,5)$ are continuous in the first argument.

The Box–Cox class utilized throughout the main portion of this paper satisfies Assumption 6 provided Assumption 1 holds, while Assumptions 7 and 8 are shown to be satisfied by Propositions 1 and 2 in Appendix 1, for any interval $\mathcal{A} = [0, T]$, with $T > 0$ arbitrary, and $\alpha,\ \beta \in \mathcal{A}$.

Proof of Theorem 1

Proof. Define $\Gamma(t) = H_\alpha\{\Sigma(t)\}$ and $\hat{\Gamma}(t) = H_\alpha\{\hat{\Sigma}(t)\}$ and let

$$R_\alpha = \int_0^1 \Gamma(t)\ \mathrm{d}t, \quad \hat{R}_\alpha = \int_0^1 \hat{\Gamma}(t)\ \mathrm{d}t.$$

By Jensen's inequality,

$$d_{\mathrm{F}}(R_\alpha, \hat{R}_\alpha)^2 = \sum_{k,l=1}^{p} \left[\int_0^1 \{\Gamma_{kl}(t) - \hat{\Gamma}_{kl}(t)\}\ \mathrm{d}t\right]^2 \leq \int_0^1 \sum_{k,l=1}^{p} \left\{\Gamma_{kl}(t) - \hat{\Gamma}_{kl}(t)\right\}^2\ \mathrm{d}t = \int_0^1 d_\alpha\{\Sigma(t), \hat{\Sigma}(t)\}^2\ \mathrm{d}t.$$

Assumptions 1 and 7 imply that there exist $\varepsilon_1,\ L_1 > 0$, both independent of t, such that, on the event $G_1 = [\sup_{0\leq t\leq 1} d_{\mathrm{F}}\{\Sigma(t), \hat{\Sigma}(t)\} < \varepsilon_1]$,

$$d_\alpha\{\Sigma(t), \hat{\Sigma}(t)\} \leq L_1 d_{\mathrm{F}}\{\Sigma(t), \hat{\Sigma}(t)\}.$$

Hence, on G_1, $d_{\mathrm{F}}(R_\alpha, \hat{R}_\alpha) \leq L_1 \left[\int_0^1 d_{\mathrm{F}}\{\Sigma(t), \hat{\Sigma}(t)\}^2\ \mathrm{d}t\right]^{1/2} = L_1 O_p(r_n)$ by Assumption 2. This assumption also implies $\mathrm{pr}(G_1) \to 1$, so that $d_\alpha(S_\alpha, \hat{S}_\alpha) = d_{\mathrm{F}}(R_\alpha, \hat{R}_\alpha) = O_p(r_n)$.

Let $s_1 \geq \cdots \geq s_p$ be the eigenvalues of S_α. Assumption 7 yields $\varepsilon_2, L_2 > 0$ such that, on the event $G_2 = \{d_\alpha(S_\alpha, \hat{S}_\alpha) < \varepsilon_2\},\ d_{\mathrm{F}}(S_\alpha, \hat{S}_\alpha) \leq L_2 d_\alpha(S_\alpha, \hat{S}_\alpha)$. Then $\mathrm{pr}(G_1 \cap G_2) \to 1$ and, on $G_1 \cap G_2$, $d_{\mathrm{F}}(S_\alpha, \hat{S}_\alpha) \leq L_1 L_2 O_p(r_n)$, implying the result. □

Proof of Theorem 2

Proof. Following the logic of the proof of Theorem 1 and using Assumptions 3 and 7, we find that $\sup_{\alpha\in\mathcal{A}} d_{\mathrm{F}}(S_\alpha, \hat{S}_\alpha) = O_p(r_n)$. Next, Assumptions 4 and 5 together imply that, for large n and with high probability, $|\alpha^* - \hat{\alpha}^*| < \eta$. In this case, it is easy to show that

$$|\alpha^* - \hat{\alpha}^*| \leq b^{-1}\{M(\alpha^*) - M(\hat{\alpha}^*)\}^{1/\beta} \leq b^{-1}\{2 \sup_{\alpha\in\mathcal{A}} |M(\alpha) - \hat{M}(\alpha)|\}^{1/\beta} = O_p(s_n^{1/\beta}).$$

Now, using Assumptions 1 and 8, one obtains $L_3 > 0$, independent of t such that $d_{\mathrm{F}}\left[H_{\alpha^*}\{\Sigma(t)\}, H_{\hat{\alpha}^*}\{\Sigma(t)\}\right] \leq L_3|\alpha^* - \hat{\alpha}^*|$, hence $d_{\mathrm{F}}\{H_{\alpha^*}(S_{\alpha^*}), H_{\hat{\alpha}^*}(S_{\hat{\alpha}^*})\} = O_p(|\alpha^* - \hat{\alpha}^*|)$. Lastly, Assumption 8 also yields constants $L_4,\ L_5 > 0$ such that

$$d_{\mathrm{F}}(S_{\alpha^*}, S_{\hat{\alpha}^*}) \leq L_4|\alpha^* - \hat{\alpha}^*| + L_5 d_{\mathrm{F}}\{H_{\alpha^*}(S_{\alpha^*}), H_{\hat{\alpha}^*}(S_{\hat{\alpha}^*})\}$$

on an event G_3 for which $\mathrm{pr}(G_3) \to 1$. Hence, $d_{\mathrm{F}}(S_{\alpha^*}, S_{\hat{\alpha}^*}) = O_p(|\alpha^* - \hat{\alpha}^*|) = O_p(s_n^{1/\beta})$, so that

$$d_{\mathrm{F}}(S_{\alpha^*}, \hat{S}_{\hat{\alpha}^*}) \leq \sup_{\alpha\in\mathcal{A}} d_{\mathrm{F}}(S_\alpha, \hat{S}_\alpha) + d_{\mathrm{F}}(S_{\alpha^*}, S_{\hat{\alpha}^*}) = O_p(r_n + s_n^{1/\beta}).$$

APPENDIX 3

Auxiliary Lemma

LEMMA 1. *Let* $X = \{X_1(t), \ldots, X_p(t); 0 \leq t \leq 1\}$, *set* $Y_{jk}(t) = X_j(t)X_k(t)$ *and let* $X^{(i)}$ $(i = 1, \ldots, n)$ *be independent and identically distributed as* X. *Suppose* $|\mathrm{d}X_j/\mathrm{d}t| < \infty$ $(j = 1, \ldots, p)$ *almost surely and that the moments* $E\left(\mathrm{d}^l X_j/\mathrm{d}t\right)$ $(l = 1, \ldots 4)$, *are continuous in* t, $E\{\sup_t |\mathrm{d}X_j/\mathrm{d}t|\} < \infty$ *and* $E\{\sup_t |\mathrm{d}Y_{jk}/\mathrm{d}t|\} < \infty$ $(j, k = 1, \ldots, p)$. *Then the estimator*

$$\hat{\Sigma}_{jk}(t) = \frac{1}{n}\sum_{i=1}^{n} Y_{jk}^{(i)}(t) - \left\{\frac{1}{n}\sum_{i=1}^{n} X_j^{(i)}(t)\right\}\left\{\frac{1}{n}\sum_{i=1}^{n} X_k^{(i)}(t)\right\}$$

satisfies Assumption 2 *with* $r_n = n^{-1/2}$.

Proof. First, we consider the convergence of $\hat{\mu}_j(t) = n^{-1}\sum_{i=1}^n X_j(t)$ to the target $\mu_j(t) = E\{X_j(t)\}$. By Fubini's theorem and the assumption that $E\left\{X_j^2(t)\right\}$ is continuous,

$$E\left(\|\hat{\mu}_j - \mu_j\|_{L^2}^2\right) = \int_0^1 E\{\hat{\mu}_j(t) - \mu_j(t)\}^2 \,\mathrm{d}t = O(n^{-1}).$$

Hence, $\|\hat{\mu}_j - \mu_j\|_{L^2} = O_p(n^{-1/2})$. For the uniform convergence, we have $\hat{\mu}_j(t) - \mu_j(t) = o_p(1)$ for all t. Furthermore, Theorem 1.5.7 of Van der Vaart & Wellner (1996) shows that the process is tight since, for $\delta > 0$,

$$\sup_{|s-t|<\delta} |\hat{\mu}_j(s) - \hat{\mu}_j(t)| \leq \frac{\delta}{n}\sum_{i=1}^n \sup_t |\mathrm{d}X_j/\mathrm{d}t|$$

and $n^{-1}\sum_{i=1}^n \sup_t |\mathrm{d}X_j/\mathrm{d}t| = O_p(1)$. Thus, Theorems 1.5.4 and 1.3.6 of Van der Vaart & Wellner (1996) imply that $\sup_t |\hat{\mu}_j(t) - \mu_j(t)| = o_p(1)$. The result for the covariance estimator $\hat{\Sigma}(t)$ follows by applying a similar argument to the processes Y_{jk}. □

REFERENCES

ALLEN, E. A., DAMARAJU, E., PLIS, S. M., ERHARDT, E. B., EICHELE, T. & CALHOUN, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex* **24**, 663–676.

ANDO, T., LI, C.-K. & MATHIAS, R. (2004). Geometric means. *Linear Algebra and its Applications* **385**, 305–334.

ARSIGNY, V., FILLARD, P., PENNEC, X. & AYACHE, N. (2007). Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications* **29**, 328–347.

BAI, F., WATSON, D. R., YU, H., SHI, Y., YUAN, Y. & ZHANG, Z. (2009). Abnormal resting-state functional connectivity of posterior cingulate cortex in amnestic type mild cognitive impairment. *Brain Research* **1302**, 167–174.

BAI, F., ZHANG, Z., YU, H., SHI, Y., YUAN, Y., ZHU, W., ZHANG, X. & QIAN, Y. (2008). Default-mode network activity distinguishes amnestic type mild cognitive impairment from healthy aging: a combined structural and resting-state functional MRI study. *Neuroscience Letters* **438**, 111–115.

BAKER, C. R. (1974). Joint measures and cross-covariance operators. *Transactions of the American Mathematical Society* **186**, 273–289.

BERRENDERO, J., JUSTEL, A. & SVARC, M. (2011). Principal components for multivariate functional data. *Computational Statistics and Data Analysis* **55**, 2619–2634.

BOSQ, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. New York: Springer-Verlag.

BUCKNER, R. L., SEPULCRE, J., TALUKDAR, T., KRIENEN, F. M., LIU, H., HEDDEN, T., ANDREWS-HANNA, J. R., SPERLING, R. A. & JOHNSON, K. A. (2009). Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. *The Journal of Neuroscience* **29**, 1860–1873.

CHIOU, J.-M., CHEN, Y.-T. & YANG, Y.-F. (2014). Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica* **24**, 1571–1596.

CHIOU, J.-M. & LI, P.-L. (2007). Functional clustering and identifying substructures of longitudinal data. *Journal of the Royal Statistical Society: Series B* **69**, 679–699.

CLAESKENS, G., HUBERT, M., SLAETS, L. & VAKILI, K. (2014). Multivariate functional halfspace depth. *Journal of the American Statistical Association* **109**, 411–423.

DRYDEN, I. L., KOLOYDENKO, A. & ZHOU, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Annals of Applied Statistics* **3**, 1102–1123.

DUBIN, J. A. & MÜLLER, H.-G. (2005). Dynamical correlation for multivariate longitudinal data. *Journal of the American Statistical Association* **100**, 872–881.

FERRATY, F., GOIA, A., SALINELLI, E. & VIEU, P. (2013). Functional projection pursuit regression. *Test* **22**, 293–320.

FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. In *Annales de l'Institut Henri Poincaré*, vol. 10.

FRISTON, K. J. (2011). Functional and effective connectivity: a review. *Brain Connectivity* **1**, 13–36.

GRADY, C. L., FUREY, M. L., PIETRINI, P., HORWITZ, B. & RAPOPORT, S. I. (2001). Altered brain functional connectivity and impaired short-term memory in Alzheimer's disease. *Brain* **124**, 739–756.

GUALTIEROTTI, A. F. (1979). On cross-covariance operators. *SIAM Journal on Applied Mathematics* **37**, 325–329.

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

HE, G., MÜLLER, H.-G. & WANG, J.-L. (2004). Methods of canonical analysis for functional data. *Journal of Statistical Planning and Inference* **122**, 141–159.

LE, H. (2001). Locating Fréchet means with application to shape spaces. *Advances in Applied Probability* **33**, 324–338.

LEE, M., SMYSER, C. & SHIMONY, J. (2013). Resting-state fMRI: a review of methods and clinical applications. *American Journal of Neuroradiology* **34**, 1866–1872.

LEURGANS, S. E., MOYEED, R. A. & SILVERMAN, B. W. (1993). Canonical correlation analysis when the data are curves. *Journal of the Royal Statistical Society: Series B* **55**, 725–740.

MAGNUS, J. R. (1985). On differentiating eigenvalues and eigenvectors. *Econometric Theory* **1**, 179–191.

OPGEN-RHEIN, R. & STRIMMER, K. (2006). Inferring gene dependency networks from genomic longitudinal data: A functional data approach. *REVSTAT - Statistical Journal* **4**, 53–65.

PIGOLI, D., ASTON, J. A., DRYDEN, I. L. & SECCHI, P. (2014). Distances and inference for covariance operators. *Biometrika* **101**, 409–422.

SERBAN, N., STAICU, A.-M. & CARROLL, R. J. (2013). Multilevel cross-dependent binary longitudinal data. *Biometrics* **69**, 903–913.

SHELINE, Y. I. & RAICHLE, M. E. (2013). Resting state functional connectivity in preclinical Alzheimer's disease. *Biological Psychiatry* **74**, 340–347.

VALENCIA, D., LILLO, R. & ROMO, J. (2014). Spearman coefficient for functions. *In: Contributions in Infinite-dimensional Statistics and Related Topics, Ed. Bongiomo, E.G., Salinelli, E., Goia, A., Vieu, P., Societa Editrice Esculapio* , 269–272.

VAN DER VAART, A. & WELLNER, J. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

VILLANI, C. (2003). *Topics in Optimal Transportation*. American Mathematical Society.

WANG, K., LIANG, M., WANG, L., TIAN, L., ZHANG, X., LI, K. & JIANG, T. (2007). Altered functional connectivity in early Alzheimer's disease: A resting-state fMRI study. *Human Brain Mapping* **28**, 967–978.

WANG, L., ZANG, Y., HE, Y., LIANG, M., ZHANG, X., TIAN, L., WU, T., JIANG, T. & LI, K. (2006). Changes in hippocampal connectivity in the early stages of Alzheimer's disease: evidence from resting state fMRI. *Neuroimaging* **31**, 496–504.

YANG, W., MÜLLER, H.-G. & STADTMÜLLER, U. (2011). Functional singular component analysis. *Journal of the Royal Statistical Society: Series B* **73**, 303–324.

ZHANG, H.-Y., WANG, S.-J., XING, J., LIU, B., MA, Z.-L., YANG, M., ZHANG, Z.-J. & TENG, G.-J. (2009). Detection of PCC functional connectivity characteristics in resting-state fMRI in mild Alzheimers disease. *Behavioural Brain Research* **197**, 103–108.

ZHOU, L., HUANG, J. & CARROLL, R. (2008a). Joint modelling of paired sparse functional data using principal components. *Biometrika* **95**, 601–619.

ZHOU, Y., DOUGHERTY, J. H., HUBNER, K. F., BAI, B., CANNON, R. L. & HUTSON, R. K. (2008b). Abnormal connectivity in the posterior cingulate and hippocampus in early Alzheimer's disease and mild cognitive impairment. *Alzheimer's & Dementia* **4**, 265–270.