ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2020)

Functional models for time-varying random objects

Paromita Dubey and Hans-Georg Müller

University of California, Davis, USA

[*Read before* The Royal Statistical Society *at a meeting organized by the* Research Section *on Wednesday, October 16th, 2019*, Professor A. Doucet *in the Chair*]

Summary. Functional data analysis provides a popular toolbox of functional models for the analysis of samples of random functions that are real valued. In recent years, samples of time-varying object data such as time-varying networks that are not in a vector space have been increasingly collected. These data can be viewed as elements of a general metric space that lacks local or global linear structure and therefore common approaches that have been used with great success for the analysis of functional data, such as functional principal component analysis, cannot be applied. We propose *metric covariance*, a novel association measure for paired object data lying in a metric space (Ω, d) that we use to define a metric autocovariance function for a sample of random Ω-valued curves, where Ω generally will not have a vector space or manifold structure. The proposed metric autocovariance function is non-negative definite when the squared semimetric d^2 is of negative type. Then the eigenfunctions of the linear operator with the autocovariance function as kernel can be used as building blocks for an *object functional principal component analysis* for Ω-valued functional data, including time-varying probability distributions, covariance matrices and time dynamic networks. Analogues of functional principal components for time-varying objects are obtained by applying Fréchet means and projections of distance functions of the random object trajectories in the directions of the eigenfunctions, leading to real-valued *Fréchet scores*. Using the notion of *generalized Fréchet integrals*, we construct *object functional principal components* that lie in the metric space Ω. We establish asymptotic consistency of the sample-based estimators for the corresponding population targets under mild metric entropy conditions on Ω and continuity of the Ω-valued random curves. These concepts are illustrated with samples of time-varying probability distributions for human mortality, time-varying covariance matrices derived from trading patterns and time-varying networks that arise from New York taxi trips.

Keywords: Fréchet integral; Functional data analysis; Metric covariance; Object data; Principal component analysis; Stochastic processes; Time-varying networks

1. Introduction

Time-varying data where one collects an independent and identically distributed sample of random functions, which take values in a general object space that does not have a linear structure, are increasingly common, whereas the statistical methodology for the analysis of such data has been lagging behind. We aim to introduce techniques that will help to fill this gap. For the case where observations consist of samples of random trajectories that take values in $\mathbb{R}^p$, the methodology of choice is often functional data analysis (FDA) (Ramsay and Silverman, 2005; Horvath and Kokoszka, 2012; Wang *et al.*, 2016), where methodology for one-dimensional $(p = 1)$ functional data is readily available. Models for functional data that consist of vector-valued processes $(p > 1)$ have been studied more recently (Zhou *et al.*, 2008; Berrendero *et al.*, 2011; Chiou *et al.*, 2014; Claeskens *et al.*, 2014; Verbeke *et al.*, 2014; Chiou *et al.*, 2016) as well as

Address for correspondence: Paromita Dubey, Department of Statistics, University of California, Davis, 399 Crocker Lane, Davis, CA 95616, USA.
E-mail: pdubey@ucdavis.edu

 1369–7412/20/82000

the case where at each time point one records a random function, i.e. function-valued stochastic processes (Park and Staicu, 2015; Chen and Müller, 2012; Chen *et al.*, 2017). In these models, the responses are in a linear space, either the Euclidean space $\mathbb{R}^p$ or the Hilbert space L^2. Functions of objects in spaces that can be locally approximated by linear spaces such as Riemannian manifolds including spheres have also been considered more recently (Lin *et al.*, 2017; Dai and Müller, 2018). The major objective of this paper is to overcome the global or local linearity assumptions that are inherent in these previous approaches. The challenge is that existing FDA methodology relies on vector operations and inner products, which are no longer available.

Functional principal component analysis (FPCA) (Kleffe, 1973; Dauxois *et al.*, 1982) has emerged as the method of choice to represent and interpret samples of random functions that take values in linear spaces. It also provides dimension reduction by expanding an underlying random process into the basis functions given by the eigenfunctions of the autocovariance operator and then truncating this expansion at a finite number of expansion terms. A related tool is the modes of variation, which enable exploration of the effects of single eigendirections (Castro *et al.*, 1986; Jones and Rice, 1992) and are useful in practical applications (Dong *et al.*, 2018). FPCA also provides a starting point for many theoretical investigations and FDA techniques such as functional clustering (Chiou and Li, 2007; Jacques and Preda, 2014; Suarez and Ghosal, 2016) or regression and classification (Yao *et al.*, 2005a; Dai *et al.*, 2017).

As we enter the era of 'big data', it has become increasingly common to observe more complex, often non-Euclidean, data on a time grid. Technological advances have made it possible to record and store efficiently time courses of image, network, sensor or other complex data. For example, neuroscientists are interested in dynamic functional connectivity, where one essentially observes samples of time-varying covariance or correlation matrices obtained from functional magnetic resonance imaging data for each subject in a sample. Time-varying network data arise in various forms, e.g. road or Internet traffic networks or time evolving social networks, and it is of interest to extract structure and patterns from such data.

To obtain efficient and interpretable summaries of the information that is contained in samples of complex observations is a major task in modern statistics that has led for example to the development of methods such as geodesic principal component analysis in the space of probability distributions on $\mathbb{R}$ (Bigot *et al.*, 2017) and on more general Hilbert spaces (Seguy and Cuturi, 2015) that utilize optimal transport geometry and geodesic curves under the Wasserstein metric. These approaches utilize geodesics to connect the random distributions with the Wasserstein barycentres. We aim here at identifying dominant directions of variation in a sample of time-varying random object trajectories, where the random objects are indexed by time and are in a general metric space. The time-varying aspect provides for a novel and little-explored setting, and to develop tools that are supported by theory and are useful for the further exploration and analysis of such data is the main motivation for this paper.

Although FPCA for samples of functions taking values in smooth Riemannian manifolds has been studied both practically and theoretically (Anirudh *et al.*, 2017; Dai and Müller, 2018), these approaches critically depend on the local Euclidean property of Riemannian manifolds and cannot be extended to functional data objects that take values in more general metric spaces that do not have a tractable and relatively simple Riemannian geometry. FPCA for doubly functional data, where the observations at each time point are functions rather than scalars (Chen *et al.*, 2017), is based on a tensor product representation of the underlying function-valued stochastic process. The functions need to be Hilbert space valued, so this approach cannot be applied to non-Hilbertian data. Because of the lack of linear structure, developing a form of FPCA for random functions taking values in a metric space, which we refer to as *functional random objects*, is a major challenge.

Consider a totally bounded separable metric space (Ω, d) and a random sample of fully observed Ω-valued functional data. Aiming to extend key tools of FDA to cover such data, we first revisit the well-established FPCA for the case of real-valued functional data. The essence of FPCA is contained in the autocovariance structure of the underlying random functions, i.e. their covariance at different time points. This leads to the question how to quantify correlation between random objects in general metric spaces that correspond to the values of the random function at different time points. An example for such an extension of Pearson correlation to the case of multivariate data is the RV-coefficient (Robert and Escoufier, 1976), which is 0 if all the vector components are uncorrelated and is strictly positive otherwise.

In this paper we introduce *metric covariance*, which is a novel association measure for paired data in general metric spaces. Metric covariance differs in key aspects from distance correlation, which is another measure of dependence between metric space data (Lyons, 2013; Székely and Rizzo, 2017), the latter being primarily suited to measure probabilistic independence rather than for quantifying the strength of 'positive' or 'negative' association, which is the primary goal of the former. Unlike distance covariance, the magnitude of metric covariance quantifies the degree of association between paired data objects. The key objective of FPCA is to decompose the variation in a sample of trajectories into orthogonal directions. An important difference between metric covariance and distance covariance, which is specifically relevant in this context, arises when considering the associated notion of variance. In contrast with distance correlation, metric covariance of a random object with itself leads to an interpretable notion of variance for data objects, as we shall demonstrate below. We also show that metric covariance is symmetric and non-negative definite whenever the squared distance d^2 is a semimetric of negative type (Sejdinovic *et al.*, 2013; Lyons, 2013; Schoenberg, 1938). The notion of *metric correlation* can then be easily derived from metric covariance and random objects will be considered to be uncorrelated if they have a metric correlation of size 0.

In FPCA for $\mathbb{R}$-valued functional data, once the autocovariance function has been determined, one defines a linear Hilbert–Schmidt operator whose eigenfunctions represent the orthonormal directions of variance for the functional data in the Hilbert space L^2. The corresponding eigenvalues represent the fraction of variance explained by the respective FPCs, which are the lengths of the projections of the functional data in the direction of each eigenfunction. How can we extend these ideas to object-valued functional data, where we do not have a linear structure or inner product? We proceed by constructing a linear Hilbert–Schmidt operator by using the proposed metric covariance as its kernel and utilize its eigenfunctions and eigenvalues. For real-valued functional data, we obtain the FPCs by the Karhunen–Loève expansion of centred functional data in the eigenbasis, where the FPCs are the inner products of the centred functional data with respect to the eigenfunctions. Unfortunately it is not possible to 'centre' object functional data living in general metric spaces and also we do not have an inner product. In the case of FDA in the Hilbert space L^2, the inner products can be expressed as integrals. Although due to the lack of linear structure there is no integral for functional random objects, the interpretation of inner products as integrals nevertheless provides a way forward that we develop in this paper. We propose two approaches for obtaining FPCs for object functional data: one in which the FPCs are scalar irrespective of the nature of the metric space in which the random objects live, and an alternative approach in which the FPCs themselves are random objects, i.e. Ω valued.

To obtain FPCs in object space, we introduce the notion of a *generalized Fréchet integral* of an Ω-valued curve with respect to a real-valued function, where this integral resides in Ω. Generalized Fréchet integrals depend on the underlying metric d in Ω and are defined under the constraint that the real-valued function in the integrand integrates to 1. We draw inspiration

from the covariance integral for multivariate functional data that was previously introduced as a Fréchet integral (Petersen and Müller, 2016a). This previous integral is a special case of the generalized Fréchet integral that is introduced here; it corresponds to the special case where Ω is the space of covariance matrices and the real-valued function in the integrand is the constant function 1. We demonstrate that the resulting object FPCs, which reside in Ω, provide useful insights about the structure of the underlying functional random objects.

For an alternative scalar approach, we extract relatively simple characteristics from the object functional data. A first step is to define a 'mean' function by using the notion of Fréchet means (Fréchet, 1948). This mean function resides in the object function space and serves as a 'central' trajectory for the object functional data. To obtain a representative scalar FPC for a specific random object trajectory and eigenfunction, we utilize projections of the distance function between the specific random object trajectory and the Fréchet mean trajectory on each of the eigenfunctions. The resulting *Fréchet scores* encapsulate variation in the departures of functional random objects from the Fréchet mean trajectory. As we illustrate in simulations and data analysis, plotting these Fréchet scores against each other often illustrates meaningful patterns in the sample of object functional data that are generally difficult to capture visually, because of their complexity and non-linearity. For example, such plots can aid in detecting the presence of clusters or outliers in functional random objects.

In this paper, we have three major objectives. First, we lay out a framework for extending FPCA to general metric-space-valued functional data. The population target parameters are the *metric autocovariance* operator, its eigenvalues and eigenfunctions and the population Fréchet mean function, which are introduced in Section 2; additionally, the object FPCs, which are generalized Fréchet integrals and the Fréchet scores (Section 3). Second, we provide sample-based estimators of these population targets and establish their asymptotic properties under mild restrictions on the metric entropy of the metric space Ω and the continuity of the object functional data (Section 4). Proofs of all results are in section S1 of the on-line supplement. Third, we illustrate our results through simulations (Section 5) and various data examples (Section 6), which include samples of time-varying probability distributions of age at death obtained from human mortality data of 32 countries, time-varying yellow taxi trip networks of different regions in Manhattan observed daily during the year 2016, and of changing trade patterns between countries that can be represented as time-varying covariance matrices, followed by a brief discussion (Section 7).

2. Metric covariance

2.1. Covariance and correlation for random objects

We consider a totally bounded separable metric space (Ω, d) where d is a metric and an Ω-valued stochastic process $X = \{X(t)\}_{t\in[0,1]}$ on the interval $[0, 1]$. With P denoting the probability measure of the random process X, we are given a sample $\{X_i = (X_i(t))_{t\in[0,1]} : i = 1, 2, \ldots, n\}$ of random Ω-valued functions on $[0, 1]$ generated by P. The simplest case is $\Omega = \mathbb{R}$ with the intrinsic Euclidean metric, where $\{X_1, X_2, \ldots, X_n\}$ is a sample of real-valued functional data. For general metric spaces Ω, we refer to $\{X_1, X_2, \ldots, X_n\}$ as a sample of functional random objects. Inspired by the approach to obtain FPCA for real-valued functional data, our first goal is to quantify the association between random objects $X(s)$ and $X(t)$ in Ω, where s and t are two arbitrary points in the domain $[0, 1]$.

For motivation, consider first the case of real random variables (U, V) with finite covariance. Imagine for a moment that we cannot add, subtract or multiply these random variables and are restricted to compute their distances $d_E(U, V) = |U - V|$. As is well known, we then can

write the variance of U by using an independent and identically distributed copy U' of U by $\mathrm{var}(U) = \frac{1}{2}E\{d_{\mathrm{E}}^2(U, U')\}$.

Interestingly, this non-algebraic construction can be extended to the covariance of U, V: let (U', V') be an independent and identically distributed copy of (U, V). We then obtain an alternative formulation of $\mathrm{cov}(U, V)$ in terms of pairwise distances as follows:

$$\begin{aligned}\mathrm{cov}(U, V) &= E\{(U - E(U))(V - E(V))\} \\ &= \tfrac{1}{4}E\{d_{\mathrm{E}}^2(U, V') + d_{\mathrm{E}}^2(U', V) - 2d_{\mathrm{E}}^2(U, V)\}.\end{aligned}$$

If (U, V) are $\mathbb{R}^d$-valued random variables with $d_{\mathrm{E}}(\cdot, \cdot)$ denoting the Euclidean distance in $\mathbb{R}^d$, a simple calculation shows that, in this case,

$$\tfrac{1}{4}E\{d_{\mathrm{E}}^2(U, V') + d_{\mathrm{E}}^2(U', V) - 2d_{\mathrm{E}}^2(U, V)\} = E\{(U - E(U))^{\mathrm{T}}(V - E(V))\},$$

which is the inner product in the Hilbert space of $\mathbb{R}^d$-valued random variables with finite $E(U^{\mathrm{T}}U)$. Next consider the case where (U, V) are $\mathcal{H}$-valued random variables, where $\mathcal{H}$ is a Hilbert space and $d_{\mathrm{E}}(\cdot, \cdot)$ is replaced by $d_{\mathcal{H}}(U, V) = \|U - V\|_{\mathcal{H}}$, the metric that arises from the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ of the Hilbert space. If the metric $d_{\mathcal{H}}(\cdot, \cdot)$ is bounded then $E(\|U\|_{\mathcal{H}}^2) < \infty$. One can show with some simple algebra and utilizing the Riesz representation theorem that

$$\tfrac{1}{4}E\{d_{\mathcal{H}}^2(U, V') + d_{\mathcal{H}}^2(U', V) - 2d_{\mathcal{H}}^2(U, V)\} = E\{\langle U - E(U), V - E(V)\rangle_{\mathcal{H}}\}, \tag{1}$$

which is the inner product in $L^2(\mathcal{H})$: the Hilbert space of $\mathcal{H}$-valued random variables U such that $E(\|U\|_{\mathcal{H}}^2) < \infty$.

What happens if (U, V) are Ω-valued random variables and we replace $d_{\mathcal{H}}$ by d where (Ω, d) is a general metric space with no vector space structure to rely on? Or, for which spaces does the function $\frac{1}{4}E\{d^2(U, V') + d^2(U', V) - 2d^2(U, V)\}$ retain desirable properties? Proposition 3 of Sejdinovic *et al.* (2013) implies that, whenever d^2 is a semimetric of negative type, there is a Hilbert space $\mathcal{H}$ and an injective map, say $f : \Omega \to \mathcal{H}$, with

$$d^2(U, V) = \|f(U) - f(V)\|_{\mathcal{H}}^2, \tag{2}$$

and therefore it follows from equation (1) that, for some 'remote' Hilbert space $\mathcal{H}$ and the unknown map $f(\cdot)$,

$$\tfrac{1}{4}E\{d^2(U, V') + d^2(U', V) - 2d^2(U, V)\} = E[\langle f(U) - E\{f(U)\}, f(V) - E\{f(V)\}\rangle_{\mathcal{H}}]. \tag{3}$$

Here a space $(\mathcal{Z}, \rho)$ with a semimetric ρ is of negative type if, for all $n \geqslant 2$, $z_1, z_2, \ldots, z_n \in \mathcal{Z}$ and $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R}$ with $\Sigma_{i=1}^n \alpha_i = 0$ we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j \rho(z_i, z_j) \leqslant 0.$$

These considerations motivate the following definition of a generalized version of covariance $\mathrm{cov}_\Omega(U, V)$ for paired random objects (U, V) that take values in $\Omega \times \Omega$, where (Ω, d) is a separable metric space:

$$\mathrm{cov}_\Omega(U, V) = \tfrac{1}{4}E\{d^2(U, V') + d^2(U', V) - 2d^2(U, V)\}, \tag{4}$$

where as above (U', V') is an independent and identically distributed copy of (U, V). We refer to $\mathrm{cov}_\Omega(U, V)$ as the *metric covariance* of U and V. The metric covariance is always finite if

the underlying metric space is bounded and coincides with the usual notion of covariance in Euclidean spaces.

We also define the *metric correlation* between two Ω-valued random variables as

$$\rho_\Omega(U,V) = \frac{\mathrm{cov}_\Omega(U,V)}{\sqrt{\{\mathrm{cov}_\Omega(U,U)\mathrm{cov}_\Omega(V,V)\}}}.$$

By the Cauchy–Schwarz inequality we have $-1 \leqslant \rho_\Omega(U,V) \leqslant 1$. Metric covariance or metric correlation depends on the choice of the metric d and different choices of d might reveal different aspects of association between random objects, depending on the underlying geometry of the metric.

2.2. Metric autocovariance operators

As in the real-valued Euclidean case, we define the metric autocovariance function $C(s,t)$ for functional random objects $\{X_1, X_2, \ldots, X_n\} \in \Omega$ as

$$C(s,t) = \mathrm{cov}_\Omega\{X(s), X(t)\},$$

for all $(s,t) \in [0,1] \times [0,1]$. Obviously, $C(s,t)$ is a symmetric kernel and therefore has real eigenvalues when used as the kernel of a linear Hilbert–Schmidt operator. The following result shows that, for metric spaces (Ω, d) for which the squared distance function d^2 is of negative type, the metric autocovariance operator is positive semidefinite.

Proposition 1. If Ω is separable and d^2 is of negative type, then $C(s,t)$ is a non-negative definite kernel.

By proposition 3 in Sejdinovic *et al.* (2013) and equation (3), $\mathrm{cov}_\Omega(U,V) = 0$ implies that there is an abstract Hilbert space $\mathcal{H}$ and an injective map $f : \Omega \to \mathcal{H}$ such that $f(U)$ and $f(V)$ are orthogonal in $L^2(\mathcal{H})$. Note that $\mathrm{var}_\Omega(U) = \mathrm{cov}_\Omega(U,U) = \frac{1}{2}E\{d^2(U,U')\}$, which for real-valued random variables equals $\mathrm{var}(U)$.

Formally, we can define the *metric autocovariance operator* as a linear Hilbert–Schmidt integral operator T_C that operates on functions $g \in L^2([0,1])$ and utilizes the metric autocovariance kernel,

$$(T_C g)(s) = \int_0^1 C(s,t)g(t)\mathrm{d}t.$$

We note that for example theorem 4.6.4 of Hsing and Eubank (2015) implies the non-negative definiteness of the kernel $C(s,t)$, in the sense that $\langle T_C f, f\rangle \geqslant 0$ for all f.

By Mercer's theorem there is an orthonormal basis $\{\phi_i\}_{i=1}^\infty$ of $L^2([0,1])$ consisting of eigenfunctions of T_C such that the corresponding sequence of eigenvalues $\{\lambda_i\}_{i=1}^\infty$, which are ordered in declining order, is non-negative, since $C(s,t)$ is positive semidefinite. The eigenfunctions corresponding to non-zero eigenvalues are continuous on $[0,1]$ and C has the representation

$$C(s,t) = \sum_{j=1}^\infty \lambda_j \phi_j(s)\phi_j(t),$$

where the convergence is absolute and uniform; see, for example, lemma 4.6.1 and theorems 4.5.2, 4.6.2, 4.6.5 and 4.6.7 of Hsing and Eubank (2015).

We have thus accomplished the first step of extending FPCA from Euclidean-valued functional data to general metric space-valued functional data. The eigenfunctions $\{\phi_j\}_{j=1}^\infty$ can be interpreted as principal directions of variation of the functional object process and will be ordered according to the size of the associated eigenvalues. We can view the eigenvalues as

representing a metric version of the 'fraction of variance explained', which is their common interpretation in the real-valued case. The only requirement for this extension is that the squared metric d^2 is of negative type but this is not a severe restriction and in the light of proposition 3 of Sejdinovic *et al.* (2013) is true for the following examples:

(a) (Ω, d) where Ω is the space of univariate probability distributions on a common compact support in $T \subset \mathbb{R}$—choices of d include the popular 2-Wasserstein metric or the L^2-metric;
(b) (Ω, d) where Ω is the space of correlation matrices of a fixed dimension r, where the choice of metrics includes the Frobenius metric, log-Frobenius metric, power Frobenius metric and Procrustes metric (Dryden *et al.*, 2009; Pigoli *et al.*, 2014; Tavakoli *et al.*, 2019);
(c) (Ω, d) where Ω is the space of networks with a fixed number, say r, of nodes—one can view networks as adjacency matrices or graph Laplacians equipped with the Frobenius metric (Ginestet *et al.*, 2017) or as resistance matrices equipped with the resistance perturbation metric (Monnig and Meyer, 2018).

We conclude that in most cases of interest the autocovariance operator and its eigenfunctions will be well defined.

2.3. Interpretation of metric covariance

When X and Y are real valued, classical Pearson correlation captures the strength and sign of linear (also monotone) associations between X and Y. From a geometrical perspective, Pearson correlation can be interpreted as the cosine of the angle between X and Y. In $\mathbb{R}^d$, angles between vectors are defined by using inner products, which can also be used for data in Hilbert space to characterize dependence. Specifically, for random functions in the metric space L^2 this idea leads to the notion of 'dynamic correlation' in FDA (Dubin and Müller, 2005), which was found to be useful for data analysis in genetics (Opgen-Rhein and Strimmer, 2006) and psychology (Liu *et al.*, 2016). Dynamic correlation turns out to be equivalent to metric covariance when the random objects are in the Hilbert space $L^2([0, 1])$, equipped with the usual L^2-metric. Metric covariance then provides a generalization beyond Hilbert spaces.

For general metric spaces, under the weak assumption that the squared metric is of negative type, the map f from object to Hilbert space in equation (2) implies that metric covariance can be derived from the inner product in an abstract Hilbert space, whereas metric correlation is obtained by standardizing metric covariance and is thus tied to the notion of an angle in an abstract space. Hence its magnitude can be interpreted as the strength of association between random objects. Although we use the existence of a map f and an associated abstract Hilbert space, we do not require knowledge about f. Metric covariance is thus a natural extension of Pearson covariance to general metric spaces.

In recent work (Petersen and Müller, 2019a), Wasserstein covariance for pairs of univariate probability distributions was introduced and was shown to have an appealing interpretation as an expected value of an inner product of optimal transport maps. More specifically, if f_1 and f_2 are the components of a random bivariate density process and $F_1^{-1}(\cdot)$ and $F_2^{-1}(\cdot)$ the corresponding random quantile functions, the squared Wasserstein distance between f_1 and f_2 is given by

$$d_{\mathrm{W}}^2(f_1, f_2) = \int_0^1 \{Q_1(t) - Q_2(t)\}^2 \mathrm{d}t$$

and the Wasserstein covariance between f_1 and f_2 was introduced as

$$\mathrm{cov}_{\mathrm{W}}(f_1, f_2) = E\left(\int_0^1 [Q_1(t) - E\{Q_1(t)\}][Q_2(t) - E\{Q_2(t)\}]\mathrm{d}t \right).$$

Wasserstein covariance is then easily seen to be a special case of metric covariance when the metric-space-valued random objects are probability distributions and the Wasserstein metric is used. This Wasserstein version of metric covariance was found to quantify the degree of synchronization of the movement of probability mass from the marginal Fréchet means of the probability distributions to the random components of a multivariate density process. In applications to functional magnetic resonance imaging data, this Wasserstein version led to new findings and insights about differences in brain connectivity of normal *versus* Alzheimer disease patients, which is a topic of special interest in neuroimaging (Petersen and Müller, 2019a). The examples of dynamic correlation for Hilbert-space-valued random variables in FDA and of Wasserstein covariance or Wasserstein correlation demonstrate the utility of metric covariance or metric correlation in non-standard spaces and its interpretability in applications. This provides evidence that metric covariance and metric correlation are indeed useful tools for data analysis in general metric spaces.

A word of caution is in order. Although metric covariance can be universally applied and in the space of distributions with the Wasserstein metric has an interpretation as an inner product of transport maps, such interpretations hinge on the specific metric space in which the random objects are located and may not be available for all spaces. In practice, interpretations for specific scenarios can be important. The choice of the metric also matters and should be considered carefully, as it will affect the interpretation of metric covariance.

Apart from the interpretation of covariance as the expectation of an inner product, the diagonal elements of the metric autocovariance surface reflect a natural notion of variance of metric-space-valued objects, as

$$\mathrm{var}_\Omega = \tfrac{1}{2} E\{d^2(U, U')\}, \tag{5}$$

where U' is an independent copy of U. This provides a variation measure that is tied to the average squared distance of objects that are independently sampled from the underlying population, which is a natural and interpretable measure of spread that is well known to coincide with conventional variance in the Euclidean case.

Since it is sensible to define variance for metric-space-valued random objects as

$$\tfrac{1}{2} E\{d^2(U, U')\} = \tfrac{1}{2} E\{d^2(U, U') - d^2(U, U)\},$$

it is then natural to extend this to a covariance measure between random objects (U, V) that reflects the difference between squared distances when sampling independently from the marginal distributions of U and V and when sampling from the joint distribution of (U, V). This simple idea provides another avenue to suggest

$$\widetilde{\mathrm{cov}}_\Omega(U, V) = E\{d^2(U, V') - d^2(U, V)\}.$$

Symmetrizing this expression and adding the factor 0.25 to match the usual definition of covariance in the Euclidean case then leads to formula (4). These arguments also lead to an interpretation of the total variance that corresponds to the trace of the proposed metric covariance operator $C(s, t)$, as an integrated squared distance between the functional random objects X and an independent copy X',

$$\sum_{j=1}^{\infty} \lambda_j = \int_0^1 \mathrm{cov}_\Omega\{X(t), X(t)\}\mathrm{d}t = \frac{1}{2} \int_0^1 E[d^2\{X(t), X'(t)\}]\mathrm{d}t. \tag{6}$$

We find in our examples and applications that the eigenfunctions that are derived from metric covariance lead to useful and often well interpretable modes of variation of the time-varying

metric random objects in the sense of Jones and Rice (1992), adding to the practical appeal of metric covariance for the analysis of functional random objects.

To conclude this discussion, we note that metric covariance differs substantially from distance correlation (Székely *et al.*, 2007; Lyons, 2013). A distinguishing feature of distance correlation is that it is equivalent to probabilistic independence between the distributions of U and V when it is 0, but we find that it is not suitable as a covariance or correlation measure for random objects in the situations that we study here. Specifically, the autocovariance operator that it generates is not useful for our purposes. For further details on this, see section S2 in the on-line supplement.

3. Functional principal components: generalized Fréchet integrals and Fréchet scores

3.1. Generalized Fréchet integrals and object functional principal components

FPCs in the case of real-valued functional data are projections of the centred process onto the directions of the eigenfunctions and therefore summarize how a function differs from the mean function along orthonormal eigenfunction directions. Formally the FPC of the ith process $X_i(t)$ and the kth eigenfunction $\phi_k(t)$ is

$$\xi_{ik} = \int_0^1 \{X_i(t) - \mu(t)\}\phi_k(t)\mathrm{d}t,$$

where $\mu(\cdot)$ is the mean process. The part of the score contributing to the variability of the functional data is $\int_0^1 X_i(t)\phi_k(t)\mathrm{d}t$, which is just a horizontal shift of the actual scores, so centring is not needed when our goal is to decompose the variability of the random processes X, which is fortuitous as one cannot 'centre' object data to obtain an analogue of $X(t) - \mu(t)$, as algebraic operations such as subtraction are not feasible in metric spaces.

In the Euclidean case for any function ϕ on $[0, 1]$, whenever $\int_0^1 \phi(t)\mathrm{d}t \neq 0$, we can obtain a scaled version of the integral of X with respect to ϕ as follows:

$$\int_0^1 X(t)\frac{\phi(t)}{\int_0^1 \phi(t)\mathrm{d}t}\mathrm{d}t = \operatorname*{arg\,inf}_{\omega\in\mathbb{R}} \int_0^1 d_{\mathrm{E}}^2\{\omega, X(t)\}\frac{\phi(t)}{\int_0^1 \phi(t)\mathrm{d}t}\mathrm{d}t.$$

This suggests defining an integral of an Ω-valued function S with respect to a real-valued function ϕ which integrates to 1. For any real-valued function ϕ with $\int_0^1 \phi(t)\mathrm{d}t = 1$, we define the *generalized Fréchet integral* of S with respect to ϕ as

$$\int_{\oplus} S(t)\phi(t)\mathrm{d}t = \operatorname*{arg\,inf}_{\omega\in\Omega} \int_0^1 d^2\{\omega, S(t)\}\phi(t)\mathrm{d}t, \tag{7}$$

provided that the integral $\int_0^1 d^2\{\omega, S(t)\}\phi(t)\mathrm{d}t$ exists as a limit of Riemann sums for all $\omega \in \Omega$ and the minimizer of the integrals over $\omega \in \Omega$ exists and is unique. A special case of the integral in equation (7) was introduced as the Fréchet integral in Petersen and Müller (2016a), where an integral for the space of covariance matrices was constructed for $\phi \equiv 1$.

The Fréchet integrals that are defined here are far more general. Generalized Fréchet integrals can be interpreted as an extension of weighted Fréchet means (Fréchet, 1948). We omit the additional term 'generalized' in what follows and note that Fréchet integrals can be interpreted as projections of functional random objects onto functions ϕ, by weighting the elements $S(t)$ according to the value of $\phi(t)$, in direct analogy to projections in the linear function space L^2. This feature motivates us to employ Fréchet integrals to obtain object FPCs.

For fixed $\omega \in \Omega$ consider the Fréchet integral function

$$I(\omega) = \int_0^1 d^2\{\omega, S(t)\}\phi(t)\mathrm{d}t,$$

which, if it exists, is the limit of Riemann sums. A sufficient condition for its existence is that $d^2\{\omega, S(t)\}\phi(t)$ is a continuous function of $t \in [0,1]$. If the metric is bounded and S and ϕ are continuous for $t \in [0,1]$, the function $d^2\{\omega, S(t)\}\phi(t)$ is a continuous function of $t \in [0,1]$ and the integral $I(\omega)$ exists for all ω. Note that, for any $\omega \in \Omega$, $I(\omega)$ is finite by the Cauchy–Schwarz inequality whenever the metric space is bounded and the L^2-norm of the function $\phi(\cdot)$ is finite.

If the integrals $I(\omega)$ exist as limits of Riemann sums, the question arises under which conditions the minimizers of the Riemann sums converge and whether the limit of the minimizers coincides with the Fréchet integral $\int_\oplus S(t)\phi(t)\,\mathrm{d}t$. Proposition 2 below addresses this question. Let $0 = x_0 < x_1 < x_2 < \ldots < x_k = 1$ be a partition $\mathcal{P}$ of $[0,1]$, where the $[x_j, x_{j+1}]$ are the subintervals of the partition and the length of the jth subinterval is $\Delta_j = x_{j+1} - x_j$. The mesh size $\epsilon_\mathcal{P}$ of the partition is given by $\epsilon_\mathcal{P} = \max_j \Delta_j$. We select $t_0, t_1, \ldots, t_{k-1}$ such that, for each j, $t_j \in [x_j, x_{j+1}]$. For each $\omega \in \Omega$, the Riemann sum $I_\mathcal{P}(\omega)$ corresponding to the partition $\mathcal{P}$ and $t_0, t_1, \ldots, t_{k-1}$ is given by

$$I_\mathcal{P}(\omega) = \sum_{j=0}^{k-1} d^2\{\omega, S(t_j)\}\phi(t_j)\Delta_j$$

and the Riemann integral $I(\omega)$ is obtained as a limit of Riemann sums as the partition becomes finer. Formally, $I(\omega) = \lim_{\epsilon_\mathcal{P} \to 0} I_\mathcal{P}(\omega)$.

We shall invoke the following assumptions for the integral function $I(\omega)$. For ease of notation, we suppress t in $\int_\oplus S(t)\phi(t)\mathrm{d}t$, writing $\int_\oplus S\phi$ in what follows.

Assumption (i). The integrand function $H(\omega, t) = d^2\{\omega, S(t)\}\phi(t)$ is uniformly equicontinuous in $t \in [0,1]$ and $\omega \in \Omega$.

Assumption (ii). $\int_\oplus S\phi = \arg\min_{\omega\in\Omega} I(\omega)$ exists and is unique, and $\inf_{d(\omega, \int_\oplus S\phi) > \delta} I(\omega) > I(\int_\oplus S\phi)$ for all $\delta > 0$.

Assumption (iii). There are constants $\beta > 0, \nu > 0$ and $C > 0$ such that

$$I(\omega) - I\left(\int_\oplus S\phi\right) \geqslant C d^\beta\left(\omega, \int_\oplus S\phi\right)$$

whenever $d(\omega, \int_\oplus S\phi) < \nu$.

Define $\Sigma_{\mathcal{P},\oplus} S\phi = \arg\min_{\omega\in\Omega} I_\mathcal{P}(\omega)$.

Proposition 2.

(a) Under assumption (i), $I_\mathcal{P}(\omega)$ converges to $I(\omega)$ uniformly in ω as $\epsilon_\mathcal{P} \to 0$.
(b) Under assumptions (i) and (ii), $\lim_{\epsilon_\mathcal{P}\to 0} d(\Sigma_{\mathcal{P},\oplus} S\phi, \int_\oplus S\phi) = 0$.
(c) If $\lim_{\epsilon_\mathcal{P}\to 0} h(\epsilon_\mathcal{P}) \sup_{\omega\in\Omega} |I_\mathcal{P}(\omega) - I(\omega)| = 0$ for a function h with $h(\delta) \to \infty$ as $\delta \to 0$, then, under assumption (iii), $\lim_{\epsilon_\mathcal{P}\to 0} h^{1/\beta}(\epsilon_\mathcal{P}) d(\Sigma_{\mathcal{P},\oplus} S\phi, \int_\oplus S\phi) = 0$.

As a continuous function on a compact interval is uniformly continuous, whenever $S(\cdot)$ is continuous and $\phi(\cdot)$ is bounded and continuous, assumption (i) holds since, for $D = \mathrm{diam}(\Omega)$,

$$\begin{aligned} |H(\omega, t_1) - H(\omega, t_2)| &= |d^2\{\omega, S(t_1)\}\phi(t_1) - d^2\{\omega, S(t_2)\}\phi(t_1) + d^2\{\omega, S(t_2)\}\phi(t_1) \\ &\quad - d^2\{\omega, S(t_2)\}\phi(t_2)| \\ &\leqslant 2D\, d\{S(t_1), S(t_2)\}|\phi(t_1)| + D^2|\phi(t_1) - \phi(t_2)|. \end{aligned}$$

Assumption (i) is sufficient to guarantee that the Fréchet integrals are well defined, whereas Assumption (iii) is a restriction on the curvature of the function $I(\omega)$ near its minimizer, implying convergence rates of the approximations of the Fréchet integrals. A few examples of spaces that satisfy assumptions (ii) and (iii) are as follows.

(a) Let (Ω, d_{W}) be the space of univariate probability distributions on a common support $T \subset R$. For any $\omega \in \Omega$, denote the corresponding random distribution and quantile functions by $Q(\omega)$. The squared 2-Wasserstein metric between distributions ω_1 and ω_2 is

$$d^2_{\mathrm{W}}(\omega_1, \omega_2) = d^2_{L^2}\{Q(\omega_1), Q(\omega_2)\} = \int_0^1 \{Q(\omega_1)(u) - Q(\omega_2)(u)\}^2 \mathrm{d}u.$$

For any $S(t)$ taking values in Ω, where we view $Q\{S(t)\}$ as the quantile function of the distribution at time $t \in [0, 1]$, writing $Q\{S(t)\}(u)$ for the uth quantile of the distribution at time t, define $Q^*(u) = \int_0^1 Q\{S(t)\}(u)\phi(t)\mathrm{d}t$. Since $\int_0^1 \phi(t)\mathrm{d}t = 1$, a simple calculation shows that, for any $\omega \in \Omega$,

$$\arg\inf_{\omega \in \Omega} I(\omega) = \arg\inf_{\omega \in \Omega} d^2_{L^2}\{Q(\omega), Q^*\};$$

therefore the minimizer exists and is unique by the convexity of the space of univariate quantile functions. By the orthogonal projection theorem the minimizer $\tilde{\omega}$ is uniquely characterized by

$$\langle Q^* - Q(\tilde{\omega}), Q(\omega) - Q(\tilde{\omega}) \rangle_{L^2} \leqslant 0,$$

for all $\omega \in \Omega$, and therefore it is enough to choose $\nu = C = 1$ and $\beta = 2$ in assumption (iii).

(b) Consider the space of graph Laplacians or graph adjacency matrices of connected, undirected and simple graphs with a fixed number r of nodes (Ω, d_{F}), equipped with the Frobenius metric d_{F}. For any $\omega \in \Omega$,

$$d^2_{\mathrm{F}}(\omega_1, \omega_2) = \sum_{j=1}^{r} \sum_{k=1}^{r} (\omega_{1,jk} - \omega_{2,jk})^2.$$

For any $S(t)$ taking values in Ω, let $S_{jk}(t)$ be the (j, k)th entry of the graph Laplacian or the graph adjacency matrix. Define $S^*_{jk} = \int_0^1 S_{jk}(t)\phi(t)\mathrm{d}t$. Since $\int_0^1 \phi(t)\mathrm{d}t = 1$, it can be easily seen that, for any $\omega \in \Omega$,

$$\arg\inf_{\omega \in \Omega} I(\omega) = \arg\inf_{\omega \in \Omega} d^2_{\mathrm{F}}(\omega, S^*),$$

and so the minimizer exists and is unique by the convexity of the space of graph Laplacians (Ginestet *et al.*, 2017) and the space of graph adjacency matrices. Again, by the orthogonal projection theorem, the minimizer $\tilde{\omega}$ is uniquely characterized by

$$\sum_{j=1}^{r} \sum_{k=1}^{r} (\tilde{\omega}_{jk} - S^*_{jk})(\tilde{\omega}_{jk} - \omega_{jk}) \leqslant 0,$$

for all $\omega \in \Omega$ and therefore it is enough to choose $\nu = C = 1$ and $\beta = 2$ in assumption (iii).

(c) The same arguments also imply that (Ω, d_{F}) satisfies assumptions (ii) and (iii) when Ω is the space of correlation matrices of a fixed dimension r.

As we have seen, for general metric spaces Ω, under mild assumptions on the boundedness of the metric and continuity of the functions $S(\cdot)$ and $\phi(\cdot)$, the Fréchet integral has nice properties if it exists and is unique. Moreover, when Ω is bounded and $\int_0^1 |\phi(t)|\mathrm{d}t < \infty$,

$$|I(\omega_1) - I(\omega_2)| \leqslant 2Dd(\omega_1, \omega_2) \int_0^1 |\phi(t)| \mathrm{d}t,$$

and therefore $I(\omega)$ is a continuous function of $\omega \in \Omega$. This ensures that the Fréchet integral always exists when Ω is compact.

We now define the FPCs corresponding to the bounded continuous eigenfunctions ϕ_k of the metric autocovariance operator in the object space by using Fréchet integrals. For this, we assume that all trajectories $\{X_i(t)\}_{t\in[0,1]}$ have continuous sample paths almost surely and the metric space Ω is bounded, and furthermore that the following assumptions hold.

Assumption 1. $\int_0^1 \phi_k(t)\mathrm{d}t \neq 0$.

Assumption 2. $\int_\oplus X_i \phi_k^*$ exists and is unique almost surely for all $i = 1, \ldots, n$, where $\phi_k^*(t) = \phi(t)/\int_0^1 \phi(t)\mathrm{d}t$.

Then *object FPCs* for X_i and ϕ_k are defined as the Fréchet integrals

$$\psi_\oplus^{ik} = \int_\oplus X_i \phi_k^*, \tag{8}$$

which are random objects in Ω. Similarly to ordinary FPCA we can choose various basis functions aiming to explain a desired percentage of variation in the data utilizing the eigenvalues of the metric autocovariance operator. If $\Omega = \mathbb{R}$, the object FPCs correspond to a location- and scale-shifted version of the ordinary FPCs.

3.2. Fréchet scores

Exploratory data analysis such as checking for clusters or outliers often benefits from plotting the FPCs against each other for the case of real-valued functional data. FPCs defined by using Fréchet integrals live in the object space Ω and therefore visualizing them is non-trivial. One approach is to obtain their projections to a lower dimensional real space by using multidimensional scaling or its variants (Kruskal, 1964; Belkin and Niyogi, 2002) and then visualizing the projections. Here we propose another approach for obtaining a scalar version of object FPCs. The resulting scalar FPCs are interpretable and can be plotted against each other and are thus useful for exploratory data analysis.

In the real-valued case, one obtains projections of the deviations of the observed random curves from the mean curve onto dominant eigenfunctions. Although the concept of a mean function can be generalized to object functional data by using Fréchet means (Fréchet, 1948), one cannot centre object data and does not have directional information. Nevertheless, it is possible to study how distances of sample curves from the mean curve project onto a few dominant eigenfunctions, in analogy to the real-valued case. Formally, given a random object process $\{X(t)\}_{t\in[0,1]}$, the population Fréchet mean function is

$$\mu_\oplus(t) = \underset{\omega\in\Omega}{\arg\min}\, E[d^2\{\omega, X(t)\}],$$

where we assume existence and uniqueness of the minimizer. For real-valued functional data under the Euclidean metric the Fréchet mean function coincides with the usual pointwise mean function. Defining distance functions

$$D_i(t) = d\{X_i(t), \mu_\oplus(t)\}$$

for sample trajectories X_i, we represent the scalar functions D_i in the eigenbasis of the metric autocovariance operator, obtaining the coefficients

$$\beta_{ik} = \int_0^1 D_i(t)\phi_k(t)\mathrm{d}t = \int_0^1 d\{X_i(t), \mu_\oplus(t)\}\phi_k(t)\mathrm{d}t. \tag{9}$$

We refer to the scalars β_{ik} as the *Fréchet scores*.

The Fréchet scores can be interpreted as decomposition of the departures of the sample elements from the 'central' Fréchet mean curve in predominant directions of variation. They can be plotted against each other and have the potential to provide interesting insights, as we shall illustrate in the data applications. Considering the existence of the Fréchet scores, with D denoting as before the diameter of the totally bounded metric space Ω, continuity of the Fréchet mean function implies that, for any $t_1, t_2 \in [0,1]$,

$$\begin{aligned}|d^2\{X_i(t_1), \mu_\oplus(t_1)\} - d^2\{X_i(t_2), \mu_\oplus(t_2)\}| &= |d^2\{X_i(t_1), \mu_\oplus(t_1)\} - d^2\{X_i(t_1), \mu_\oplus(t_2)\} \\ &\quad + d^2\{X_i(t_1), \mu_\oplus(t_2)\} - d^2\{X_i(t_2), \mu_\oplus(t_2)\}| \\ &\leqslant 2D\,[d\{\mu_\oplus(t_1), \mu_\oplus(t_2)\} + d\{X_i(t_1), X_i(t_2)\}].\end{aligned}$$

Thus, for a continuous eigenfunction ϕ_k, the function $d^2\{X_i(t), \mu_\oplus(t)\}\phi_k(t)$ is a continuous function of $t \in [0,1]$ almost surely and therefore the Fréchet scores will exist. Proposition 3 shows that under the following assumption 3 the Fréchet mean function is indeed continuous.

Assumption 3. For each $t \in [0,1]$, the pointwise Fréchet mean $\mu_\oplus(t)$ exists and is unique, and

$$\inf_{d\{\omega, \mu_\oplus(t)\} > \gamma} E[d^2\{\omega, X(t)\}] > E[d^2\{\mu_\oplus(t), X(t)\}]$$

for any $\gamma > 0$.

Proposition 3. If the random object process $\{X(t)\}_{t\in[0,1]}$ has almost surely continuous paths, then $\mu_\oplus(\cdot)$ is continuous under assumption 3.

Assumption 3 is satisfied for the space (Ω, d_W) of univariate probability distributions with the 2-Wasserstein metric and also for the space (Ω, d_F), where Ω is the space of covariance matrices or alternatively graph Laplacians of fixed dimension with the Frobenius metric d_F (Dubey and Müller, 2019; Petersen and Müller, 2019b).

4. Estimation and theory

Having defined suitable population targets, our goal now is to construct appropriate estimators, starting with a sample of functional random objects. An empirical estimator of the metric autocovariance operator $C(s,t)$ as defined in Section 2 is given by

$$\hat{C}(s,t) = \frac{1}{4n(n-1)} \sum_{i \neq j} f_{s,t}(X_i, X_j), \tag{10}$$

where

$$f_{s,t}(X_i, X_j) = d^2\{X_i(s), X_j(t)\} + d^2\{X_j(s), X_i(t)\} - d^2\{X_i(s), X_i(t)\} - d^2\{X_j(s), X_j(t)\}.$$

Observe that, for each $s,t \in [0,1]$, $\hat{C}(s,t)$ is a U-statistic and the class $\{\hat{C}(s,t) : s,t \in [0,1]\}$ is a family of U-statistics.

Noting that $\hat{C}(s,t)$ can be viewed as a stochastic process indexed by the function class $\mathcal{F} = \{f_{s,t}(\cdot,\cdot) : s,t \in [0,1]\}$, where

$$f_{s,t}(x,y) = d^2\{x(s), y(t)\} + d^2\{y(s), x(t)\} - d^2\{x(s), x(t)\} - d^2\{y(s), y(t)\},$$

enables us to apply the theory of U-processes (Nolan and Pollard, 1987, 1988; Arcones and Giné, 1993) for weak convergence (Billingsley, 1968; van der Vaart and Wellner, 1996). For the uniform convergence of $\{\hat{C}(s,t) : s,t \in [0,1]\}$, we need an assumption on the rate of continuity of the functional random objects.

Assumption 4. The process $X(\cdot)$ is almost surely α Hölder continuous for some $0 < \alpha \leqslant 1$, where the Hölder constant has a finite second moment, i.e. for some non-negative function $G(X)$ we have

$$d\{X(s), X(t)\} \leqslant G(X)|s-t|^{\alpha},$$

where $E\{G(X)\}^2 < \infty$.

Theorem 1. Under assumption 4, the sequence of stochastic processes

$$U_n(s,t) = \sqrt{n}\{\hat{C}(s,t) - C(s,t)\}$$

converges weakly to a Gaussian process with mean 0 and covariance function

$$R_{(s,t),(u,v)} = \mathrm{cov}\{f_{s,t}(X,X'), f_{u,v}(X,X')\}.$$

where X' is an independent and identically distributed copy of X.

Writing $\hat{\lambda}_j$ and $\hat{\phi}_j$ for the eigenvalues and eigenfunctions of $\hat{C}(s,t)$, uniform convergence and rates of convergence of these estimates of the eigenvalues and eigenfunctions of the metric autocovariance operator to their targets are obtained as a direct consequence of proposition 1 under the following assumption on the spacings of the eigenvalues.

Assumption 5. For each $j \geqslant 1$, the eigenvalue λ_j has multiplicity 1, i.e it holds that $\delta_j > 0$, where $\delta_j = \min_{1 \leqslant l \leqslant j}(\lambda_l - \lambda_{l+1})$.

Corollary 1 (Bosq, 2000). Under assumptions 4 and 5,

$$|\hat{\lambda}_j - \lambda_j| = O_P(1/\sqrt{n}).$$

$$\sup_{s \in [0,1]} |\hat{\phi}_j(s) - \phi_j(s)| = O_P\{1/(\delta_j\sqrt{n})\}.$$

As in classical FDA, the eigenfunctions ϕ_j are uniquely identifiable only up to a sign change. For theoretical considerations such as the convergence in corollary 1, we may always assume that true and estimated eigenfunctions are aligned in the sense that $\langle \hat{\phi}_j, \phi_j \rangle \geqslant 0$. Our next objective is to obtain sample estimators for the object FPCs (8) that were defined in Section 3.1. For each j, consider the following estimators of $\phi_j^*(t)$:

$$\hat{\phi}_j^*(t) = \frac{\hat{\phi}_j(t)}{\int_0^1 \hat{\phi}_j(t)\mathrm{d}t}.$$

A natural estimator for the Fréchet integral $\psi_{\oplus}^{ik}$ is then

$$\hat{\psi}_{\oplus}^{ik} = \int_{\oplus} X_i \hat{\phi}_j^* = \arg\min_{\omega \in \Omega} \int_0^1 d^2\{\omega, X_i(t)\}\hat{\phi}_j^*(t)\mathrm{d}t. \tag{11}$$

To obtain convergence of $\hat{\psi}_{\oplus}^{ik}$ to its population target, we make the following assumptions.

Assumption 6. For every i and k, $\psi_{\oplus}^{ik}$ and $\hat{\psi}_{\oplus}^{ik}$ exist and are unique almost surely. Moreover, for any $\varepsilon > 0$, $c_\varepsilon = \inf_{d(\omega, \psi_{\oplus}^{ik}) > \varepsilon}[\int_0^1 d^2\{\omega, X_i(t)\}\phi^*(t)\mathrm{d}t - \int_0^1 d^2\{\psi_{\oplus}^{ik}, X_i(t)\}\phi^*(t)\mathrm{d}t] > 0$ almost surely.

Assumption 7. There are constants $\beta_1 > 1$, $\nu' > 0$ and $C' > 0$ such that, almost surely,

$$\left[\int_0^1 d^2\{\omega, X_i(t)\}\phi^*(t)\mathrm{d}t - \int_0^1 d^2\{\psi_\oplus^{ik}, X_i(t)\}\phi^*(t)\mathrm{d}t\right] \geqslant C' d^{\beta_1}(\omega, \psi_\oplus^{ik}),$$

whenever $d(\omega, \psi_\oplus^{ik}) < \nu'$.

Assumption 6 on the existence and uniqueness of the Fréchet integrals is used to establish consistency. Assumption 7 is a restriction on the local behaviour of the integrals around the minimizer and determines the rate of convergence.

Theorem 2. Under assumptions 1, 2, 4 and 6,

$$d(\hat{\psi}_\oplus^{ik}, \psi_\oplus^{ik}) = o_P(1).$$

If additionally assumption 7 holds, then

$$d(\hat{\psi}_\oplus^{ik}, \psi_\oplus^{ik}) = O_P(n^{-1/(2\beta_1)}).$$

Here we choose $\hat{\phi}_j$ to be such that $\langle \hat{\phi}_j, \phi_j \rangle \geqslant 0$ which ensures matching signs for the true and estimated eigenfunctions in the computation of $\hat{\psi}_\oplus^{ik}$ and $\psi_\oplus^{ik}$.

Next we provide estimates of the Fréchet scores and study their asymptotics. The starting point is the following estimator of the population Fréchet mean function:

$$\hat{\mu}_\oplus(t) = \underset{\omega \in \Omega}{\arg\min} \frac{1}{n} \sum_{i=1}^n d^2\{X_i(t), \omega\}. \tag{12}$$

We need the following assumptions.

Assumption 8. The Fréchet mean function estimate $\hat{\mu}_\oplus(t)$ exists and is unique almost surely for all $t \in [0, 1]$. Additionally, for every $\varepsilon > 0$, there exists $\tau(\varepsilon) > 0$ such that

$$\lim_{n\to\infty} P\left(\inf_{s\in[0,1]} \inf_{d\{\omega, \hat{\mu}_\oplus(s)\} > \varepsilon} \frac{1}{n} \sum_{l=1}^n [d^2\{X_i(s), \omega\} - d^2\{X_i(s), \hat{\mu}_\oplus(s)\}] \geqslant \tau(\varepsilon)\right) = 1.$$

Assumption 9. There are small $\delta > 0$ and constants $0 < \nu_\delta \leqslant 1$ and $H_\delta > 0$, such that for all Ω-valued functions $\omega(\cdot)$ with $d_\infty(\omega, \mu_\oplus) < \delta$, where $d_\infty(\omega, \mu_\oplus) = \sup_{s\in[0,1]} d_\infty\{\omega(s), \mu_\oplus(s)\}$, the functions $\omega(\cdot)$ are ν_δ Hölder continuous with Hölder constant bounded above by H_δ, i.e.

$$d\{\omega(s), \omega(t)\} \leqslant H_\delta |s - t|^{\nu_\delta}.$$

Assumption 10. For $I(\delta) = \int_0^1 \sup_{s\in[0,1]} \sqrt{\log(N[A\varepsilon\delta, B_\delta\{\mu_\oplus(s)\}, d])}\mathrm{d}\varepsilon$, it holds that $I(\delta) = O(1)$ as $\delta \to 0$ for all sufficiently small $\delta > 0$ and for any constant $A > 0$. Here $B_\delta\{\mu_\oplus(s)\} = \{\omega \in \Omega : d\{\omega, \mu_\oplus(s)\} < \delta\}$ is the δ-ball around $\mu_\oplus(s)$ and $N[\varepsilon, B_\delta\{\mu_\oplus(s)\}, d]$ is the covering number, i.e. the minimum number of balls of radius ε required to cover $B_\delta\{\mu_\oplus(s)\}$ (van der Vaart and Wellner, 1996).

Assumption 11. There are $\alpha > 0$, $D > 0$ and $\beta_2 > 1$ such that

$$\inf_{s\in[0,1]} \inf_{d\{\omega, \mu_\oplus(s)\} < \alpha} (E[d^2\{X(s), \omega\}] - E[d^2\{X(s), \mu_\oplus(s)\}] - D d^{\beta_2}\{\omega, \mu_\oplus(s)\}) \geqslant 0.$$

Proposition 4. Under assumptions 3 and 8,

$$\sup_{t\in[0,1]} d\{\hat{\mu}_\oplus(t), \mu_\oplus(t)\} = o_P(1).$$

Assumptions 4 and 9–11 are required to obtain an entropy condition for the space of functional random objects (lemma 1 below), which is used to establish the rate of convergence of the sample Fréchet mean function. We note that assumption 9, where we assume that in a sufficiently close neighbourhood of the true Fréchet mean function $\mu_\oplus(t)$ all object functions have a common rate of Hölder continuity and a common Hölder constant, is weaker than assumptions that have been required in classical FDA (see for example Müller *et al.* (2006)), where one deals with real-valued random functions. Assumption 10 is a bound on the covering number of the object metric space and is satisfied by common instances for random objects that include the examples that were discussed at the end of Section 3.2.

We write $\omega(\cdot)$ for Ω-valued functions $[0,1]\to\Omega$ and define

$$V_n(\omega, s) = \frac{1}{n}\sum_{i=1}^{n}[d^2\{X_i(s), \omega(s)\} - d^2\{X_i(s), \mu_\oplus(s)\}],$$

$$V(\omega, s) = E[d^2\{X(s), \omega(s)\} - d^2\{X(s), \mu_\oplus(s)\}].$$

Here $\hat{\mu}_\oplus(\cdot)$ is the minimizer of $V_n(\omega, s)$ and $\mu_\oplus(\cdot)$ is the minimizer of $V(\omega, s)$. We refer to $\hat{\mu}_\oplus(\cdot)$, $\mu_\oplus(\cdot)$ and $\omega(\cdot)$ as $\hat{\mu}_\oplus$, $\mu_\oplus$ and ω in what follows. To derive the rate of convergence of $\hat{\mu}_\oplus$, we first obtain a bound for $E\{\sup_{s\in[0,1]}\sup_{d_\infty(\omega,\mu_\oplus)<\delta}|V_n(\omega,s) - V(\omega,s)|\}$ for small $\delta > 0$, where $d_\infty(\omega, \mu_\oplus) = \sup_{s\in[0,1]} d\{\omega(s), \mu_\oplus(s)\}$. For this, we define function classes

$$\mathcal{F}_\delta = \{f_{\omega,s}(x) = d^2\{x(s), \omega(s)\} - d^2\{x(s), \mu_\oplus(s)\} : s\in[0,1], d_\infty(\omega, \mu_\oplus) < \delta\}. \tag{13}$$

It is easy to see that an envelope function for this class is the constant function $F(x) = 2M\delta$, where M is the diameter of Ω. The L^2-norm of this envelope function is $\|F\|_2 = 2M\delta$. By theorem 2.14.2 of van der Vaart and Wellner (1996) we have

$$E\left\{\sup_{s\in[0,1]}\sup_{d_\infty(\omega,\mu_\oplus)<\delta}|V_n(\omega,s) - V(\omega,s)|\right\} \leqslant \frac{2M\delta\ J_{[]}\{1, \mathcal{F}_\delta, L^2(P)\}}{\sqrt{n}}, \tag{14}$$

where $J_{[]}\{1, \mathcal{F}_\delta, L^2(P)\}$ is the bracketing integral of the function class $\mathcal{F}_\delta$:

$$J_{[]}\{1, \mathcal{F}_\delta, L^2(P)\} = \int_0^1 \sqrt{(1 + \log[N\{\varepsilon\|F\|_2, \mathcal{F}_\delta, L^2(P)\}])}\,\mathrm{d}\varepsilon.$$

Here $N\{\varepsilon\|F\|_2, \mathcal{F}_\delta, L^2(P)\}$ is the minimum number of balls of radius $\varepsilon\|F\|_2$ required to cover the function class $\mathcal{F}_\delta$ under the $L^2(P)$ norm. Lemma 1 provides the behaviour of the bracketing integral of the function class $\mathcal{F}_\delta$: a key step for the proof of theorem 3.

Lemma 1. Under assumptions 4, 9 and 10, it holds for the function class $\mathcal{F}_\delta$ as defined in expression (13) that $J_{[]}\{1, \mathcal{F}_\delta, L^2(P)\} = O\{\sqrt{\log(1/\delta)}\}$ as $\delta\to 0$.

Theorem 3. Under assumptions 3, 4 and 8–11,

$$\sup_{s\in[0,1]} d\{\hat{\mu}_\oplus(s), \mu_\oplus(s)\} = O_P\left[\left\{\frac{\sqrt{\log(n)}}{n}\right\}^{1/\beta_2}\right].$$

Setting $\hat{D}_i(t) = d\{X_i(t), \hat{\mu}_\oplus(t)\}$, an application is the convergence of the estimated Fréchet scores

$$\hat{\beta}_{ik} = \int_0^1 \hat{D}_i(t)\hat{\phi}_k(t)\mathrm{d}t. \tag{15}$$

Corollary 2. Under assumptions 3–5 and 8–11,

$$|\hat{\beta}_{ik} - \beta_{ik}| = O_P\left[n^{-1/2} + \left\{\frac{\sqrt{\log(n)}}{n}\right\}^{1/\beta_2}\right].$$

Following the widely adopted convention, we assume throughout that true and estimated eigenfunctions are aligned in the sense that $\langle\hat{\phi}_j, \phi_j\rangle \geqslant 0$, as the scores are identifiable only up to a change in sign.

5. Simulations

We illustrate the utility of the proposed methods through simulations for two settings. In the first setting, the space Ω consists of univariate probability distributions equipped with the 2-Wasserstein metric and, in the second setting, Ω consists of networks with fixed number of nodes, represented as graph adjacency matrices and equipped with the Frobenius metric.

5.1. Time-varying probability distributions

We generated random samples of sizes $n = 25$, 50, 100 of 'distribution'-valued curves on the domain $[0, 1]$, where, for each $t \in [0, 1]$, $X_i(t)$ is a normal distribution with mean $\mu_i(t)$ and variance $\sigma_i^2(t)$ with

$$\mu_i(t) = 1 + U_i\phi_1(t) + V_i\phi_3(t), \qquad U_i \sim N(0, 12), \quad V_i \sim N(0, 1),$$

$$\sigma_i(t) = 3 + W_i\phi_2(t) + Z_i\phi_3(t), \qquad W_i \sim \sqrt{72}\,U(0, 1), \quad Z_i \sim \sqrt{9}\,U(0, 1),$$

with $\phi_1(t) = (t^2 - 0.5)/0.3416$, $\phi_2(t) = \sqrt{3}t$ and $\phi_3(t) = (t^3 - 0.3571t^2 - 0.6t + 0.1786)/0.0895$, where ϕ_1, ϕ_2 and ϕ_3 are orthonormal on $[0, 1]$. We use the 2-Wasserstein metric for the distribution space Ω. For these specifications, the metric autocovariance function is

$$C(s, t) = 12\phi_1(s)\phi_1(t) + 6\phi_2(s)\phi_2(t) + 1.75\phi_3(s)\phi_3(t), \tag{16}$$

and $\phi_1(\cdot)$, $\phi_2(\cdot)$ and $\phi_3(\cdot)$ are the first three eigenfunctions.

We applied the proposed method to estimate the metric autocovariance operator to the simulated data and obtained its eigenvalues and eigenfunctions. Denoting the estimated metric autocovariance surface and the estimated jth eigenvalue and eigenfunction obtained at the kth simulation run respectively by $\hat{C}_k(s, t)$ $\hat{\lambda}_{j,k}$ and $\hat{\phi}_{j,k}$, we computed mean integrated squared errors (MISE)

$$\mathrm{MISE}(C) = \frac{1}{100}\sum_{k=1}^{100}\int_0^1\int_0^1 \{\hat{C}_k(s, t) - C(s, t)\}^2\mathrm{d}s\mathrm{d}t,$$

$$\mathrm{MISE}(\phi_j) = \frac{1}{100}\sum_{k=1}^{100}\int_0^1 \{\hat{\phi}_{j,k}(s) - \phi_j(s)\}^2\mathrm{d}s,$$

$$\mathrm{MISE}(\lambda_j) = \frac{1}{100}\sum_{k=1}^{100}(\hat{\lambda}_{j,k} - \lambda_j)^2.$$

Fig. 1 shows the true and estimated metric autocovariance surfaces and their eigenfunctions for one randomly chosen simulation run for $n = 25$ and $n = 100$. We find that the method proposed has negligible bias as the sample size increases. The MISEs are reported in Table 1 and are seen to decrease with increasing sample sizes.

To illustrate the nature of the simulated random density trajectories, four density-valued random functions that are part of a sample of density-valued random functions as generated in one Monte Carlo run are displayed in Fig. 2, reflecting variation in means and variances of the Gaussian distributions as a function of time for the four selected subjects. The estimated object FPCs, i.e. the Fréchet integrals of the object curves along the first two eigenfunctions for a Monte Carlo run are in Fig. 3 for sample size 50. Here the first object FPCs reflect variation in location of the distributions and the second object FPCs variation in the variance of the distributions, which is what we expect in view of how these data were generated. The object FPCs are found to be useful for discovering the underlying modes of variation for distributions as functional random objects.

5.2. *Time-varying networks*

In each iteration, we generated random samples of sizes $n = 25, 50, 100$ of time-varying random networks with 10 nodes each in the time interval $[0, 1]$. For generating the edge weights, we

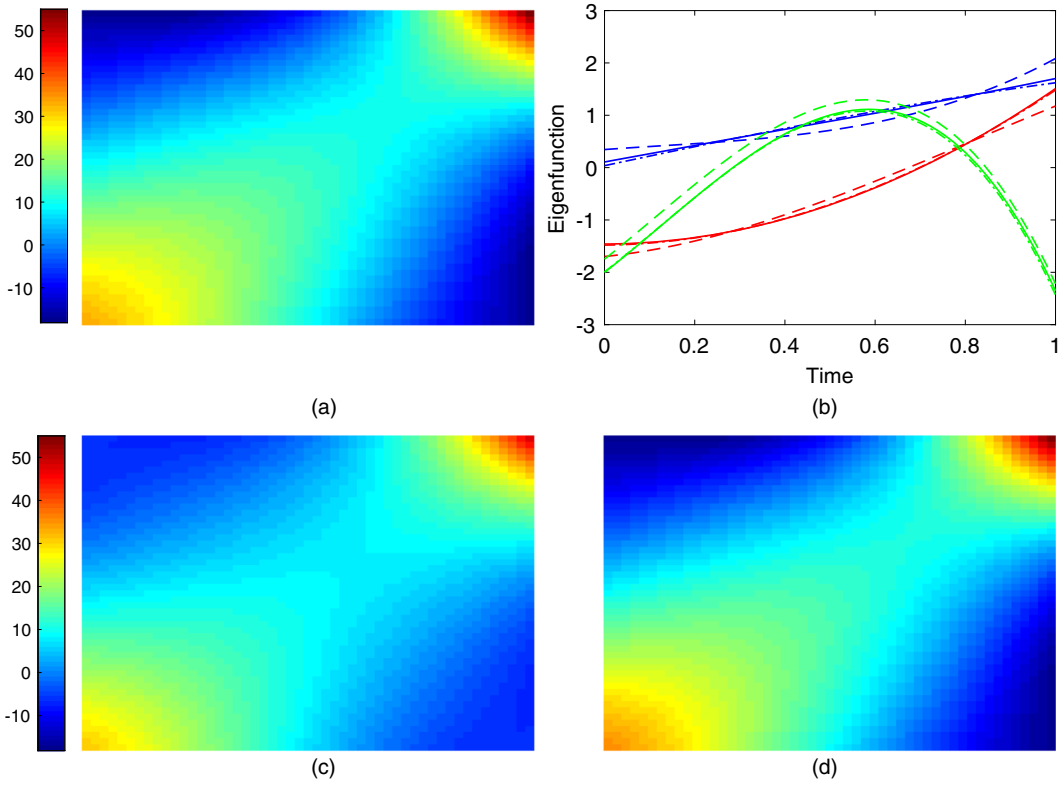

Fig. 1. (a) True and (c) ($n = 25$), (d) ($n = 100$) estimated metric autocovariance contour plots (10) for simulation with distributions as functional random objects: (b) ——, – – –, · – · –, first eigenfunction; ——, – – –, · – · –, second eigenfunction; ——, – – –, · – · –, third eigenfunction, depicting true (——, ——, ——) and estimated (– – –, – – –, – – –, $n = 25$; · – · –, · – · –, · – · –, $n = 100$) eigenfunctions

Table 1. MISEs for the estimators of the metric autocovariance kernel C and the eigenfunctions ϕ_1 and ϕ_2 in dependence on sample size when the functional random objects are distributions

n	C	ϕ_1	ϕ_2	ϕ_3	λ_1	λ_2	λ_3
25	12.2709	1.2841	1.5351	1.5202	10.8634	7.8471	3.5798
50	8.6598	0.0504	0.0201	0.0030	0.9748	0.6482	0.3680
100	4.0697	0.0158	0.0084	0.0047	0.1239	0.0607	0.0314

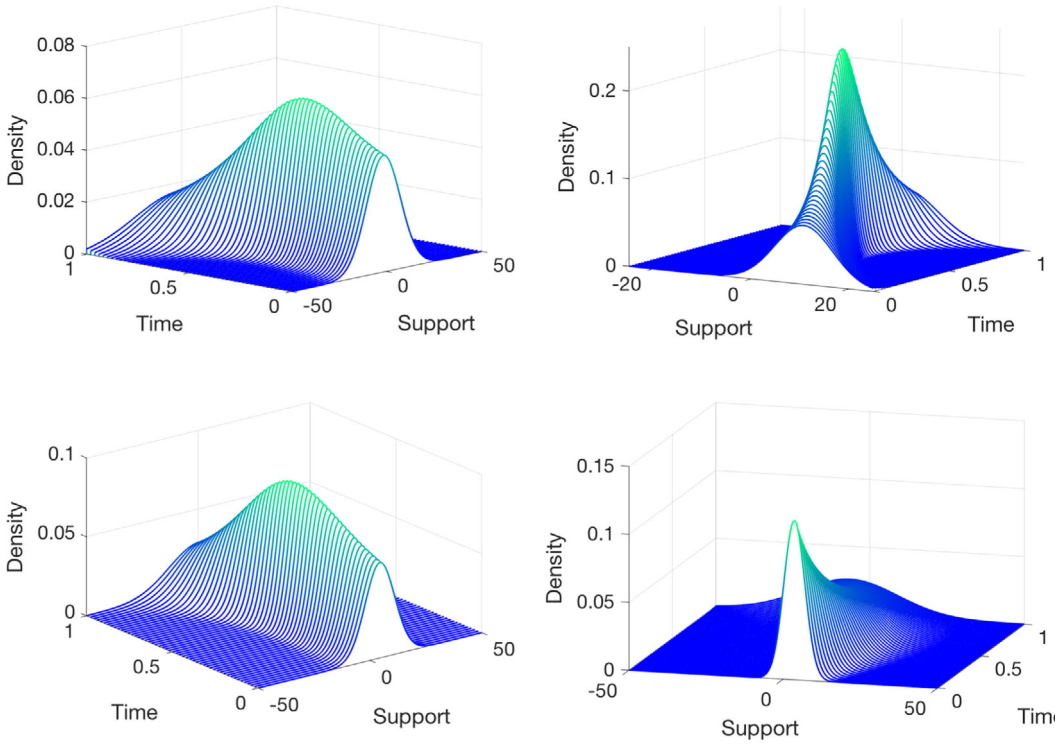

Fig. 2. Four randomly chosen observations of density-valued trajectories, selected from the sample of distributions generated by one of the Monte Carlo runs: the densities are plotted as a function of time

followed the model that is described below. We assumed that the network has two communities: the first five nodes belonging to one community and the second five nodes to the other. For each fixed time t, the edge weights within each community and also those between the communities are the same, where the latter are smaller than the within-community edge weights. Formally, if $p_{1,i}(t)$, $p_{2,i}(t)$ and $p_{12,i}(t)$ denote the edge weight at time $t \in [0,1]$ for the first community, the second community and between communities, for the ith network-valued curve we generated

$$\left.\begin{aligned} p_{1,i}(t) &= 0.5 + U_i\phi_1(t) + V_i\phi_3(t), \\ p_{2,i}(t) &= 0.5 + W_i\phi_2(t) + Z_i\phi_3(t), \\ p_{12,i}(t) &= 0.1. \end{aligned}\right\} \tag{17}$$

Here the U_i, V_i, W_i and Z_i were generated from the uniform distributions $U(0, 0.4)$, $U(0, 0.1)$,

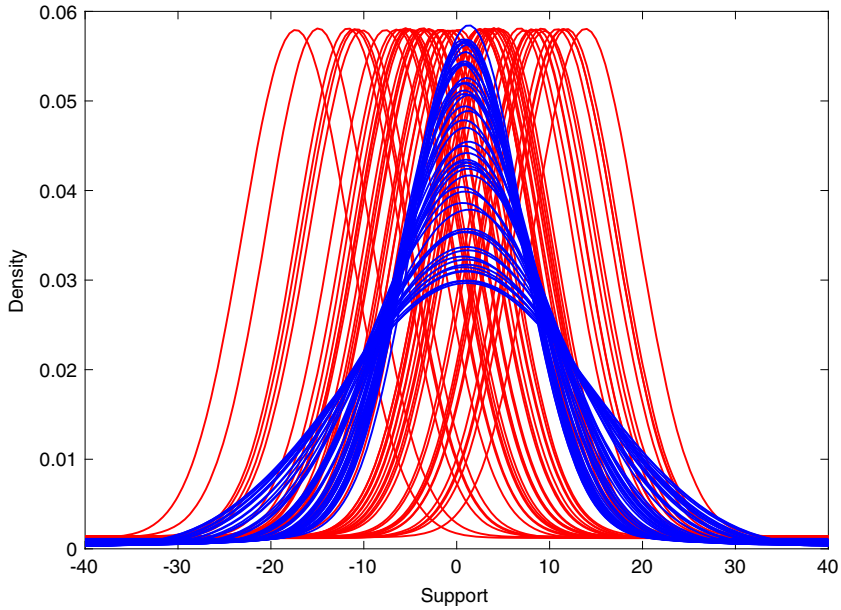

Fig. 3. Estimated Fréchet integrals (11) for the first (———) and second (———) eigenfunction for the sample elements, for simulated time-varying probability distributions

Table 2. MISEs for the estimators of the metric autocovariance kernel C and eigenfunctions–eigenvalues $\phi_j, \lambda_j, j = 1, 2, 3$, in dependence on sample size for samples of functional random objects that correspond to time-varying networks

n	C	ϕ_1	ϕ_2	ϕ_3	λ_1	λ_2	λ_3
25	0.0039	0.0039	0.0017	0.0007	0.0025	0.0010	0.0007
50	0.0017	0.0093	0.0046	0.0021	0.0007	0.0003	0.0001
100	0.0010	0.0130	0.0063	0.0028	0.0001	0.0001	0.0001

$U(0, 0.3)$ and $U(0, 0.1)$ respectively. The functions $\phi_1(t)$, $\phi_2(t)$ and $\phi_3(t)$ are orthonormal polynomials derived from Jacobi polynomials $P_n^{(\alpha,\beta)}(x)$ (Totik, 2005), which are classical orthogonal polynomials for $\alpha, \beta > 1$. They are orthogonal with respect to the basis $(1+x)^\beta(1-x)^\alpha$ on $[-1, 1]$. With a suitable change of basis, one can obtain a version of the Jacobi polynomials on $[0, 1]$ which are orthonormal with respect to the weight function $x^\beta(1-x)^\alpha$ on $[0, 1]$. We selected $\phi_1(t)$, $\phi_2(t)$ and $\phi_3(t)$ as

$$\phi_j(t) = \frac{P_{2j}^{(4,3)}(2t-1)t^{1.5}(1-t)^2}{[\int_0^1 \{P_{2j}^{(4,3)}(2t-1)\}^2 t^{1.5}(1-t)^2 \mathrm{d}t]^{1/2}} \qquad \text{for } j = 1, 2, 3.$$

The weighted networks are represented as graph adjacency matrices with the Frobenius metric. Here the true metric autocovariance function is

$$C(s,t) = 0.266\phi_1(s)\phi_1(t) + 0.15\phi_2(s)\phi_2(t) + 0.0417\phi_3(s)\phi_3(t), \tag{18}$$

and $\phi_1(\cdot)$, $\phi_2(\cdot)$ and $\phi_3(\cdot)$ are the first three eigenfunctions.

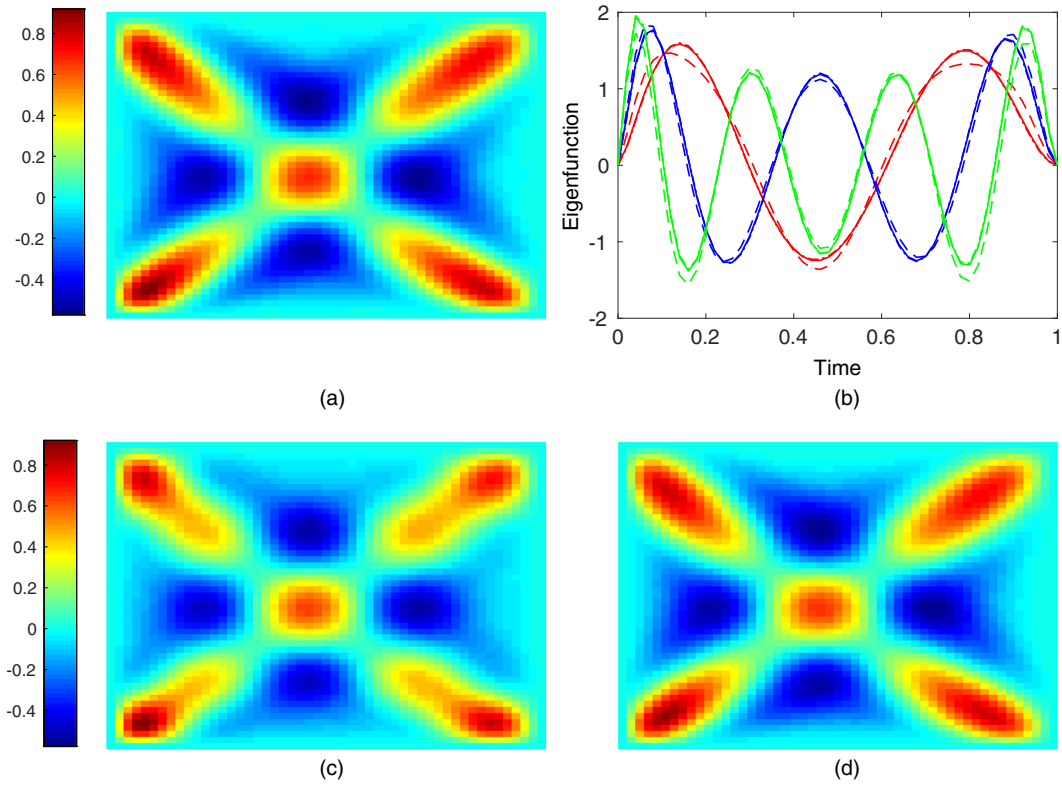

Fig. 4. (a) True and (c) ($n = 25$), (d) ($n = 100$) estimated metric autocovariance contour plots (10) for simulation with networks as functional random objects: (b) ———, – – –, – · –, first eigenfunction; ———, – – –, – · –, second eigenfunction; ———, – – –, – · –, third eigenfunction, depicting true (———, ———, ———) and estimated (– – –, – – –, – – –, $n = 25$; – · –, – · –, – · –, $n = 100$) eigenfunctions

We estimated the metric autocovariance operator from the simulated data and obtained its eigenfunctions for different sample sizes. Fig. 4 displays the true and estimated metric autocovariance surfaces and corresponding eigenfunctions for one randomly chosen simulation run for $n = 25$ and $n = 100$. The MISEs were computed as described for the previous simulation setting and are reported in Table 2. They decrease with increasing sample sizes. The method proposed is seen to work very well.

The object FPCs were obtained by using Fréchet integrals (11). For visualization they are presented as 'networks.mov' in the on-line supplementary materials. In the movie the leftmost plot corresponds to Fréchet integrals for the first eigenfunction which, as expected because of the true model, shows variation only in the edge weights of the first community. The middle plot corresponds to Fréchet integrals for the second eigenfunction and indicates variation only in the edge weights of the second community. The rightmost plot corresponds to Fréchet integrals for the third eigenfunction where variation in both the first and the second community edge weights can be discerned.

6. Data applications

6.1. Mortality data

The human mortality database provides life table data differentiated by gender and is avail-

able from www.mortality.org. Currently the mortality database contains life table data for 37 countries spanning over five decades. One can obtain histograms from life tables and smooth these with local least squares to obtain estimated probability density functions for age at death. We carried this out for the age interval $[0, 80]$ years. The mortality data can then be viewed as samples of time-varying univariate probability distributions, for a sample of 32 countries, where the time axis corresponds to calendar years between 1960 and 2009 and the observation that is made at each calendar year for each country corresponds to the age-at-death distribution for that year. We included the 32 countries which had complete records over the entire calendar period. For each country and year, we used the Hades package that is available from https://stat.ucdavis.edu/hades/ for smoothing the histograms and used bandwidth $= 2$ to obtain the age-at-death densities. For illustration, the time-varying age-at-death distributions represented as density functions for the age interval $[0, 80]$ years and indexed by calendar year are displayed for four selected countries, the USA, Ukraine, Russia and Portugal, for males in Fig. 5 and for females in Fig. 6.

Choosing the 2-Wasserstein metric for the probability distributions space, the estimated metric autocovariance surfaces for males and females can be inspected in Fig. 7 and the eigenfunctions of the corresponding autocovariance operators in Fig. 8. The autocovariance functions and eigenfunctions indicate that there are systematic differences between males and females.

The resulting object FPCs, i.e. the Fréchet integrals, are illustrated in Fig. 9 for the first two eigenfunctions. The object FPCs are distributions that are represented as densities for males and females. The eastern European countries that are included in the database, namely

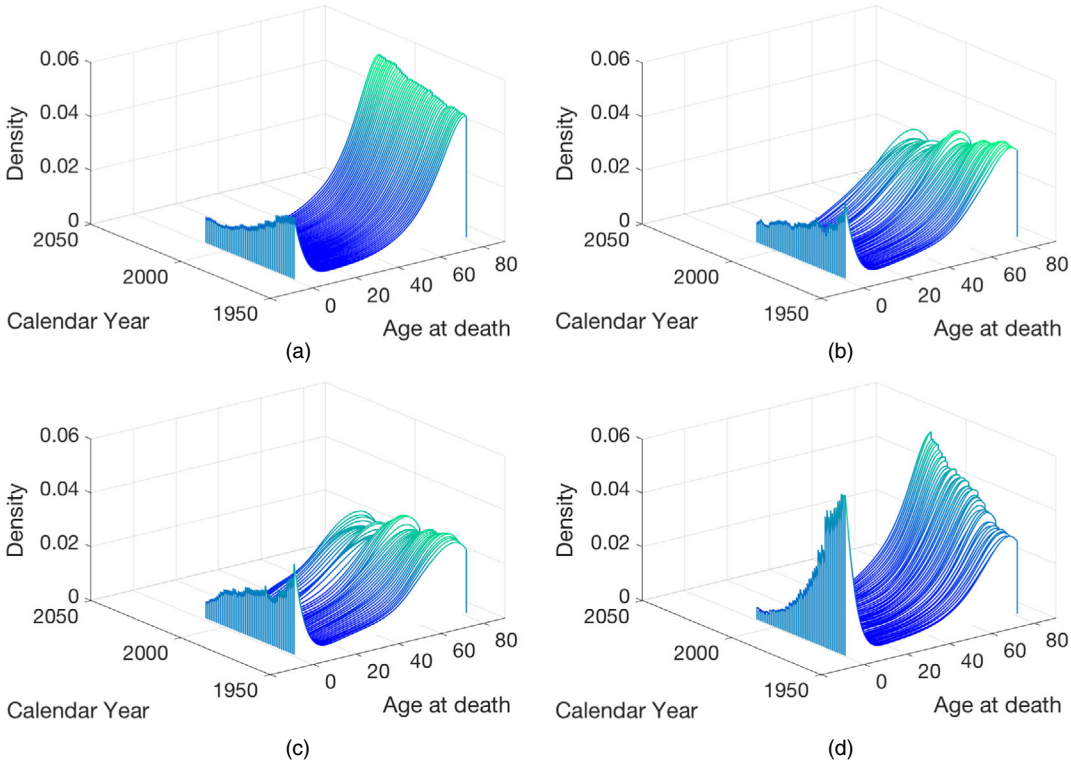

Fig. 5. Time-varying age-at-death density functions for the age interval $[0, 80]$ years for males in (a) the USA, (b) Ukraine, (c) Russia and (d) Portugal

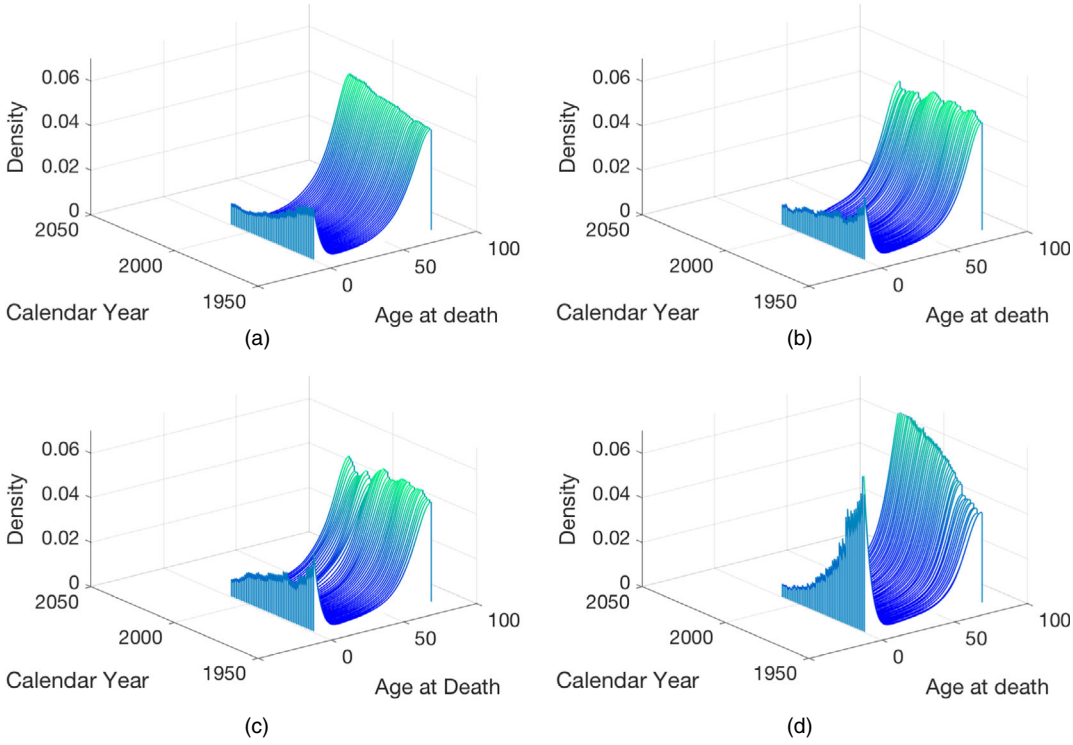

Fig. 6. Time-varying age-at-death density functions for the age interval $[0, 80]$ years for females in (a) the USA, (b) Ukraine, (c) Russia and (d) Portugal

Belarus, Bulgaria, the Czech Republic, Hungary, Latvia, Lithuania, Poland, Slovakia, Ukraine, Russia and Estonia, underwent major political upheaval due to the end of Communist rule in these regions during the period between the late 1980s and early 1990. This is reflected in clear distinctions between the eastern European countries (red) and the rest (blue) in the Fréchet integrals for the males but much less so for the females, which indicates that particularly male mortality was affected by the political upheavals.

The sample Fréchet mean function at a particular calendar year corresponds to the sample average of the quantile functions of the various countries at that calendar year and is illustrated in the movies 'mean_males.mov' and 'mean_females.mov' in the on-line supplementary materials. Fig. 10 illustrates the scalar FPCs, i.e. the Fréchet scores for the second eigenfunction plotted against the Fréchet scores for the first eigenfunction for males and females. Russia is an outlier for the first eigenfunction for males and Portugal is an outlier for the second eigenfunction, even though it does not belong to the above list of eastern European countries. One could speculate that this might be related to the fact that Portugal in 1974 moved to a democratic government after four decades of authoritarian dictatorship. Figs 5 and 6 suggest higher infant mortality for both males and females in Portugal during the earlier era. Another interesting observation is that the order of outliers is reversed for females, as Russia turns out to be an outlier for females for the second eigenfunction and Portugal for the first. The plots of the Fréchet scores against each other indicate there are clear distinctions between the two groups of countries and Portugal.

6.2. Time-varying networks for New York taxi data

The New York City Taxi and Limousine Commission provides records on pick-up and drop-

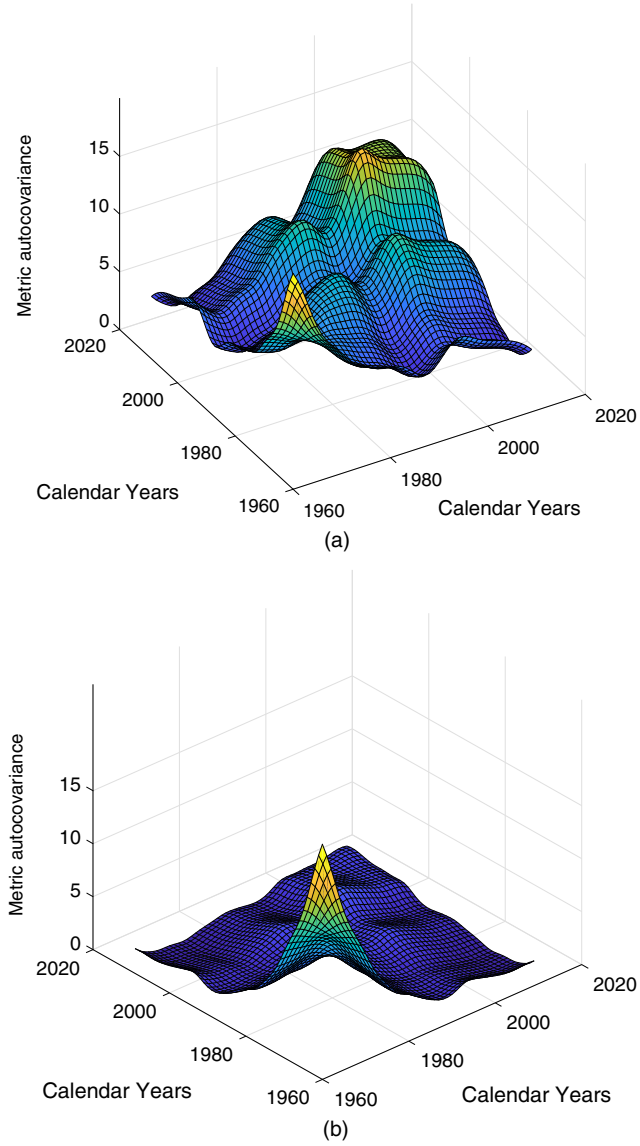

Fig. 7. Estimated metric autocovariance surfaces (10) for (a) males and (b) females for densities as functional random objects, as obtained for the mortality data

off dates and times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types and driver-reported passenger counts for yellow and green taxis which are available from `http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml`. The time resolution of these data is of the order of seconds. Of interest are networks which represent how many people travelled between places of interest and the evolution of these networks during a typical day. To study this, we constructed samples of time-varying networks where the sample elements are the recordings for each day in the year 2016. Three days (January 23rd and 24th and March 13th) were excluded from the study because of incomplete records.

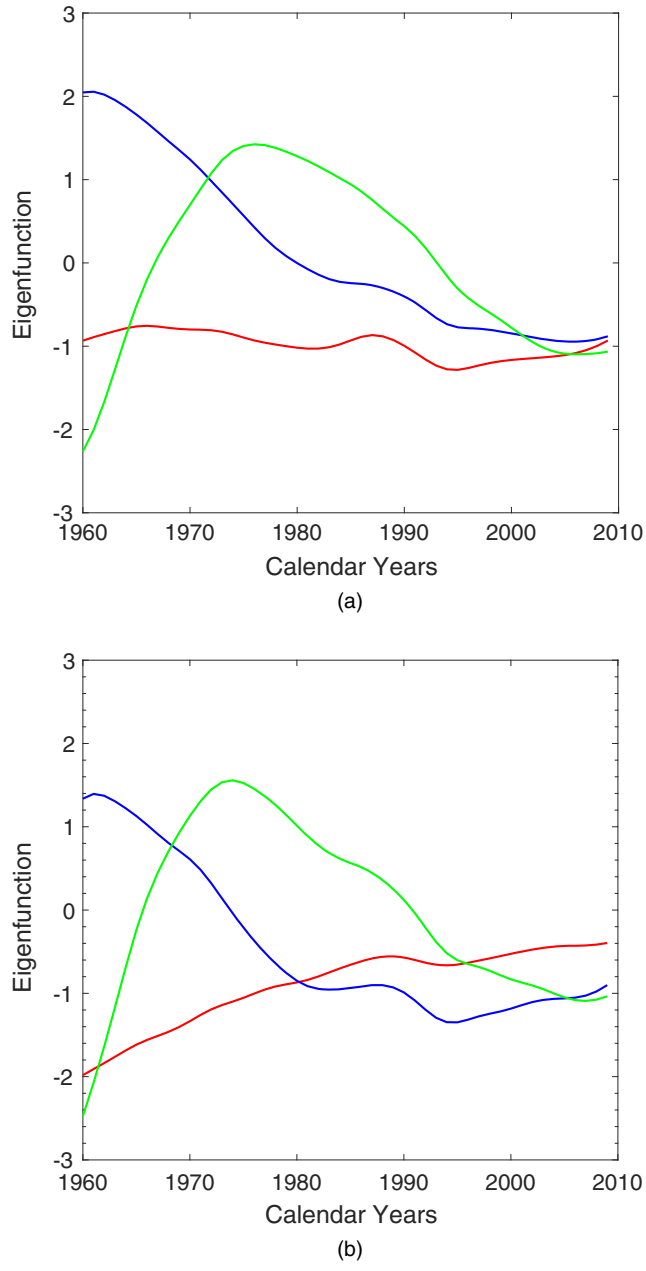

Fig. 8. Eigenfunctions of the estimated metric autocovariance surface for (a) males (———, 78.32%; ———, 16.06%; ———, 3.47%) and for (b) females (———, 65.17%; ———, 23.03%; ———, 7.92%) for the mortality data

We focus on the Manhattan area, which has the highest traffic, and split the area according to the provided location shape files into 10 zones, which form the regions of interest. Details about the zones are in section S6.1 of the on-line supplement. Yellow taxis provide the predominant taxi service in Manhattan. We divided each day into 20-min intervals, and for each interval

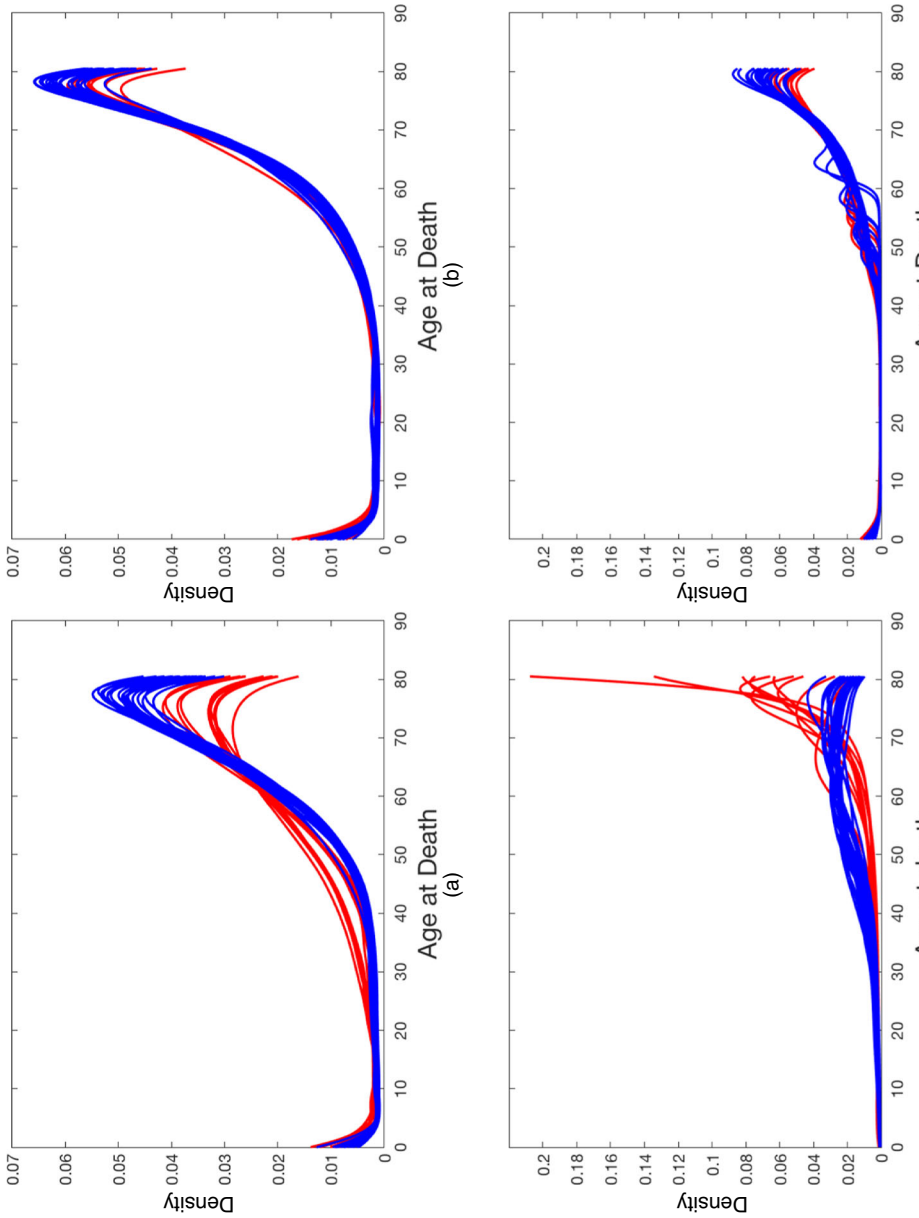

Fig. 9. Fréchet integrals (11) for the first eigenfunction for (a) males and (b) females and for the second eigenfunction for (c) males and (d) females for the mortality data: ——, eastern European countries, namely Belarus, Bulgaria, the Czech Republic, Hungary, Latvia, Lithuania, Poland, Slovakia, Ukraine, Russia and Estonia; ——, the other countries

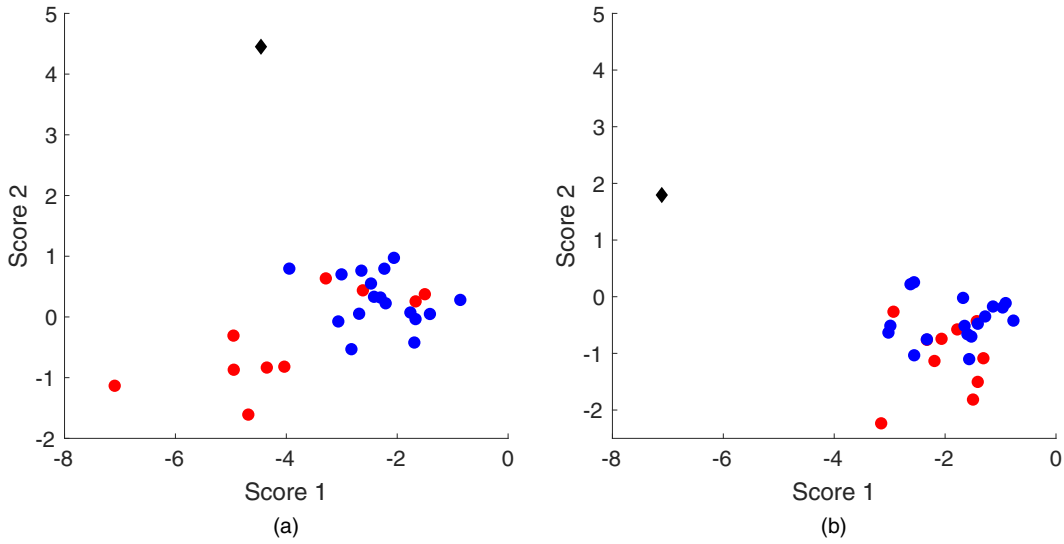

Fig. 10. Fréchet scores (15) for the first and the second eigenfunctions plotted against each other for (a) males and for (b) females for the mortality data: ◆, Portugal; ●, eastern European countries, namely Belarus, Bulgaria, the Czech Republic, Hungary, Latvia, Lithuania, Poland, Slovakia, Ukraine, Russia and Estonia; ●, the other countries

constructed a network with nodes corresponding to the 10 selected zones and edge weights representing the number of people who travelled between the zones connecting the edges within the 20-min interval. The edge weights were normalized by the maximum edge weight for each day so that they lie in [0, 1]. We thus have a time-varying network for each of the 363 days in 2016 for which complete records are available, where the time points at which the network-valued functions are evaluated correspond to the 20-min intervals of a 24-h day. The observations at each time point correspond to a 10-dimensional graph adjacency matrix which characterizes the network between the 10 zones of Manhattan for the particular 20-min interval.

We choose the Frobenius metric as metric between the graph adjacency matrices. The sample Fréchet mean function at a particular time point therefore corresponds to the sample average of the graph adjacency matrices of 363 networks corresponding to different days for that time point. It is illustrated in the movie 'mean_NY.mov' in the on-line supplementary materials. Fig. 11 illustrates the estimated autocovariance function and associated eigenfunctions. The plots of the Fréchet scores for the second, third and fourth eigenfunction against the scores for the first eigenfunction can be found in Fig. 12, where the blue dots correspond to Mondays–Thursdays, the green dots to Fridays and the red dots to Saturdays and Sundays. Several interesting patterns emerge: weekdays and weekends form clearly distinguishable clusters. Special holidays show similar patterns to those of weekends. Several outliers can be identified by using the projection scores for the eigenfunctions, which turn out to be special days: for the first eigenfunction, the outliers correspond to New Year's day and November 6th, 2016, which is the day when daylight saving ends. March 13th, 2016, is the day on which the daylight saving begins but was excluded as it did not have complete records. For the second eigenfunction, an outlying point is Independence Day, July 4th, 2016, and, for the third eigenfunction, February 14th, 2016, which is Valentine's day. Another day that stands out is September 18th, 2016. On further investigation it was found that between September 17th and 19th, 2016, three bombs exploded and several unexploded bombs were found in

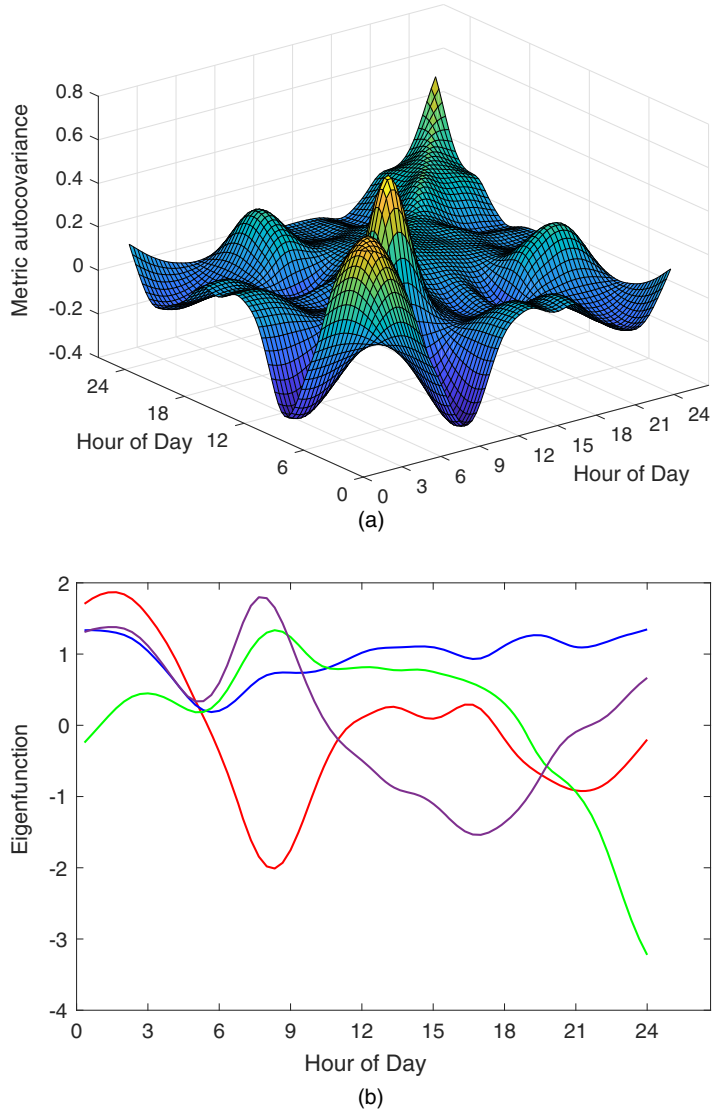

Fig. 11. (a) Estimated metric autocovariance surface (10) and (b) the corresponding eigenfunctions for the New York taxi data, viewed as time-varying networks: ——, 44.7%; ——, 32.72%; ——, 10.56%; ——, 8.42%

the New York metropolitan area (`https://en.wikipedia.org/wiki/2016_New_York_and_New_Jersey_bombings`).

We then repeated the analysis separately for three groups of days, namely the weekdays Monday–Thursday (group 1), Fridays and weekends (group 2) and holidays (group 3). We present the results in Fig. 17 (in the on-line supplement) and in several movies whose descriptions can be found in sections S3 and S5 of the on-line supplement.

6.3. World trade data

The Center for International Data at the University of California, Davis (`http://cid.econ.ucdavis.edu/nberus.html`), provides detailed documentation of United Nations trade

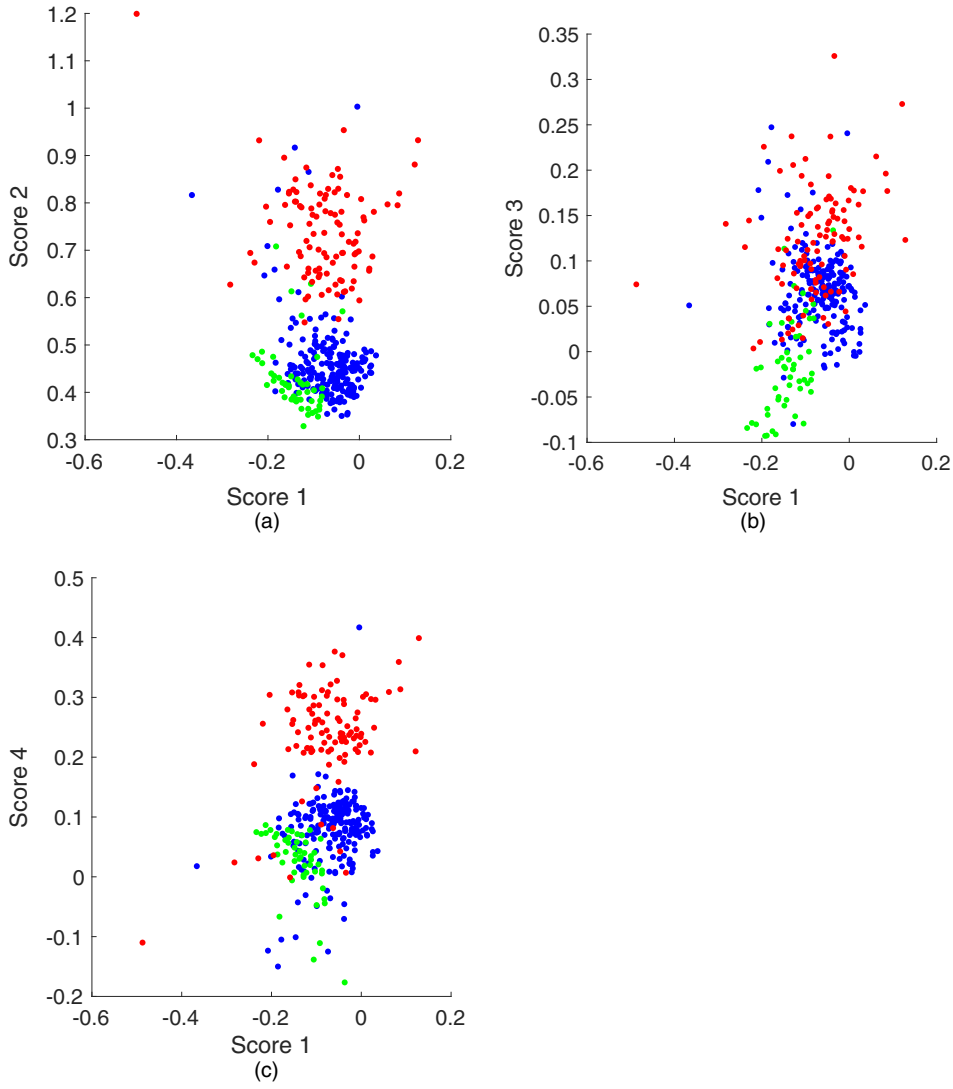

Fig. 12. Fréchet scores (15) for (a) the second, (b) third and (c) fourth eigenfunctions on the y-axis plotted against Fréchet scores for the first eigenfunction on the x-axis, for the New York taxi data: ●, Mondays–Thursdays; ●, Fridays; ●, Saturdays and Sundays

data for the years 1962–2000. The data set, which is publicly available from `www.nber.org`, contains bilateral trade data during this time period for several commodities and countries. We studied the time period 1970–1999 for 46 actively trading countries and the 26 most common types of commodity. The list of chosen countries and commodities can be found in section S6.2 of the on-line supplement. For each country, commodity and year, we represent current trade as the ratio of the amount of total trade, i.e. import–export value (in thousands of US dollars), to the amount of total trade recorded for the same commodity and country in the year 2000, yielding a 26-dimensional vector of trade ratios.

Viewing the countries as sampling units, we obtain for each country and calendar year t

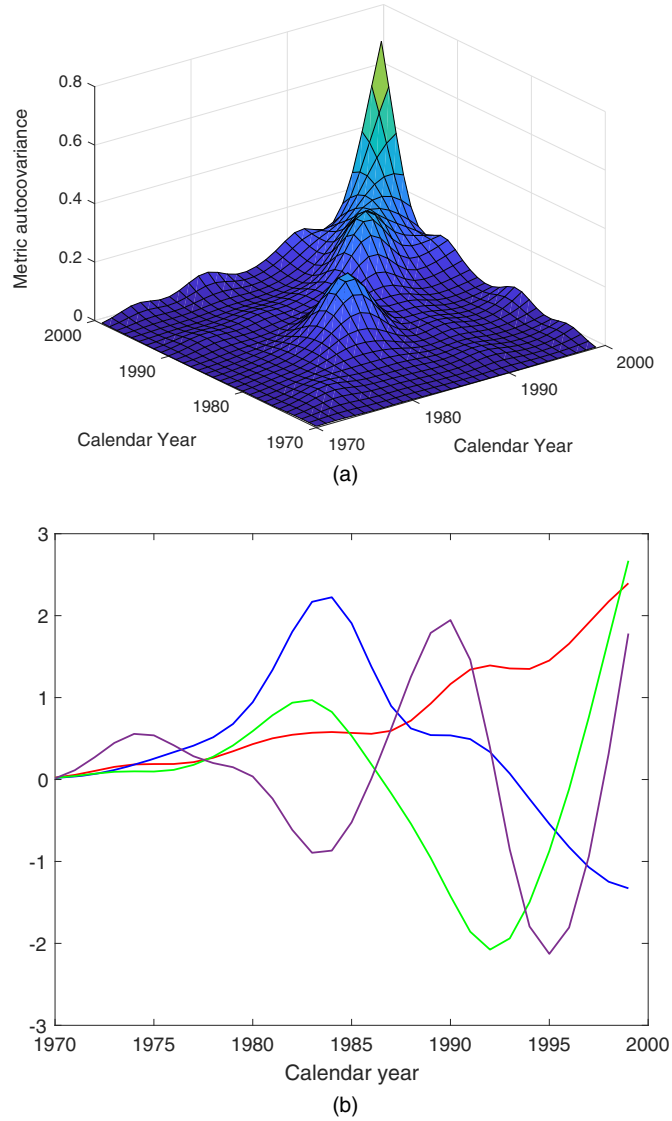

Fig. 13. (a) Estimated metric autocovariance surface and (b) corresponding first four eigenfunctions for the trade data: ———, 47.95%; ———, 20.67%; ———, 15.64%; ———, 6.79%

a (26×26)-dimensional raw covariance matrix of commodities trade ratios as $\tilde{\Sigma}(t) = (Q(t) - \bar{Q}(t))(Q(t) - \bar{Q}(t))^{\mathrm{T}}$, where $Q(t)$ is the country-specific 26-dimensional vector of commodities trade ratios for year t and the mean vector $\bar{Q}$ is obtained as a cross-sectional average over all 46 countries. These raw time-varying raw covariances were then smoothed by using local Fréchet regression with a Gaussian kernel (Petersen and Müller, 2019b; Petersen *et al.*, 2019) to obtain samples of smooth time-varying 26-dimensional covariance matrices between the components of commodities trade for each of the 46 countries over the time period 1970–2000, yielding time-varying covariance matrices over the time period 1970–2000 as functional random objects.

When adopting the Frobenius metric, the sample Fréchet mean function at calendar year t corresponds to the sample average of the smoothed covariance across 46 countries for year t. Fig.

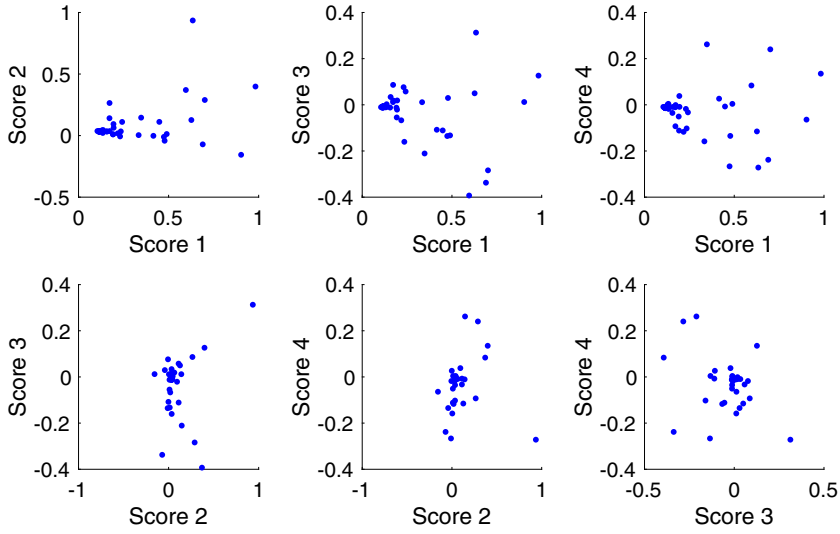

Fig. 14. Fréchet scores for various eigenfunctions plotted against each other for the trade data

13(a) illustrates the estimated metric autocovariance function and Fig. 13(b) its eigenfunctions. The metric autocovariance and its eigenfunctions provide insights about world trade patterns over the time period 1970–1999. The first eigenfunction represents increased variability due to overall expansion in world trade over the years from 1970 to 1999. The slope of the first eigenfunction is more gradual before 1985 but increases sharply starting from 1985, stagnates a little around 1990 and then again picks up. This can be connected to the boom in world trade towards the last decade of the new millennium. The second eigenfunction corresponds to a contrast before 1990 and after 1990. The peak in the second eigenfunction between 1980 and 1985 could be related to a major economic downturn caused by recession affecting several countries in the data set during the early 1980s. The recession began in the USA in 1981 and continued until 1982 and affected many of the developed western countries. The third eigenfunction captures effects of the early 1990s recession, which compared with the 1980s recession was much milder.

In Fig. 14, the Fréchet scores for the first four eigenfunctions are plotted against each other. Thailand and Egypt have high Fréchet scores for the first eigenfunction and Saudi Arabia ranks the highest for the second eigenfunction. Chile, Israel, Hong Kong and Bulgaria turn out to figure prominently in the third eigenfunction. Further visualization can be found in section S4 of the on-line supplement, including a movie that is described in section S5 of the on-line supplement that demonstrates the object FPCs.

7. Discussion

We propose an extension of functional data methods to the case of functional random objects. The basis of our approach is metric covariance: a novel covariance measure for paired metric-space-valued data. Eigenfunctions of the metric covariance operator for time-varying object data aid in creating a version of object FPCA, where the object FPCs in the metric space Ω are obtained as Fréchet integrals, which are a general and versatile concept. Alternatively, components of variation can be quantified by Fréchet scores, which are real numbers. For the precursor problem, where we have non-functional time-varying object data, i.e. we have observations for just one random object function over time, methods for metric-space-valued regression have

been considered previously (Steinke *et al.*, 2010; Faraway, 2014; Petersen and Müller, 2019b), often under the special assumption that the regression responses are on a Riemannian manifold (Shi *et al.*, 2009; Fletcher, 2013; Hinkle *et al.*, 2012; Su *et al.*, 2012; Yuan *et al.*, 2012; Cornea *et al.*, 2017). However, the more general object function case, which is characterized by samples of random functions that are object valued, is considerably more challenging, as the absence of a linear structure in the object space both globally and locally imposes serious limitations on the methods that can be applied.

The tools that we propose here for functional random objects, namely metric covariance, the metric autocovariance operator and its eigenfunctions, the Fréchet integrals and the Fréchet scores, make it possible to obtain compact summaries, visualizations and interpretations of the observed samples of time-varying object data that in themselves are highly complex and difficult to quantify. These tools can provide insights into the patterns of variability of the object trajectories, as we demonstrated in the simulations and data examples. The quantification of functional random objects can also be used for other tasks. For example the object FPCs that we introduce reside in the object space and can serve as responses for a regression model, where predictors are Euclidean vectors and responses are random object trajectories, which are summarized by these object FPCs. Implementing such a regression approach is analogous to the principal component approach for function-to-function regression (Yao *et al.*, 2005b). Various regression models can then be implemented through Fréchet regression (Petersen and Müller, 2019b). For the case where functional random objects feature as predictors in a regression setting, one can employ the vector of Fréchet scores that summarize each random object trajectory as predictors. The ensuing regression, classification and clustering models will be interesting topics for future research.

A core challenge that one faces when modelling and analysing samples of random object trajectories is that, in contrast with the situation for real-valued processes, we cannot expect to represent object-valued processes in terms of an analogue to the Karhunen–Loève expansion, because of the lack of a linear structure in the object space Ω. In some special cases such expansions are possible, e.g. through a transformation method, whenever random objects can be transformed to a linear space, as exemplified for the case of objects that are probability distributions (Petersen and Müller, 2016b) or for the case of Riemannian manifold-valued objects (Dai and Müller, 2018). Apart from such special cases, it is an open problem whether more general useful representations of functional random objects can be found. Another open problem is inference for such data, e.g. comparing two groups or testing for structural features of autocovariance. Here the metric autocovariance operator that we introduce in this paper and also the Fréchet mean function could prove useful for the extension of tests that have been considered for real-valued functional data (for some recent examples, see Aston *et al.* (2017), Constantinou *et al.* (2017), Chen and Lynch (2018) and Choi and Reimherr (2018)). These and many other open problems in this area indicate that there is ample potential for future research.

Acknowledgements

We thank the reviewers for their comprehensive and helpful comments that led to numerous improvements and new perspectives. This research was supported by National Science Foundation grant DMS-1712864 and National Institutes of Health grant 5UH3OD023313.

References

Anirudh, R., Turaga, P., Su, J. and Srivastava, A. (2017) Elastic functional coding of Riemannian trajectories. *IEEE Trans. Pattn Anal. Mach. Intell.*, **39**, 922–936.

Arcones, M. A. and Giné, E. (1993) Limit theorems for U-processes. *Ann. Probab.*, **21**, 1494–1542.
Aston, J. A., Pigoli, D. and Tavakoli, S. (2017) Tests for separability in nonparametric covariance operators of random surfaces. *Ann. Statist.*, **45**, 1431–1461.
Belkin, M. and Niyogi, P. (2002) Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems* (eds T. G. Dietterich, S. Becker and Z. Gharamani), pp. 585–591. Cambridge: MIT Press.
Berrendero, J., Justel, A. and Svarc, M. (2011) Principal components for multivariate functional data. *Computnl Statist. Data Anal.*, **55**, 2619–2634.
Bigot, J., Gouet, R., Klein, T. and López, A. (2017) Geodesic PCA in the Wasserstein space by convex PCA. *Ann. Inst. H. Poincaré* B, **53**, 1–26.
Billingsley, P. (1968) *Convergence of Probability Measures*. New York: Wiley.
Bosq, D. (2000) *Linear Processes in Function Spaces: Theory and Applications*. New York: Springer.
Castro, P. E., Lawton, W. H. and Sylvestre, E. A. (1986) Principal modes of variation for processes with continuous sample curves. *Technometrics*, **28**, 329–337.
Chen, K., Delicado, P. and Müller, H.-G. (2017) Modelling function-valued stochastic processes, with applications to fertility dynamics. *J. R. Statist. Soc.* B, **79**, 177–196.
Chen, K. and Lynch, B. (2018) A test of weak separability for multi-way functional data, with application to brain connectivity studies. *Biometrika*, **105**, 815–831.
Chen, K. and Müller, H.-G. (2012) Modeling repeated functional observations. *J. Am. Statist. Ass.*, **107**, 1599–1609.
Chiou, J.-M., Chen, Y.-T. and Yang, Y.-F. (2014) Multivariate functional principal component analysis: a normalization approach. *Statist. Sin.*, **24**, 1571–1596.
Chiou, J.-M. and Li, P.-L. (2007) Functional clustering and identifying substructures of longitudinal data. *J. R. Statist. Soc.* B, **69**, 679–699.
Chiou, J.-M., Yang, Y.-F. and Chen, Y.-T. (2016) Multivariate functional linear regression and prediction. *J. Multiv. Anal.*, **146**, 301–312.
Choi, H. and Reimherr, M. (2018) A geometric approach to confidence regions and bands for functional parameters. *J. R. Statist. Soc.* B, **80**, 239–260.
Claeskens, G., Hubert, M., Slaets, L. and Vakili, K. (2014) Multivariate functional half-space depth. *J. Am. Statist. Ass.*, **109**, 411–423.
Constantinou, P., Kokoszka, P. and Reimherr, M. (2017) Testing separability of space-time functional processes. *Biometrika*, **104**, 425–437.
Cornea, E., Zhu, H., Kim, P. and Ibrahim, J. G. (2017) Regression models on Riemannian symmetric spaces. *J. R. Statist. Soc.* B, **79**, 463–482.
Dai, X. and Müller, H.-G. (2018) Principal component analysis for functional data on Riemannian manifolds and spheres. *Ann. Statist.*, **46**, 3334–3361.
Dai, X., Müller, H.-G. and Yao, F. (2017) Optimal Bayes classifiers for functional data and density ratios. *Biometrika*, **104**, 545–560.
Dauxois, J., Pousse, A. and Romain, Y. (1982) Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *J. Multiv. Anal.*, **12**, 136–154.
Dong, J. J., Wang, L., Gill, J. and Cao, J. (2018) Functional principal component analysis of glomerular filtration rate curves after kidney transplant. *Statist. Meth. Med. Res.*, **27**, 3785–3796.
Dryden, I. L., Koloydenko, A. and Zhou, D. (2009) Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Statist.*, **3**, 1102–1123.
Dubey, P. and Müller, H.-G. (2019) Fréchet analysis of variance for random objects. *Biometrika*, **106**, 805–821.
Dubin, J. A. and Müller, H.-G. (2005) Dynamical correlation for multivariate longitudinal data. *J. Am. Statist. Ass.*, **100**, 872–881.
Faraway, J. J. (2014) Regression for non-Euclidean data using distance matrices. *J. Appl. Statist.*, **41**, 2342–2357.
Fletcher, P. T. (2013) Geodesic regression and the theory of least squares on Riemannian manifolds. *Int. J. Comput. Visn*, **105**, 171–185.
Fréchet, M. (1948) Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré*, **10**, 215–310.
Ginestet, C. E., Li, J., Balachandran, P., Rosenberg, S. and Kolaczyk, E. D. (2017) Hypothesis testing for network data in functional neuroimaging. *Ann. Appl. Statist.*, **11**, 725–750.
Hinkle, J., Muralidharan, P., Fletcher, P. T. and Joshi, S. (2012) Polynomial regression on Riemannian manifolds. In *Computer Vision* (eds A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato and C. Schmid), pp. 1–14. New York: Springer.
Horvath, L. and Kokoszka, P. (2012) *Inference for Functional Data with Applications*. New York: Springer.
Hsing, T. and Eubank, R. (2015) *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. New York: Wiley.
Jacques, J. and Preda, C. (2014) Model-based clustering for multivariate functional data. *Computnl Statist. Data Anal.*, **71**, 92–106.
Jones, M. C. and Rice, J. A. (1992) Displaying the important features of large collections of similar curves. *Am. Statistn*, **46**, 140–145.

Kleffe, J. (1973) Principal components of random variables with values in a separable Hilbert space. *Statistics*, **4**, 391–406.
Kruskal, J. (1964) Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, **29**, 115–129.
Lin, L., St Thomas, B., Zhu, H. and Dunson, D. B. (2017) Extrinsic local regression on manifold-valued data. *J. Am. Statist. Ass.*, **112**, 1261–1273.
Liu, S., Zhou, Y., Palumbo, R. and Wang, J.-L. (2016) Dynamical correlation: a new method for quantifying synchrony with multivariate intensive longitudinal data. *Psychol. Meth.*, **21**, 291–308.
Lyons, R. (2013) Distance covariance in metric spaces. *Ann. Probab.*, **41**, 3284–3305.
Monnig, N. D. and Meyer, F. G. (2018) The resistance perturbation distance: a metric for the analysis of dynamic networks. *Discr. Appl. Math.*, **236**, 347–386.
Müller, H.-G., Stadtmüller, U. and Yao, F. (2006) Functional variance processes. *J. Am. Statist. Ass.*, **101**, 1007–1018.
Nolan, D. and Pollard, D. (1987) U-processes: rates of convergence. *Ann. Statist.*, **15**, 780–799.
Nolan, D. and Pollard, D. (1988) Functional limit theorems for U-processes. *Ann. Probab.*, **16**, 1291–1298.
Opgen-Rhein, R. and Strimmer, K. (2006) Inferring gene dependency networks from genomic longitudinal data: a functional data approach. *Revstat*, **4**, 53–65.
Park, S. Y. and Staicu, A.-M. (2015) Longitudinal functional data analysis. *Stat*, **4**, 212–226.
Petersen, A., Deoni, S. and Müller, H.-G. (2019) Fréchet estimation of time-varying covariance matrices from sparse data, with application to the regional co-evolution of myelination in the developing brain. *Ann. Appl. Statist.*, **13**, 393–419.
Petersen, A. and Müller, H.-G. (2016a) Fréchet integration and adaptive metric selection for interpretable covariances of multivariate functional data. *Biometrika*, **103**, 103–120.
Petersen, A. and Müller, H.-G. (2016b) Functional data analysis for density functions by transformation to a Hilbert space. *Ann. Statist.*, **44**, 183–218.
Petersen, A. and Müller, H.-G. (2019a) Wasserstein covariance for multiple random densities. *Biometrika*, to be published.
Petersen, A. and Müller, H.-G. (2019b) Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.*, **47**, 691–719.
Pigoli, D., Aston, J. A., Dryden, I. L. and Secchi, P. (2014) Distances and inference for covariance operators. *Biometrika*, **101**, 409–422.
Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*, 2nd edn. New York: Springer.
Robert, P. and Escoufier, Y. (1976) A unifying tool for linear multivariate statistical methods: the RV-coefficient. *Appl. Statist.*, **25**, 257–265.
Schoenberg, I. J. (1938) Metric spaces and positive definite functions. *Trans. Am. Math. Soc.*, **44**, 522–536.
Seguy, V. and Cuturi, M. (2015) Principal geodesic analysis for probability measures under the optimal transport metric. In *Advances in Neural Information Processing Systems* (eds C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama and R. Garnett), pp. 3312–3320. Cambridge: MIT Press.
Sejdinovic, D., Sriperumbudur, B., Gretton, A. and Fukumizu, K. (2013) Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.*, **41**, 2263–2291.
Shi, X., Styner, M., Lieberman, J., Ibrahim, J. G., Lin, W. and Zhu, H. (2009) Intrinsic regression models for manifold-valued data. In *Medical Image Computing and Computer-assisted Intervention* (eds G. Zhong, D. Hawkes, D. Rueckert, A. Noble and C. Taylor), pp. 192–199. New York: Springer.
Steinke, F., Hein, M. and Schölkopf, B. (2010) Nonparametric regression between general Riemannian manifolds. *SIAM J. Imgng Sci.*, **3**, 527–563.
Su, J., Dryden, I. L., Klassen, E., Le, H. and Srivastava, A. (2012) Fitting smoothing splines to time-indexed, noisy points on nonlinear manifolds. *Im. Visn Comput.*, **30**, 428–442.
Suarez, A. J. and Ghosal, S. (2016) Bayesian clustering of functional data using local features. *Baysn Anal.*, **11**, 71–98.
Székely, G. J. and Rizzo, M. L. (2017) The energy of data. *A. Rev. Statist. Appl.*, **4**, 447–479.
Székely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007) Measuring and testing dependence by correlation of distances. *Ann. Statist.*, **35**, 2769–2794.
Tavakoli, S., Pigoli, D., Aston, J. A. and Coleman, J. (2019) A spatial modeling approach for linguistic object data: analysing dialect sound variations across Great Britain. *J. Am. Statist. Ass.*, **114**, 1081–1096.
Totik, V. (2005) Orthogonal polynomials. *Surv. Approximn Theory*, **1**, 70–125.
van der Vaart, A. and Wellner, J. (1996) *Weak Convergence and Empirical Processes*. New York: Springer.
Verbeke, G., Fieuws, S., Molenberghs, G. and Davidian, M. (2014) The analysis of multivariate longitudinal data: a review. *Statist. Meth. Med. Res.*, **23**, 42–59.
Wang, J.-L., Chiou, J.-M. and Müller, H.-G. (2016) Functional data analysis. *A. Rev. Statist. Appl.*, **3**, 257–295.
Yao, F., Müller, H.-G. and Wang, J.-L. (2005a) Functional data analysis for sparse longitudinal data. *J. Am. Statist. Ass.*, **100**, 577–590.
Yao, F., Müller, H.-G. and Wang, J.-L. (2005b) Functional linear regression analysis for longitudinal data. *Ann. Statist.*, **33**, 2873–2903.
Yuan, Y., Zhu, H., Lin, W. and Marron, J. S. (2012) Local polynomial regression for symmetric positive definite matrices. *J. R. Statist. Soc.* B, **74**, 697–719.

Zhou, L., Huang, J. and Carroll, R. (2008) Joint modelling of paired sparse functional data using principal components. *Biometrika*, **95**, 601–619.

Discussion on the paper by Dubey and Müller

Shahin Tavakoli (*University of Warwick, Coventry*)

Overview

I congratulate Dr Dubey and Professor Müller for their inspiring and thought-provoking paper. The paper considers time-varying random objects, which fall into the remit of functional data analysis (FDA). FDA is interested in the analysis of data points that are complex, such as curves, images, shapes, trees, movies, spectra, sounds or covariance matrices or operators (e.g. Lu *et al.* (2014)). The space Ω (following the paper's notation) in which such data points lie falls (roughly) into four categories.

(a) Ω is a (separable) Hilbert space: distances between points, moving along specific directions and inner products are defined globally.
(b) Ω is a connected Riemannian manifold: distances between points are defined globally, but moving along specific directions and inner products are defined only locally.
(c) Ω is a Banach space: distances between points and moving along a specific direction are defined globally, but there is no inner product.
(d) Ω is a metric space: only distances between points are defined.

The backbone of FDA is functional principal component analysis (FPCA) (Hsing and Eubank, 2015), which is analogous to principal component (PC) analysis: if $X \in H$ is a random element of a real Hilbert space H with inner product '$\langle \cdot, \cdot \rangle$', norm '$\| \cdot \|$', $\mathbb{E}\|X\|^2 < \infty$, and, assuming that $\mathbb{E}(X) = 0$ for simplicity, FPCA solves the following iterative problem. For $k = 1, 2, \ldots,$ solve

$$v_k \in \arg \max_{v \in H : \|v\| = 1} \operatorname{var}(\langle v, X \rangle) \tag{19}$$

$$\text{such that} \quad \operatorname{cov}(\langle v_k, X \rangle, \langle v_j, X \rangle) = 0, \qquad j = 1, \ldots, k-1, \tag{20}$$

where the covariance constraint is omitted for $k = 1$. The v_ks are in the *same space as* X ($v_k \in H$) and are called PC loadings hereafter. The PC scores $\xi_k = \langle v_k X \rangle$, which are sometimes just called PCs or just scores, are *real valued*. Defining $u \otimes v$ by $(u \otimes v) f = \langle f, v \rangle u$, for $u, v, f \in H$, letting $C = \mathbb{E}[X \otimes X]$ be the covariance operator of X, and writing its spectral decomposition as $C = \Sigma_{k \geqslant 1} \lambda_k \varphi_k \otimes \varphi_k$, where the φ_ks are orthonormal eigenvectors of C, with associated eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant 0$, a standard result (Hsing and Eubank, 2015) is that the PC loadings satisfy $v_k = \varphi_k, \forall\, k$. Furthermore,

$$X = \sum_{k \geqslant 1} \xi_k \varphi_k, \tag{21}$$

with convergence in expected squared norm. Keeping only the first K terms in the sum in equation (21) yields $\Sigma_{k=1}^{K} \xi_k \varphi_k$, which is the solution to the problem of finding a subspace of dimension K onto which the orthogonal projection of X has maximal expected squared norm, or the best approximation of X in expected squared norm. An important special case is $H = L^2([0, 1], \mathbb{R})$, the Hilbert space of square integrable functions on $[0, 1]$, and $X(\cdot) \in L^2([0, 1], \mathbb{R})$ is mean square continuous (*continuity is key* here: this assumption seems to have been omitted in the paper), then equation (21) and the spectral decomposition of C hold with stronger conditions (e.g. $t \mapsto \varphi_k(t)$s are continuous), and they are called the Karhunen–Loève expansion of X and Mercer's theorem respectively. FPCA is popular because the PC scores can be plotted (e.g. pairs plots) for exploratory data analysis, and the PC loadings provide interpretation of PC scores (or of the modes of variations of X), and typically $X \approx \Sigma_{k=1}^{K} \xi_k \varphi_k$ is very good for $K = 2, 3$.

In all of the data categories (a)–(d) mentioned above, we can compute distances between any two points. If 'moving along a specific direction' and inner products are (locally) defined, then we can perform (local) FPCA. However, in the absence of an inner product, FPCA is not an option, as it hinges on linear

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Online supplement for "Functional models for time-varying random objects"'.

projections. Even if we are given φ_ks, equation (21) is meaningless if moving along a direction is not defined. Hence, for data points in a general metric space, it is not clear how to obtain PC scores, nor what 'modes of variations' means.

The major contribution of the paper is to propose *an* answer to these problems if data points are time-varying random objects, i.e. $X(t) \in \Omega$ for $t \in [0, 1]$, where (Ω, d) is a bounded metric space. To construct an alternative to FPCA, which hinges on the definition of covariance (which is not available in this set-up), the key observation of the paper is that, for random variables $U, V \in \mathbb{R}$ (and U', V' independent and identically distributed copies),

$$\begin{aligned}\operatorname{cov}(U, V) &= \mathbb{E}[(U - \mathbb{E}[U])(V - \mathbb{E}[V])] \\ &= \tfrac{1}{4}\mathbb{E}[d_{\mathrm{E}}^2(U, V') + d_{\mathrm{E}}^2(U', V) - 2d_{\mathrm{E}}^2(U, V)],\end{aligned}$$

where $d_{\mathrm{E}}(U, V) = |U - V|$. Based on this, if U, V are a random object in metric space (Ω, d), the paper introduces the *metric covariance* $\operatorname{cov}_\Omega(U, V)$ between U and V, defined as

$$\operatorname{cov}_\Omega(U, V) := \tfrac{1}{4}\mathbb{E}[d^2(U, V') + d^2(U', V) - 2d^2(U, V)],$$

and defines the *metric autocovariance kernel* $c(s, t) := \operatorname{cov}_\Omega\{X(s), X(t)\} \in \mathbb{R}$, which summarizes the co-variations of $X(\cdot)$. This kernel is symmetric but not non-negative definite (unlike 'standard' covariance kernels); however, if the metric (Ω, d) is of negative type (see the definition in the paper), then there is an abstract Hilbert space $(H, \langle\cdot\rangle)$ and an injective mapping $h : \Omega \to H$ such that

$$d(U, V) = \langle h(U) - h(V), h(U) - h(V)\rangle;$$

hence $\operatorname{cov}_\Omega(U, V) = \mathbb{E}[\langle h(U) - \mathbb{E}[h(U)], h(V) - \mathbb{E}[h(V)]\rangle]$, which implies that $c(s, t)$ is non-negative definite. Mercer's theorem is now in force *provided that* $c(s, t)$ is continuous, and the eigenfunction $\phi_k(\cdot) \in L^2([0, 1], \mathbb{R})$ of the integral operator induced by $c(\cdot, \cdot)$ is continuous, and they can be interpreted as the modes of variation of X. A decomposition such as equation (21) is still not possible (and note that ϕ_k is *not* a function $[0, 1] \to \Omega$), but the paper proposes two versions of PC scores.

(a) *Object FPCs*: summarize $X(\cdot)$ by $\psi_{\oplus}^k \in \Omega, k \in \{1, \ldots, K\}$ (scores in the metric space Ω).
(b) *Fréchet scores*: summarize $X(\cdot)$ by $\beta_k \in \mathbb{R}, k \in \{1, \ldots, K\}$ (scores are real valued).

Typically one hopes that $K = 2, 3$ provides reasonable summaries of X. These hinge on computing Fréchet integrals and Fréchet means, which are minimizers of some functional (described in the paper). This requires that Ω is a *complete* metric space (an assumption that seems to have been omitted in the paper). These new 'scores' enable summarizing each time-varying object by a few objects FPCs (in Ω), or by a few Fréchet scores (in $\mathbb{R}$), which helps exploratory analysis. The ϕ_ks, object FPCs and Fréchet scores are, in my opinion, the major contribution of the paper. Sample versions of them are proposed, and are backed up with consistency results.

Some critical thoughts

Going back to the second page of the paper, it is written that

> 'We aim here at identifying dominant directions of variation ... where the random objects are indexed by time and in a general metric space'.

So the goal was

(a) to deal with general metric spaces and
(b) to obtain dominant modes of variation (which hinges on measuring co-variations and maximum variations).

However, all the examples that were considered in the paper (Ω is a space of densities, networks or covariances) have a structure that is *much richer* than a metric space. If only a metric is available, $\operatorname{cov}_\Omega(s, t)$ and its eigenfunctions can be computed; however, object FPCs or Fréchet scores will *not* be (easily) computable because they require *solving minimization problems in the metric space* Ω.

Regarding (b), note crucially that, for random elements $U, V \in H$ with finite expected square norm,

$$\operatorname{cov}_\Omega(U, V) = \mathbb{E}[\langle U - \mathbb{E}[U], V - \mathbb{E}[V]\rangle] = \operatorname{tr}\{\operatorname{cov}(U, V)\} \neq \operatorname{cov}(U, V),$$

where $\mathrm{cov}(U, V)$ is the cross-covariance operator between U and V (a trace class operator on H). Although the trace is a good measure of magnitude *for self-adjoint non-negative definite* operators, it is *not* a good measure for general operators. Here is a toy example to illustrate this: let $U(t) \in \mathbb{R},\ t \in [0, \frac{1}{2}],$ be random with mean 0 and $\mathrm{cov}\{U(t), U(t)\} = 1,$ and define

$$X(t) = \begin{cases} \begin{pmatrix} U(t) \\ 0 \end{pmatrix}, & t \in [0, \frac{1}{2}], \\ \begin{pmatrix} 0 \\ U(t - \frac{1}{2}) \end{pmatrix}, & t \in (\frac{1}{2}, 1]. \end{cases}$$

We have that $\mathrm{cov}_\Omega\{X(t), X(t + \frac{1}{2})\} = 0$ for $t \in [0, \frac{1}{2}],$ but the true covariance matrix is

$$\mathrm{cov}\{X(t), X(t + \tfrac{1}{2})\} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and is non-zero. Hence metric covariance fails to measure the perfect linear association between $X(t)$ and $X(t + \frac{1}{2})$.

Let us now compare the usual PC loading in the case where $\Omega = H$, a separable Hilbert space. Then $X : [0, 1] \to H$, assuming that $\int_0^1 \mathbb{E}\|X(t)\|^2 \mathrm{d}t < \infty$ and glossing over measurability issues and other technicalities, the first PC loading of FPCA is the H-valued function $\varphi : [0, 1] \to H$ with $\int_0^1 \|\varphi(t)\|^2 \mathrm{d}t = 1$ that maximizes

$$\mathrm{var}(\langle \varphi, X \rangle) = \int_0^1 \langle \underbrace{\mathrm{cov}\{X(s), X(t)\}}_{\text{operator on } H} \varphi(t), \varphi(s) \rangle \mathrm{d}s\, \mathrm{d}t,$$

where we emphasize that H can be *any* separable Hilbert space, and $\mathrm{cov}\{X(s), X(t)\}$ is an *operator* on H for each fixed $s, t \in [0, 1]$.

The first eigenfunction of the metric autocovariance is the function $\phi : [0, 1] \to \mathbb{R}$ with $\int_0^1 \phi^2(t) \mathrm{d}t = 1$ that maximizes

$$\int c(s, t) \phi(s)\, \phi(t) \mathrm{d}s\, \mathrm{d}t = \int \phi(s) \phi(t) \mathrm{tr}[\mathrm{cov}\{X(s), X(t)\}] \mathrm{d}s\, \mathrm{d}t.$$

So, unless $H = \mathbb{R}$, the variance maximization interpretation of FPCA is lost with object FPCA.

Let us now look at the Fréchet scores for the case $\Omega = \mathbb{R}$. We directly obtain

$$\beta_k := \int d\{X(t), \mu_\oplus(t)\} \phi_k(t)\, \mathrm{d}t = \int |X(t) - \mu(t)| \phi_k(t)\, \mathrm{d}t;$$

hence Fréchet scores are not directly comparable with usual PC scores if $X \in L^2([0, 1], \mathbb{R})$.

Having said that, the methods that are proposed in the paper are universally applicable off-the-shelf methods, without the need to think about tailoring FPCA to each specific Ω.

Possible extensions

I propose two suggestions for extensions in addition to those mentioned in the paper.

(a) $t \in [0, 1]$ could be generalized to $t \in E$. For instance, we could have $E =$ Great Britain, $X(t)$ is a covariance matrix at location $t \in E$ and E is equipped with a non-Euclidean metric, as in Tavakoli *et al.* (2019).

(b) The implicit injection $h : \Omega \to H$ into an abstract Hilbert space H to work with full covariance operator $\mathcal{C} := \mathbb{E}[\{h(U) - \mathbb{E}[h(U)]\} \otimes \{h(V) - \mathbb{E}[h(V)]\}]$ on H, which is related to the metric covariance by

$$\mathrm{cov}_\Omega(U, V) = \mathrm{tr}(\mathcal{C}).$$

It is, however, unclear how to work with $\mathcal{C}$ since h is defined only implicitly.

To conclude, I believe that this is a great paper: it proposes novel and universally applicable approaches for exploratory data analysis for cases where FPCA is not directly applicable. It is an inspiring paper, and I am glad to propose the vote of thanks.

Dino Sejdinovic (*University of Oxford*)
I congratulate Dubey and Müller on several substantial conceptual and theoretical contributions which promise to lead to a widely applicable methodology. One of them is a new association measure between paired random objects in a metric space, termed *metric covariance*. I shall focus my discussion on this notion and on its relationship with other similar concepts which have previously appeared in the literature, including *distance covariance* (Székely *et al.*, 2007, 2009; Lyons, 2013) as well as its generalizations which rely on the formalism of reproducing kernel Hilbert spaces (RKHSs) (Sejdinovic *et al.*, 2013).

If (Ω, d) is a metric space such that d^2 is of negative type, then metric covariance takes the form

$$\mathrm{cov}_\Omega(X, Y) = \tfrac{1}{4}\mathbb{E}_{XY}[\mathbb{E}_{X'Y'}\{d^2(X, Y') + d^2(X', Y) - 2d^2(X, Y)\}].$$

A negative type of d^2 implies that we can find a Hilbert space $\mathcal{H}$ and a *feature map* $\phi : \Omega \to \mathcal{H}$ such that

$$d^2(X, Y) = \|\phi(X) - \phi(Y)\|^2_{\mathcal{H}},$$

and hence

$$\begin{aligned}\mathrm{cov}_\Omega(X, Y) &= \mathbb{E}_{XY}[\langle \phi(X) - \mathbb{E}_X\{\phi(X)\}, \phi(Y) - \mathbb{E}_Y\{\phi(Y)\}\rangle_{\mathcal{H}}] \\ &= \mathbb{E}_{XY}\{\langle \phi(X), \phi(Y)\rangle_{\mathcal{H}}\} - \mathbb{E}_{XY'}\{\langle \phi(x), \phi(Y')\rangle_{\mathcal{H}}\},\end{aligned}$$

corresponding to the discrepancy between expected inner products of features of X and Y under the joint and under the product of the marginals, measuring whether X and Y are on average more similar (as measured by feature maps) in the coupled or in the uncoupled regime. Importantly, metric covariance can take both positive and negative values.

In contrast, distance covariance takes the form

$$\begin{aligned}\Xi(X, Y) = {} & \mathbb{E}_{XY}[\mathbb{E}_{X'Y'}\rho_{\mathcal{X}}\{(X, X')\rho_{\mathcal{Y}}(Y, Y')\}] \\ & + \mathbb{E}_X[\mathbb{E}_{X'}\{\rho_{\mathcal{X}}(X, X')\}]\mathbb{E}_Y[\mathbb{E}'_Y\{\rho_{\mathcal{Y}}(Y, Y')\}] \\ & - 2\mathbb{E}_{XY}\{\mathbb{E}_{X'}\rho_{\mathcal{X}}(X, X')\mathbb{E}_{Y'}\rho_{\mathcal{Y}}(Y, Y')\},\end{aligned}$$

where $(\mathcal{X}, \rho_{\mathcal{X}})$ and $(\mathcal{Y}, \rho_{\mathcal{Y}})$ are two semimetric spaces of negative type (we allow random objects X and Y to take values in different domains) and semimetrics $\rho_{\mathcal{X}}$ and $\rho_{\mathcal{Y}}$ take the role of d^2. This expression appears less intuitive and without an obvious link to metric covariance.

An alternative way to introduce distance covariance, however, is through the lens of RKHSs. Consider random objects X and Y taking values on $\mathcal{X}$ and $\mathcal{Y}$ respectively, and any two positive definite kernel functions $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and $l : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ which are associated with RKHSs $\mathcal{H}_k$ and $\mathcal{H}_l$. Define the *cross-covariance operator* $\Sigma_{YX} : \mathcal{H}_k \to \mathcal{H}_l$ such that

$$\langle g, \Sigma_{YX} f\rangle_{\mathcal{H}_l} = \mathrm{cov}\{f(X), g(Y)\}, \qquad \forall\ f \in \mathcal{H}_k, \quad g \in \mathcal{H}_l.$$

The Hilbert–Schmidt independence criterion (HSIC), which is a notion (up to a constant factor) that is equivalent to distance covariance (Sejdinovic, 2013), is given by

$$\Xi(X, Y) = \|\Sigma_{YX}\|^2_{\mathrm{HS}}, \tag{22}$$

i.e. it is simply the squared Hilbert–Schmidt norm of feature space cross-covariance. For a broad class of choices of k and l—in particular, *characteristic kernels* (Sriperumbudur *et al.*, 2011)—the HSIC fully characterizes statistical dependence. These kernels include a widely used Gaussian kernel $k(x, x') = \exp(-\{1/(2\sigma^2)\}\|x - x'\|_2^2)$ and the Matérn family. Distance covariance can be recovered from the HSIC by considering 'distance'

$$d^2_{\mathcal{X}}(x, x') = k(x, x) + k(x', x') - 2k(x, x') \tag{23}$$

on $\mathcal{X}$ and similarly for $\mathcal{Y}$. Conversely, given any d^2 of negative type, we can construct the corresponding kernel

$$k(x,x') = \tfrac{1}{2}\{d^2(x,w) + d^2(x',w) - d^2(x,x')\} \tag{24}$$

where w is an arbitrary anchor point.

Is there also an RKHS interpretation of metric covariance? Recall that the domains of X and Y in this context coincide and are given by a metric space (Ω, d) with d^2 of negative type. We associate with it a positive definite kernel in equation (24) with RKHS $\mathcal{H}_k$ and define the cross-covariance operator Σ_{YX}. We claim that $\mathrm{cov}_\Omega(X,Y) = \mathrm{tr}(\Sigma_{YX})$. Indeed,

$$\begin{aligned}
\mathrm{tr}(\Sigma_{YX}) &= \mathrm{tr}[\mathbb{E}_{XY}\{k(\cdot,X)\} \otimes k(\cdot,Y) - \mathbb{E}_{XY}\{k(\cdot,X) \otimes k(\cdot,Y')\}] \\
&= \mathbb{E}_{XY}[\mathrm{tr}\{k(\cdot,X) \otimes k(\cdot,Y)\}] - \mathbb{E}_{XY'}[\mathrm{tr}\{k(\cdot,X) \otimes k(\cdot,Y')\}] \\
&= \mathbb{E}_{XY}\{\langle k(\cdot,X), k(\cdot,Y)\rangle_{\mathcal{H}_k}\} - \mathbb{E}_{XY'}\{\langle k(\cdot,X), k(\cdot,Y')\rangle_{\mathcal{H}_k}\} \\
&= \mathbb{E}_{XY}\{k(X,Y)\} - \mathbb{E}_{XY'}\{k(X,Y')\} \\
&= \tfrac{1}{2}[\mathbb{E}_{XY'}\{d^2(X,Y')\} - \mathbb{E}_{XY}\{d^2(X,Y)\}].
\end{aligned}$$

Recall that HSIC or distance covariance can be understood as

$$\Xi(X,Y) = \|\Sigma_{YX}\|^2_{\mathrm{HS}} = \mathrm{tr}(\Sigma_{YX}\Sigma_{XY}),$$

so indeed the two notions are closely related. To interpret the connection further, we can take a Mercer basis $\{\sqrt{\lambda_j} e_j\}_{j\in J}$ of $\mathcal{H}_k$. Then

$$\mathrm{cov}_\Omega(X,Y) = \sum_{j\in J} \lambda_j \langle e_j, \Sigma_{YX} e_j\rangle_{\mathcal{H}_k} = \sum_{j\in J} \lambda_j \mathrm{cov}\{e_j(X), e_j(Y)\},$$

i.e. metric covariance considers how evaluations at the *same* basis function covary and it can be zero if positive and negative covariances between basis function evaluations 'cancel out'. In contrast, the HSIC or distance covariance considers covariances of all pairs of basis function evaluations:

$$\Xi(X,Y) = \sum_{i\in J}\sum_{j\in J} \lambda_i\lambda_j \mathrm{cov}\{e_i(X), e_j(Y)\}^2.$$

We shall now consider some cases where the metric covariance is zero even though the variables are dependent. A straightforward example is to consider the case where there is dependence between X and Y but their feature representations live in orthogonal subspaces, e.g. if we take a *linear* kernel on $\mathbb{R}^2$ and $X = (Z,0), Y = (0,Z)$. A perhaps more interesting example, also in $\mathbb{R}^2$, is as follows: take $Z \sim \mathrm{Bern}(\frac{1}{2})$, and

$$X \sim \begin{cases} \mathcal{N}([-1,1], \sigma^2 I), & \text{if } Z = 0, \\ \mathcal{N}([1,-1], \sigma^2 I), & \text{if } Z = 1, \end{cases}$$

$$Y \sim \begin{cases} \mathcal{N}([-1,-1], \sigma^2 I), & \text{if } Z = 0, \\ \mathcal{N}([1,1], \sigma^2 I), & \text{if } Z = 1. \end{cases}$$

We have here coupled the 'mixing variable' so that X_1 and Y_1 are positively correlated, whereas X_2 and Y_2 are negatively correlated. It is readily shown, however, that $\|X - Y\| =^{\mathrm{d}} \|X - Y'\|$. Hence, metric covariance computed with *any radial kernel*, i.e. where $k(x,y)$ depends on x and y through $\|x-y\|$ only, which includes Gaussian and Matérn families known to be characteristic, will not be able to detect such dependence between X and Y. To be able to detect dependence we would require looking into individual dimensions, which may become impractical for higher dimensional problems.

In summary, although the authors demonstrate that distance covariance or the HSIC is not suitable for use in the developed framework of object functional principal component analysis, it is worth noting that metric covariance is a strictly weaker statistical dependence measure and it is possible that it misses certain types of multivariate associations. For a generic choice of metric, the corresponding feature map ϕ is defined implicitly and may not be straightforward to interpret whereas which forms of dependence are captured by metric covariance does depend on the form of ϕ and hence on the associated kernel k. Finally, we believe that the RKHS interpretation that is described here may give rise to different estimation methods of metric covariance and to its novel uses.

The vote of thanks was passed by acclamation.

Wicher Bergsma (*London School of Economics and Political Science*)

In an inspiring paper Dubey and Müller (DM) extend principal component analysis (PCA) to the case that observations are metric-valued functions. As an alternative, we develop a kernel PCA (KPCA) (Schölkopf *et al.*, 1998) approach, which we show is closely related to the DM approach. Whereas kernel principal components (KPCs) are simply defined, DM require added complexity in the form of 'object FPCs' and 'Fréchet scores'.

Kernel principal component analysis

Suppose that observations $X_1, \ldots, X_n$ take values in an arbitrary set $\mathcal{X}$. Let $\mathcal{H}$ be a Hilbert space and consider an embedding function $e: \mathcal{X} \to \mathcal{H}$. KPCA is essentially PCA on the embedded observations $e(X_1), \ldots, e(X_n)$, whose sample covariance kernel is

$$C = \frac{1}{n} \sum_{i=1}^{n} \{e(X_i) - \bar{e}\} \otimes \{e(X_i) - \bar{e}\} \in \mathcal{H} \otimes \mathcal{H}, \tag{25}$$

where $\bar{e} = n^{-1} \Sigma e(X_i)$. With φ_k the kth eigenvector of C, the kth KPC of $x \in \mathcal{X}$ is

$$\langle \varphi_k, e(x) \rangle_{\mathcal{H}}. \tag{26}$$

The embedding function need not be computed explicitly as the KPCs of the observations are given by the eigenvectors of the $n \times n$ kernel matrix with elements

$$K_{\mathcal{X}}(X_i, X_j) = \langle e(X_i) - \bar{e}, e(X_j) - \bar{e} \rangle_{\mathcal{H}}.$$

Kernel principal component analysis for observations that are metric-valued functions

If $(\mathcal{X}, d_{\mathcal{X}})$ is a metric space of negative type, an isometry $e: \mathcal{X} \to \mathcal{H}$ exists, and a corresponding unique centred kernel

$$\begin{aligned} K_{\mathcal{X}}(x, x') &= \langle e(x) - \bar{e}, e(x') - \bar{e} \rangle_{\mathcal{H}} \\ &= \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \{ \|e(x) - e(X_i)\|_{\mathcal{H}}^2 + \|e(x') - e(X_j)\|_{\mathcal{H}} - \|e(x) - e(x')\|_{\mathcal{H}}^2 - \|e(X_i)^2 - e(X_j)\|_{\mathcal{H}}^2 \} \\ &= \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \{ d_{\mathcal{X}}^2(x, X_i) + d_{\mathcal{X}}^2(x', X_j) - d_{\mathcal{X}}^2(x, x') - d_{\mathcal{X}}^2(X_i, X_j) \}. \end{aligned}$$

Let (Ω, d_Ω) be a metric space of negative type and let $\mathcal{X}$ be the space of Ω-valued functions over an index set $\mathcal{T}$ for which

$$d_{\mathcal{X}}^2(x, x') = \int_{\mathcal{T}} d_\Omega^2 \{x(t), x'(t)\} \, \mathrm{d}t$$

is finite. Then $(\mathcal{X}, d_{\mathcal{X}})$ is also of negative type, with corresponding kernel

$$\begin{aligned} K_{\mathcal{X}}(x, x') &= \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \{ d_{\mathcal{X}}^2(x, X_i) + d_{\mathcal{X}}^2(x', X_j) - d_{\mathcal{X}}^2(x, x') - d_{\mathcal{X}}^2(X_i, X_j) \} \\ &= \frac{1}{2} \int_{\mathcal{T}} [d_\Omega^2 \{x(t), X_i(t)\} + d_\Omega^2 \{x'(t), X_j(t)\} - d_\Omega^2 \{x(t), x'(t)\} - d_\Omega^2 \{X_i(t), X_j(t)\}] \, \mathrm{d}t \\ &= \int_{\mathcal{T}} K_\Omega \{x(t), x'(t)\} \mathrm{d}t. \end{aligned}$$

We reanalysed the mortality data (Human Mortality Database, 2019) using L_2-distance between mortality cumulative distribution functions. Fig. 15 shows the first two KPCs for both males and females (explained variances around 40% and 12% respectively in both cases). The separation between eastern European countries and others is clearer than in Fig. 10 of the authors' paper, which displays Fréchet scores.

Comparison between kernel principal component analysis and Dubey–Müller method

Let $e_\Omega: \Omega \to \mathcal{H}_\Omega$ be an isometry of (Ω, d_Ω) into a Hilbert space $\mathcal{H}_\Omega$. Assume that $\mathcal{H}$ consists of $\mathcal{H}_\Omega$-valued functions over $\mathcal{T}$ and

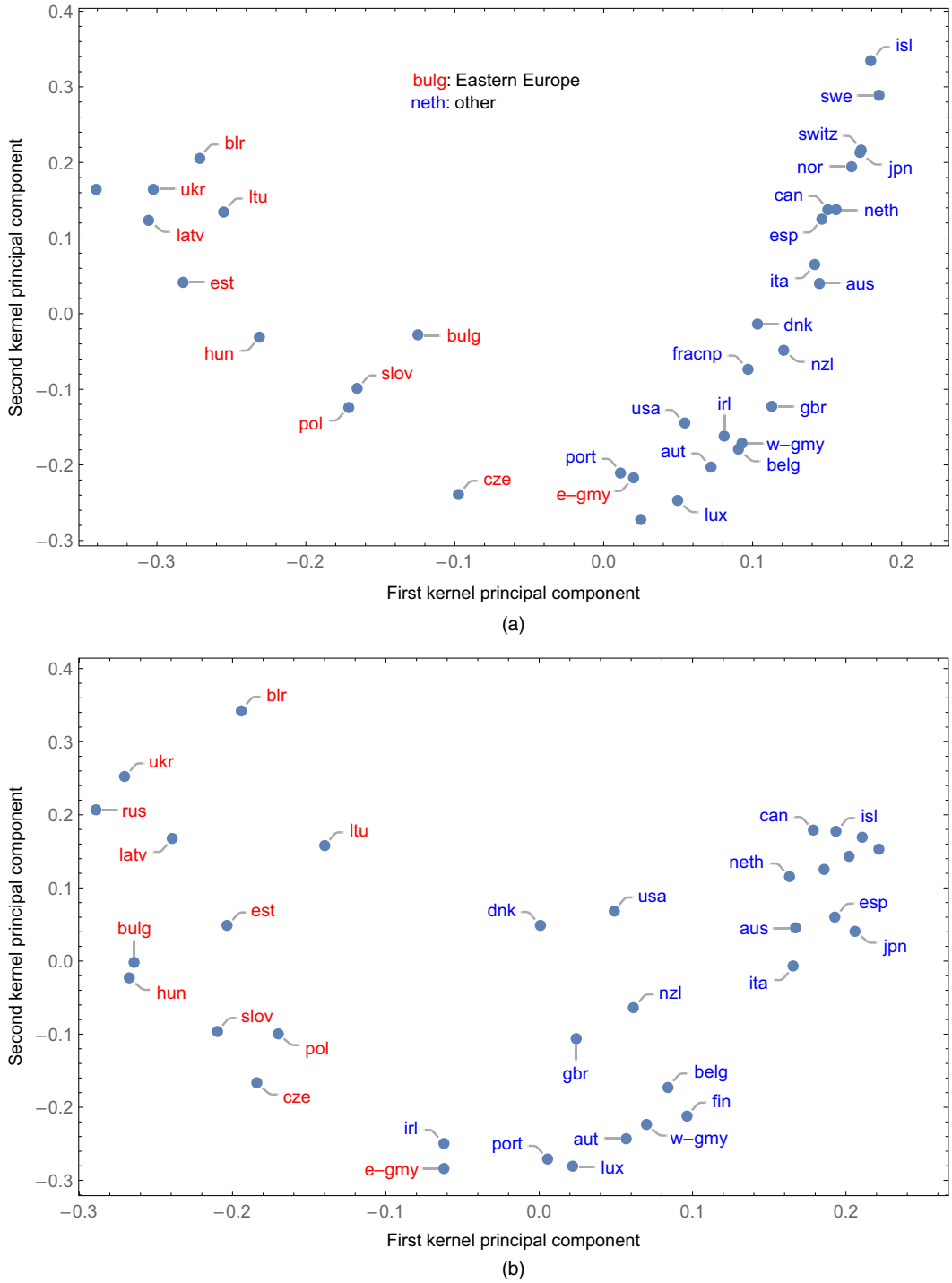

Fig. 15. First two KPCs for the mortality data: (a) males; (b) females

$$e(x)(t) = e_\Omega\{x(t)\}, \qquad \forall\, t \in \mathcal{T}, x \in \mathcal{X}.$$

Then C given by equation (25) is a *kernel-valued kernel* over $\mathcal{T} \times \mathcal{T}$, given for $s, t \in \mathcal{T}$ by

$$\begin{aligned} C(s,t) &= \frac{1}{n}\sum_{i=1}^{n}\{e(X_i)(s) - \bar{e}(s)\} \otimes \{e(X_i)(t) - \bar{e}(t)\} \\ &= \frac{1}{n}\sum_{i=1}^{n}[e_\Omega\{X_i(s)\} - \bar{e}(s)] \otimes [e_\Omega\{X_i(t)\} - \bar{e}(t)]. \end{aligned}$$

Instead, the DM method is based on the metric covariance kernel $C_{\mathrm{DM}} : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$:

$$C_{\mathrm{DM}}(s,t) = \frac{1}{n}\sum_{i=1}^{n}\langle e_\Omega\{X_i(s)\} - \bar{e}(s), e_\Omega\{X_i(t)\} - \bar{e}(t)\rangle.$$

We immediately have

$$C_{\mathrm{DM}}(s,t) = \mathrm{tr}\{C(s,t)\},$$

i.e. C_{DM} entails some information loss relative to C.

The eigenvectors of C are in the 'correct' space of $\mathcal{H}_\Omega$-*valued functions* over $\mathcal{T}$ and KPCs are naturally defined by equation (26). However, the eigenvectors of C_{DM} are in the 'wrong' space of *real-valued functions* over $\mathcal{T}$, and it is not obvious how metric-valued functional observations load on an eigenvector of C_{DM}. For this reason, DM needed to add some complexity with the new concepts named object FPCs and Fréchet scores, which can be avoided with our approach.

Yoav Zemel (*University of Cambridge*)

I congratulate Dubey and Müller for their thought-provoking contribution. The problem that is considered in this paper interweaves two challenging facets: firstly, the inherent infinite dimensionality of functional data; secondly, the non-linearity of the underlying object space.

One aspect that distinguishes functional data is the possible effect of time warping. For example, in the mortality data one clearly notices that the changes in the distribution are more rapid in Portugal than in the USA, so it seems that each country has its own timescale. Ignoring this so-called phase variation of the data is known to degrade the estimation of the covariance structure, and this may be the reason for the multimodal nature of the covariance surface in Fig. 7.

Another issue that deserves attention is the problem of discrete measurements. In practice, the data would only be observed on a grid, and some sort of smoothing needs to be performed, either at the level of the curves themselves or through smoothing of the covariance operator. The latter approach seems particularly appealing here, as the metric covariance can be estimated from the potentially sparse data.

I was also wondering about the negative-type assumption on the squared metric. For a given metric (e.g. sphere, shape space or Wasserstein metric on $\mathbb{R}^2$), is there a recipe to check whether this assumption holds? Some of the examples given on the seventh page are already Hilbertian, and the same is true for the simulations and the data analysis. The authors also mention the Procrustes metric, which has an interpretation of a Wasserstein distance in finite or infinite dimensions (Masarotto *et al.*, 2019).

Another instance of a problem combining phase variation and discrete measurements on infinite dimensional and non-linear spaces is the registration of spatial point processes (Zemel and Panaretos, 2019). The context is rather different, however: there, the data are elements on an infinite dimensional non-linear space (Ω, d), whereas in the present paper the data are functions from [0, 1] to the metric space (Ω, d).

Lastly, I would like to point out that the abstract Hilbert space H is not latent: it is simply the reproducing kernel Hilbert space of the kernel

$$\tfrac{1}{2}\{d^2(x, w_0) + d^2(y, w_0) - d^2(x, y)\}, \qquad x, y \in \Omega,$$

for some fixed $w_0 \in \Omega$ (Berg *et al.* (1984), page 84). One reason that the construction of Dubey and Müller is neat is that, although H is not unique (it depends on the arbitrary w_0), their definitions of metric covariance and Fréchet scores are intrinsic to (Ω, d) and do not depend on the embedding space.

Xiongtao Dai (*Iowa State University, Ames*)

Dubey and Müller should be congratulated for their timely contribution to the methodology for analysing

time-varying object data. The proposed object-valued functional principal component analysis models are commendable in their general applicability to metric spaces which may not even have a co-ordinate frame. The main ingredient is a new metric covariance that measures the association between two random objects, which furthers the concept of total variation (Mardia *et al.*, 1979) and dynamic correlation (Dubin and Müller, 2005) defined for Euclidean and Hilbert space-valued random elements. The proposed generalized Fréchet integrals and Fréchet scores offer new perspectives for analysing time-varying random objects and are distinct from the analogous versions in a classical functional principal component analysis even when specialized to the case of Euclidean-valued functional data.

To illustrate the difference between the metric covariance and the Fréchet variance, consider the case when Ω is a Riemannian manifold, e.g. the p-dimensional sphere S^p (Dai and Müller, 2018) with positive curvature, regarded as a metric space with the geodesic distance d. Let U be a random object that takes values in $\Omega = S^p$, $p \geqslant 2$, and assume that the Fréchet mean $\mu = \arg\min_{\omega \in \Omega} E\{d^2(\omega, U)\}$ of U exists and is unique. With an independent copy U' of U, the metric covariance $\mathrm{cov}_\Omega(U, U) = E\{d^2(U, U')\}/2 \leqslant E\{d^2(\mu, U)\}$, where the rightmost term is referred to as the Fréchet variance (Dubey and Müller, 2019) and the last inequality is due to the positive curvature of the sphere. The inequality is strict if U does not concentrate on a great circle. The metric covariance may thus be interpreted as measuring the association as seen from pairwise disparities, whereas the Fréchet variance measures the variation from the central μ.

For exploratory analysis, real-valued Fréchet scores are demonstrated to be useful in the application to the mortality, taxi networks and world trade data. The Fréchet scores that are produced by projecting the distance function from the object trajectory to the Fréchet mean trajectory on the eigenfunctions summarize deviation in terms of distance from the mean. As there might be additional variation such as rotation around the mean function that is not reflected in the distance to the centre, the Fréchet scores could be complemented by, for example, the multi-dimensional scaling co-ordinates that are produced with distance measure $\rho(x, y) = \int_0^1 d\{x(t), y(t)\}\mathrm{d}t$ between time-varying objects $x, y : [0, 1] \to \Omega$.

The following contributions were received in writing after the meeting.

Rajendra Bhansali (*Imperial College London and University of Liverpool*)

Dubey and Müller are to be congratulated on the originality of thinking shown in this impressive paper, especially for combining novel mathematical techniques with pertinent real life data analysis. I have two comments. It was interesting to see that, although the authors are dealing with highly non-Gaussian data, they have introduced 'linear' notions such as the autocovariance function, albeit a 'metric' function. It would be useful to have analogous, but 'robust' notions such as the median and upper and lower quartiles, as applicable to the type of data that are considered in the paper. An example of a situation where such notions are needed is given in Bhansali (1997), where the robustness of the auto-regressive spectral estimate for linear processes with infinite variance is demonstrated. In the simulations reported there, the median, upper and lower quartiles of the curves of the estimated spectra are shown. However, as such notions are not well defined, the approach taken was simply to evaluate each relevant summary statistic pointwise, individually at all frequencies, for which the spectra were estimated and then to draw a curve connecting these points. An obvious difficulty with such an approach is that the 'median' curve, for example, may not actually have been observed! Secondly, for the class of Gaussian stationary processes, the Hilbert space of random variables is known to be isomorphic to the space of complex exponentials. Although this isomorphism does not extend to the class of linear processes with infinite variance that was considered in Bhansali (1997), it turns out that the standard Gaussian estimates, such as the auto-correlation function, may still be defined for this class, as these estimates still reside on the space of complex exponentials. As a consequence, it is possible to examine the behaviour of such estimates in the latter space of complex exponentials. Bhansali (1988) showed, for example, that the standard Akaike information criterion provides a consistent estimator of the order of a finite auto-regressive process with stable innovations. It would analogously be useful to find a 'convenient' Hilbert space which is isomorphic to the Hilbert space that was introduced by the authors in Section 2.

Pedro Delicado (*Universitat Politècnica de Catalunya, Barcelona*)

I congratulate Dubey and Müller for this magnificent paper. I am sure that the *metric covariance* they define will have great influence in the future, as well as the *object functional principal component analysis* derived from it.

Taking into account that multi-dimensional scaling (MDS) is a powerful tool to deal with random objects in metric spaces (see, for example, Delicado (2011) and Boj *et al.* (2016)), reading Dubey and Müller's

paper suggested to me an MDS-based alternative approach to analyse time-varying random objects (not so elegant as theirs!).

Let $\{X_i = (X_i(t))_{t\in[0,1]} : i = 1, \ldots, n\}$, be a sample. Assume that $X_1, \ldots, X_n$ have been observed at the same *dense* subset $\{t_1, \ldots, t_m\} \subset [0, 1]$. Let $X_i(t_j)$, $j = 1, \ldots, m$, $i = 1, \ldots, n$, be the nm objects that we have actually observed. Proceed as follows.

Step 1: compute MDS from the nm observed objects to obtain K principal directions. Let $\xi_i^k(t_j)$ be the score of the object $X_i(t_j)$ at the kth principal direction. (In Chen *et al.* (2017) an analogous approach was taken.)

Step 2: for $k = 1, \ldots, K$, apply functional principal component analysis (FPCA) to $\{\xi_i^k = (\xi_i^k(t))_{t\in[0,1]} : i = 1, \ldots, n\}$, using the available observations $\xi_i^k(t_j)$, $J = 1, \ldots, m$, $i = 1, \ldots, n$, to obtain eigenfunctions $\psi_h^k(t)$, $h = 1, \ldots, H_k$.

My conjecture is that these eigenfunctions are related to those obtained following Dubey and Müller's methodology, as the following example indicates.

Time-varying normal distributions

Consider the example in Section 5.1 of the paper. The random objects $X_i(t)$ are univariate normal distributions $N\{\mu_i(t), \sigma_i^2(t)\}$ and the 2-Wasserstein metric d_W is used. It is known (Dowson and Landau, 1982; Olkin and Pukelsheim, 1982) that

$$d_W^2\{N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)\} = (\mu_1 - \mu_2)^2 + (\sigma_1 - \sigma_2)^2.$$

Therefore the $nm \times 2$ matrix with generic row $(\mu_i(t_j), \sigma_i(t_j))$ is an MDS solution. So, step 1 leads to $K = 2$, $\xi_i^1(t_j) = \mu_i(t_j)$ and $\xi_i^2(t_j) = \sigma_i(t_j)$. In step 2 the FPCA of $\xi_i^1(t_j) = \mu_i(t_j)$ reveals that, by the definition of $\mu_i(t)$, the principal functions are $\phi_1(t)$ and $\phi_3(t)$ (defined in the paper). Analogously, the principal functions obtained by FPCA of $\xi_i^2(t_j) = \sigma_i(t_j)$ are $\phi_2(t)$ and $\phi_3(t)$. So in this example we obtain the eigenfunctions $\phi_1(t)$, $\phi_2(t)$ and $\phi_3(t)$ given by Dubey and Müller's methodology.

Amira Elayouty, Marian Scott and Claire Miller (*University of Glasgow*)

The paper proposes a novel metric autocovariance function that can measure the association for paired random object data lying in a metric space (Ω, d), where Ω does not necessarily have a vector space or manifold structure. The eigenfunctions of the linear operator associated with the autocovariance function proposed are then used to obtain object functional principal components for Ω-valued functional data, including time-varying probability distributions, covariance matrices and time dynamic networks. In the context of high frequency environmental time series (Elayouty *et al.*, 2016) functional principal components have been used often to reduce the dimensionality of the data. These can be extended to dynamic functional principal components introduced by Hörmann *et al.* (2015) and updated by Elayouty *et al.* (2018) for non-stationary functional time series. These dynamic functional principal components are obtained in the frequency domain, to account for the serial dependence between the functional data, through the eigendecomposition of the spectral density function that contains full information on the autocovariance functions at the different time lags. Elayouty *et al.* (2018) adapted them to allow smooth changes with the changes in the autocovariance structure between the functional data.

The object functional principal components proposed by Dubey and Müller though they bring useful insights about the structure of the underlying functional random objects may miss the time dependence between those objects. This is because the autocovariance function proposed measures the association between Ω-valued functional data only at lag 0, ignoring the cross-correlations at the different time lags. Therefore, an extension of the proposed metric autocovariance function to accommodate the full information on the autocovariances at the different time lags may prove useful.

Young Kyung Lee (*Kangwon National University, Chuncheon*) **and Byeong U. Park** (*Seoul National University*)

The major hurdle in the analysis of functional data taking values in a general metric space is the lack of vector operation. Dubey and Müller are to be congratulated on making a pioneering step forward to solving the difficulty. They extend functional principal component analysis (FPCA) for real-valued random functions to the case of functional random objects by introducing an autocovariance operator acting on the space of real-valued functions. We wish to comment on the so-called metric autocovariance operator on which the associated eigenfunctions and the functional principal components are based.

The extension of FPCA to Hilbert-space-valued functional data is quite straightforward. Let $\mathbb{H}$ be a separable Hilbert space with an inner product '$\langle\cdot,\cdot\rangle$' and the corresponding norm '$\|\cdot\|$'. Let $L_2^{\mathbb{H}}$ denote

the space of Hilbertian functions $f:[0,1]\to\mathbb{H}$ such that $\int_0^1 \|f(t)\|^2 \mathrm{d}t<\infty$. We endow $L_2^{\mathbb{H}}$ with the inner product $\langle\cdot,\cdot\rangle_2$ defined by $\langle f,g\rangle_2=\int_0^1\langle(f(t),g(t)\rangle \mathrm{d}t$. Assume that $E(X)=0$ for simplicity and $E\|X\|_2^2=E\{\int_0^1\|X(t)\|^2\mathrm{d}t\}<\infty$. Then, we may define a covariance operator $\mathscr{C}_X:L_2^{\mathbb{H}}\to L_2^{\mathbb{H}}$ by

$$\mathscr{C}_X(f)=E\{\langle X,f\rangle_2\odot X(\cdot)\}=E\left\{\int_0^1\langle X(t),f(t)\rangle \mathrm{d}t\odot X(\cdot)\right\},$$

where '$\odot$' is scalar multiplication for $\mathbb{H}$. The operator $\mathscr{C}_X$ is non-negative definite and admits an eigen-decomposition and a Karhunen–Loève expansion. We note that the eigenfunctions of $\mathscr{C}_X$ lie in the space $L_2^{\mathbb{H}}$ where the data objects reside.

The above construction of $\mathscr{C}_X$ relies on an inner product and vector operations. Dubey and Müller present a thought-provoking approach to FPCA without making use of them to deal with general metric-space-valued random functions. An immediate extension that comes to our minds is for multivariate functional random objects $\mathbf{X}=(X_1,\ldots,X_d)$ with each X_j being an Ω_j-valued stochastic process and Ω_j having a metric d_j. A direct application of T_C to the induced metric $d(\mathbf{X},\mathbf{Y})=\Sigma_{j=1}^d d_j(X_j,Y_j)$ does not seem to accommodate the cross-covariances $\mathrm{cov}_{\Omega_j,\Omega_k}\{X_j(s),X_k(t)\}$ for $1\leqslant j\neq k\leqslant d$. Instead, the covariance matrix $\mathbf{C}(s,t)=(\mathrm{cov}_{\Omega_j,\Omega_k}\{X_j(s),X_k(t)\})$ and the metric covariance operator $T_{\mathbf{C}}:L_2^{\mathbb{R}^d}\to L_2^{\mathbb{R}^d}$ defined by $(T_{\mathbf{C}}\mathbf{f})(s)=\int_0^1\mathbf{C}(s,t)\mathbf{f}(t)\mathrm{d}t$, $\mathbf{f}\in L_2^{\mathbb{R}^d}$, would work. Another possible, but much more difficult, extension that arises in comparison with $\mathscr{C}_X$ would be to design a covariance operator whose eigenfunctions reside in the same space as the data objects. We wonder whether the existence of a Hilbert space $\mathbb{H}$ and a distance preserving injective map $\iota:\Omega\to\mathbb{H}$ is of any help in solving this challenging problem.

Jorge Mateu (*University Jaume I, Castellón*)

Now that we have already entered the era of 'big data', we can record such large amounts of information that was basically impossible a few years ago. It is now common to observe complex data in a variety of supports (thus many of them non-Euclidean, as stated by Dubey and Müller) in the form of temporal evolution of images or networks. It is thus my pleasure to congratulate the authors on this interesting, timely and certainly attractive paper full of ideas for the coming future. The statistical treatment and analysis of complex data living on networks, for example, is necessary and essential in the field of data science and data analytics. In particular, when we come into the field of spatial statistics we face problems along the lines of what the authors have introduced in this paper. Among all the open ideas and problems that this paper poses, I am particularly interested in, and focus the next comments on, problems related to time-varying network data.

New technological advances in the field of urban planning science enable us to report trajectories of people moving around the city in combination with trajectories of individual vehicles (taxis, as in the example in the paper) and/or public transport. These trajectories live in a network support. which can be considered a directed graph, and very well fit the methodological approach presented in the paper as they can be considered time-varying random objects. We note that we can also treat trajectories as random functions defining a functional mark over a spatial point pattern defined as the locations of events at particular time instants (see Ghorbani *et al.* (2019)). Following Muller *et al.* (2019) we can use a transform–transport metric that generalizes the Wasserstein metric to obtain barycentres for spatially distributed point patterns, and it would be nice to compute the metric covariance under this new metric.

In this context, and in connection with the authors' approach, it would be nice to develop Fréchet regression techniques where the responses can be barycentres of point patterns, or random object trajectories, and also to perform analysis of variance where the entries come from random trajectories and the levels of a potential factor could be human, public transport and private vehicles trajectories. Group testing and inference in this new context would be a welcome contribution for modern statistical analysis.

Katie E. Severn, Ian L. Dryden and Simon P. Preston (*University of Nottingham*)

This interesting paper provides an approach to study covariance properties of general data objects and includes several useful methodological and practical developments. In object-oriented data analysis (Wang and Marron, 2007; Marron and Alonso, 2014) a first question to ask is what are the data objects? As a focus we consider the New York City taxi network data. Should we consider the networks as points on a manifold, or network-valued functions on a time interval or network-valued functions on a circle representing a day, week or year? The representation of the data objects and the methods for analysis will differ in each case. Second, what covariance model is appropriate for the data objects?

Methodology for the analysis of samples of networks as graph Laplacian manifold-valued data has been developed by Severn *et al.* (2019) using Euclidean power metrics, which are of negative type, and

a Procrustes metric. Estimation of extrinsic mean networks and tangent space covariance structure of networks is carried out, together with inference for regression models and interpretation of network principal components (PCs). The mean network of Severn *et al.* (2019) is a common mean μ (either extrinsic or Fréchet) and the covariance structure is then estimated in the tangent space at μ. Another alternative is to consider a time-varying mean function $\mu(t)$ which one can estimate by using a Nadaraya–Watson estimator (Severn, 2019) or a Riemannian smoothing spline (Kim *et al.*, 2019) and unwrap and unroll the data to a base tangent space.

If v is the pT-vector of p-dimensional tangent co-ordinates observed over T times, then some candidate covariance models for v are

(a) $I_p \otimes \Sigma_T$,
(b) $\Sigma_p \otimes I_T$,
(c) $\Sigma_p \otimes \Sigma_T$ and
(d) Σ_{pT}

where I_M is the $M \times M$ identity matrix, Σ_M is an $M \times M$ symmetric positive semidefinite covariance matrix and '$\otimes$' is the Kronecker product. The eigenvectors of Σ_p are the network PCs, the eigenvectors of Σ_T are temporal PCs and the eigenvectors of Σ_{pT} are network temporal PCs. The covariance structure implied in the methods in Dubey and Müller's paper is analogous to model (a). The uncorrected time model (b) is that of Severn *et al.* (2019) and the factored PCs in model (c) can be estimated by using an alternating algorithm (Dutilleul, 1999; Dryden *et al.*, 2009, 2017). It would be interesting to fit models (a)–(d) to the New York City taxi network data, and with different length time intervals, to examine the resulting cluster analyses. Realistically we might expect different network covariance structures at different times, as in model (d), and it would be good to explore this issue.

Han Lin Shang (*Australian National University, Canberra*)
This is a nicely written paper on an interesting and popular topic. Dubey and Müller are to be congratulated on their achievement. Many powerful ideas are developed, but I shall comment on only the following aspect of this work.

Random objects, such as density functions, often have a constrained integral and thus do not constitute a vector space. Implementation of established functional modelling techniques, such as functional principal component analysis, is therefore problematic for such non-linear data. In some special cases, through a transformation method such as log-quantile transformation or compositional data analysis, random objects can be transformed to linear space, as exemplified for the case of objects that are probability distributions (Petersen and Müller, 2016; Kokoszka *et al.*, 2019). It seems natural not to perform any transformation for modelling random objects. However, what could be the disadvantages of the transformation methods?

Through the Wasserstein covariance, the authors quantify a covariance measure for random objects and study functional principal component decomposition under Fréchet integrals. Other dimension reduction techniques, such as the maximum auto-correlation function, can be extended, and a finite sample comparison may be desirable.

Throughout the paper, the authors consider the 2-Wasserstein distance with the usual Euclidean norm as the distance function. Do the results hold for other norms, such as absolute and supremum norms?

In the human mortality data example, the authors consider age-at-death distributions for 32 countries. The analysis was carried out for ages [0, 80] years. Using a histogram, the authors reconstruct a time series of probability density functions. Instead of estimating probability density functions, the authors may consider life table death counts for all available age groups. In the life table death counts, the age interval is from 0 to 109 years in a single year of age, and the last age group from 110 years and older. The life table death count at each year starts from customarily 100000 at infant age and ends at 0 at the last age group 110 years and older. The data themselves naturally present a probability distribution function without any smoothing.

It will, of course, be interesting to see whether these random thoughts could be implemented in practice and, if yes, how good or bad they will perform.

Again, the authors are to be congratulated on such a thought-provoking paper.

Lingxuan Shao and Fang Yao (*Peking University, Beijing*)
Dubey and Müller are to be congratulated on this fine piece of work that laid a fundamental framework for analysing time-varying random objects from the perspective of functional data. Given rapidly emerging

applications involving non-Euclidean data types, the ideas and methods proposed are timely and important. A main advantage in the framework is to define variance and covariance operators that rely only on the metric, not requiring any linear structure or local Euclidean structure. Moreover, the definition of autocovariance as well as the eigenpairs and object principal components do not rely on the Fréchet mean μ, and so nor do their estimators.

To guarantee that the autocovariance is a non-negative definite kernel, the authors assume that the square metric d^2 is of negative type. According to proposition 3 in Sejdinovic *et al.* (2013), the negative-type metric space can be embedded in a Hilbert space, i.e. it can be considered as a subset of a Hilbert space mathematically. This indicates that the space that this paper deals with is not a general metric space, but a subset of a Hilbert space. To be precise, it is better to say that this paper unifies many existing and important research areas, including the L^2-space, 2-Wasserstein space, Riemannian manifolds and others listed in the paper. Although the negative-type metric space can be embedded in a Hilbert space, the information of the injective map is not necessary in the development of estimation and theory, which is a main advantage. In Euclidean space, principal component analysis is mainly a decomposition $X(\omega) = \Sigma_k \beta_k(\omega)\phi_k$. It seems that the generalized eigenfunctions and scores defined here do not satisfy this decomposition or other type of extensions. Some discussion on this would be appreciated. Lastly, assumptions 6 and 8 in the paper impose requirements on $\hat{\phi}$ and $\hat{\mu}$. We wonder whether one can make assumptions on the original distributions of samples but not the samples or the estimators.

The **authors** replied later, in writing, as follows.

We thank the discussants for their contributions and insightful thoughts and are pleased to see that there is general agreement on the importance of developing methodology and theory for time-varying random objects. For brevity, we apologize that we cannot address all the issues raised by the discussants. In what follows, we briefly reply to some of the key points.

Metric covariance: trace of cross-covariance operator and assumptions

As pointed out by Tavakoli and also Sejdinovic and Bergsma, the proposed metric covariance between two random vectors U and V in $\mathbb{R}^p$ summarizes the information contained in the cross-covariance matrix by only reflecting its trace. Unless we have an orthonormal basis in the setting of Hilbert-valued random variables or develop a case-specific custom analysis utilizing the richer structure of a space when available, it seems, however, not possible in general metric spaces to obtain a finer resolution of the dependence structure. Metric covariance quantifies the overall magnitude of association along with a direction of the association, which in the case of random vectors U and V gives an idea about their overall alignment.

Considering the toy examples that were provided by Tavakoli and Sejdinovic, where we have random vectors $X = (Z, 0)$ and $Y = (0, Z)$ with $\mathrm{cov}_\Omega(X, Y) = 0$, while X and Y are dependent in the classical sense, we emphasize that metric covariance being zero does not indicate independence in analogy with classical covariance. In the special situation of the example, X and Y are contained in orthogonal subspaces: an aspect that is crucial for any principal component analysis (PCA) type of orthogonalization approach. This orthogonal support feature implies that metric covariance is zero in these examples. This may be a desirable feature in some contexts. Although one cannot recover the true cross-covariance structure without additional structural assumptions on Ω, metric covariance reveals the orthogonality of the vectors in the toy example.

Tavakoli also mentions that a requirement for the spectral decomposition of C, which is crucial to obtain the eigenfunctions, is the continuity of the functions $X(\cdot)$. In this regard, we have assumed in the paper that the stochastic processes $X(\cdot)$ are α-Hölder continuous in assumption 4, which then implies the required continuity. Shao and Yao point out that assumptions 6 and 8 impose restrictions on the behaviour of the estimates $\hat{\mu}$. It is not difficult to replace these assumptions by stronger requirements on the underlying space, which then will imply them. For spaces with non-positive sectional curvature, these assumptions are always true. They are also true for spaces of non-negative curvature under certain conditions (Ahidar-Coutrix *et al.*, 2019).

Can one check the assumption that the space is of negative type? This is a requirement for our approach to work and is mentioned by both Shang and Zemel. In case the objects are a subset of a Hilbert space, they are automatically of negative type as then the injective transformation can play the role of the map f. A semimetric ρ is of negative type if $\exp(-\lambda\rho)$ is a positive definite kernel (Schoenberg, 1938). For any space and metric one can check whether this holds case by case; it is satisfied for the spaces that we consider but it seems to be an open question whether there is a more general characterization.

Metric covariance: reproducing kernel Hilbert space representation

We are grateful to Sejdinovic for pointing out the interesting correspondence of metric covariance and the trace of the kernel cross-covariance operator, whereas distance covariance, which is equivalent to the Hilbert–Schmidt independence criterion, corresponds to the Hilbert–Schmidt norm of the same kernel cross-covariance operator. For a broad class of kernels, the Hilbert–Schmidt independence criterion or distance covariance when evaluated by using the metric derived from the kernel characterizes probabilistic independence between random objects. This is different from the target of metric covariance, which is not probabilistic independence or dependence, but rather aims to quantify the extent of alignment between random objects. That being said, the connections between metric covariance and its reproducing kernel Hilbert space (RKHS) representation pointed out by Sejdinovic are enlightening and open up avenues for future research by taking advantage of the kernel representations.

Bergsma suggests an alternative in the form of kernel PCA which embeds objects into a latent feature space. In this approach one does not need to know the latent feature space as the kernel principal components (PCs) can be recovered from the eigenvectors of the kernel matrix and may serve as an effective summary in applications, as illustrated for the mortality data example. When using this method, the time dynamics that are inherent in the eigenfunctions of the metric autocovariance between functional random objects cannot, however, be recovered without further knowledge of the embedding. When obtaining the scores for the mortality data example, the distances between the time trajectories are integrated over time. This has the undesirable consequence that we lose details of the time dynamics of the underlying behaviour of the distance trajectories. A common issue for data analysis with kernel approaches is that the kernels rely on additional tuning parameters, for example, through scaling of the distances in Gaussian kernels, or through starting points like w_0 as mentioned by Zemel. In our experience, these tuning parameters can have a major effect on the conclusions, as kernel methods are often highly sensitive to their choice.

Fréchet integrals and scores: variance decomposition

If one works with random objects in a metric space, one does not have available the classical and convenient decomposition of functional variation into orthonormal components that makes functional PCA (FPCA) in Hilbert spaces so immensely useful as a tool for dimension reduction and data analysis. When Ω is such that d^2 is of negative type and $\int_0^1 \phi(t)\mathrm{d}t \neq 0$, with $\tilde{\phi} = \phi / \int \phi$ as defined in the paper, there is an additional interpretation of the Fréchet integrals through the map f that maps Ω to the latent Hilbert space $\mathcal{H}$. Namely, the projection of $X(t)$ onto the function $\tilde{\phi}$ as quantified by the Fréchet integral is

$$\int_{\oplus} X\tilde{\phi} = \arg\min_{\omega\in\Omega} \left\| f(\omega) - \int_0^1 f\{X(t)\}\,\tilde{\phi}(t)\,\mathrm{d}t \right\|_{\mathcal{H}}.$$

Accordingly, $\int_{\oplus} X\tilde{\phi}$ can be interpreted as the element ω in Ω that minimizes the distance between $f(\omega)$ and the usual linear projection $\int_0^1 f\{X(t)\}\tilde{\phi}(t)\,\mathrm{d}t$ of $f\{X(t)\}$ onto $\tilde{\phi}(t)$ and this implies a variance decomposition in $\mathcal{H}$. However, this variance decomposition is primarily of theoretical interest and provides only a partial answer to the issues that were raised by Shao and Yao, since in general neither $\mathcal{H}$ nor the map f is known.

This is also related to the suggestion by Tavakoli to consider more fully the cross-covariance operator in $\mathcal{H}$, which can be done according to the above if we know the map f. But, even then, for spaces such as networks this could lead to computationally difficult optimization problems. This is why the problem of obtaining an eigenbasis in the space of longitudinal random objects or at least in the latent Hilbert space with a known isomorphic map is challenging.

Following up on the discussion of Lee, Park and Shang, there are a few special cases where an isomorphism with a Hilbert space can indeed be explicitly constructed. An instance for this are random distributions, where under mild conditions classes of isomorphic maps to the Hilbert space L^2 can be constructed, such as the log-quantile-density transformation, which is advocated by Shang. Once this or a similar transformation has been applied, variance decomposition can be carried out in this Hilbert space, and mapping back can provide meaningful modes of variation in the original distribution space (Petersen and Müller, 2016).

Fréchet integrals are simple extensions of weighted Fréchet means, do not require knowledge of the map f and can usually be computed without difficulties for any space Ω for which we have an efficient algorithm to compute Fréchet means. Fréchet integrals provide substitutes for projections of trajectories of random objects not only on eigenfunctions but more generally on any function ψ with $\int \psi \neq 0$. For example, suppose that we are interested to project the observed random object trajectories on linear functions, which in the real case could be done by fitting simple least squares lines to the data of each subject. For

longitudinally observed random objects one can obtain the equivalent to a fitted line $L(t)$ at predictor time t as Fréchet integral

$$L(t) = \arg\min_{\omega \in \Omega} \int d^2\{X(u), \omega\}\, \psi_t(u)\, \mathrm{d}u.$$

These Fréchet integrals project trajectories $X(u)$ onto the functions $\psi_t(u) = s(u, t)$, where the functions $s(\cdot, \cdot)$ are as defined in equation (2.6) on page 695 in Petersen and Müller (2019) and correspond to those used for linear Fréchet regression. Here we have $\int \psi_t(u)\mathrm{d}u = 1$ for all t. Similarly, one can project on the equivalent of polynomial functions and other function spaces.

Although the scalar-valued eigenfunctions of the metric autocovariance operator, similar to the metric autocovariance itself, do not reflect fine details of the cross-covariance operator, this information compression is to some extent mitigated by the application of Fréchet integrals that take values in the same space as the time-varying random objects and serve as substitutes for the scalar PCs that one obtains by using approaches like the kernel PCA discussed by Bergsma. For instance, if we have time-varying networks like the New York taxi data, the Fréchet integrals are networks themselves that can uncover the nodes in the network that contribute to the dominant modes of variation in the direction of the dominant eigenfunctions of metric autocovariance; some illustrations are provided in the movies that accompany the paper. In the same way as functional principal components (FPCs), the Fréchet integrals can be used in subsequent statistical analysis in various ways, as mentioned by Mateu. For example, if we are interested in the effect of baseline covariates on the trajectories of random objects, we can first construct the Fréchet integrals for all subjects, and then use these as responses in a Fréchet regression or in a Fréchet analysis of variance.

Time warping and time series case

Zemel pointed out the possible effect of time warping, which is also often referred to as curve registration or alignment in the context of functional data (Marron *et al.*, 2015; Wagner and Kneip, 2019). This leads to an interesting research problem. Some members of our University of California, Davis, research team have been working on such an approach in on-going research. One can actually use ideas of pairwise warping from functional data analysis (Tang and Müller, 2008), which, it turns out, can be adapted to distance-based pairwise comparisons, extending the corresponding functional approach to the time warping of random object trajectories.

Prompted by the comments of Elayouty, Scott and Miller, we reiterate that our approaches rely on obtaining consistent estimates of the metric autocovariance operator from observed data. For this one needs to have available an independent and identically distributed sample of functional random objects, which is not so in the time series case where one typically works with only one realization of an underlying process. Similarly to the time series case, the metric autocovariance operator indeed includes lags as it quantifies $\mathrm{cov}_\Omega\{X(t), X(s)\}$ for any s and t and specifically does not require an analogue of stationarity.

Multivariate functional random objects and visualization

The case of multivariate functional random objects has practical relevance; for example one might consider the joint development of left brain and right brain connectivity networks in neurodevelopment studies. At some level it is quite straightforward to extend the proposed methodology to this case, as suggested by Lee and Park, who propose the metric $d(X, Y) = \Sigma_{j=1}^{d} d_j(X_j, Y_j)$, which provides a flexible choice. The multivariate scenario presents interesting open areas for future research, including the construction of joint models for vectors of different types of object data that might reside in different spaces and might also have different temporal domains.

One of the challenges when working with object-valued PCs is visualization, where in classical FPCA plotting the FPCs against each other often provides useful information about the structure of the data. Instead of the object-valued Fréchet integrals, one could resort to the scalar-valued Fréchet scores and employ them in pairwise plots. As pointed out by Tavakoli, the Fréchet scores are, however, not equivalent to the FPC scores as obtained in classical FPCA. They reflect the decomposition of the individual variance trajectories $d\{X(t), \mu_\oplus(t)\}$ into the dominant directions obtained from the metric autocovariance operator and their interpretation is related to temporal patterns of distance of individual trajectories from the mean trajectory. Further quantifications of such distances and the corresponding scores could be a topic for future research.

An alternative for visualization is to work with multi-dimensional scaling co-ordinates as proposed by both Dai and Delicado. It would be interesting to compare these representations with those provided by the Fréchet scores. The two-step algorithm proposed by Delicado uses multi-dimensional scaling co-

ordinates as the starting point for collapsing the information that is contained in time-varying objects into time-varying functional data. This is an attractive proposal that deserves further study.

Flexibility through metric selection

The metric covariance proposed can lead to vastly different analyses of longitudinal object data if different metrics are chosen. Metric selection is thus a crucial first step before starting the analysis. It will often be based on practical considerations such as the feasibility of computations and the interpretability of results. For example, in the case of distributional data, the Wasserstein metric has emerged as a favourite when working with one-dimensional distributions (Bolstad *et al.*, 2003; Zhang and Müller, 2011; Panaretos and Zemel, 2019; Bigot, 2019); however it is not clear that it is the universally best choice for samples of trajectories of random distributions.

For multivariate distributions, the situation is much more complex and there are major computational and interpretational difficulties when using the Wasserstein metric. Optimal transport that underpins the Wasserstein metric is a mathematically nice optimization target but might not always be a good criterion from a data analysis perspective. If the random objects are covariance matrices or graph Laplacians, a multitude of metrics are available and the choice may depend on a trade-off between simplicity, where the simplest choice would be the Frobenius metric, and avoiding undesirable features such as the swelling effect (Lin, 2019).

For the case of time-varying networks, Mateu suggests treating the trajectories of the New York taxi data as random functions defining a functional mark over a spatial point pattern equipped with the transport–transform metric. This provides a promising perspective for future research. Mateu also considers a very interesting class of random objects which are point patterns. Shang observes that, in the mortality data example, instead of dealing with smooth probability distributions of age at death, one might instead consider the life table counts directly with an appropriate metric. Another alternative would be to transform the densities to log-hazard functions which are in L^2 (Petersen and Müller, 2016) and where one can connect the analysis to previous approaches that focused on log-hazard functions (Chiou and Müller, 2009).

Riemannian manifold approach

Severn, Dryden and Preston mention that, with regard to time-varying networks, alternative approaches can be devised by using graph Laplacians and exploiting tangent plane representations and extrinsic metrics for some of the space–metric combinations. Such manifold representations can also be applied in other special cases of random objects, where local linearizations can be used to advantage. In the tangent spaces, linear methods can be applied, though often with limitations; in the case of manifold-valued trajectories one can still take advantage of Karhunen–Loève expansions (Dai and Müller, 2018). However, for space–metric combinations of interest it may be tedious to compute the logarithmic and exponential maps; such maps may not exist or not be invertible, or underlying manifolds may not be Riemannian or may be infinite dimensional, which can lead to additional complications. If the ambient data space is high dimensional, manifold approaches can also be affected by the curse of dimensionality, which then may require invoking sparsity assumptions (Ginestet *et al.*, 2017), adding to the complexity of these approaches.

On the plus side, the manifold approach can provide illuminating complementary insights to those which we may obtain with the proposed metric approach and may also yield interpretable visualizations. In the end, it will be beneficial to have a rich toolbox with multiple available approaches for analysing functional random objects, which would include the manifold approach, where applicable, and the proposed metric approach as a generic method with a wide range of applicability. We agree with Severn, Dryden and Preston that it would be of substantial interest to compare the different available approaches for time-varying networks by using one standard data set, say the New York taxi data. Expanding on a suggestion by Shang, a similar comparison of transformation, compositional and metric approaches would be of interest for time-varying distributions.

Open problems

Since the modelling of trajectories of random objects has been motivated by recent new types of complex time-varying data, there are still a large number of open problems in this area, some of which have been already mentioned. In addition to the exploration of the connections of metric covariance with RKHS approaches and kernel PCA as suggested in the discussions of Sejdinovic and Bergsma, interesting problems that are currently open and have been brought to the fore by the stimulating contributions by the discussants include the following problems.

Robustness and quantiles. An interesting direction for future research is robustness, as proposed by Bhansali, and includes even simple notions that can serve as building blocks for robust procedures, such as

medians. Quantiles for random objects remain a largely open problem. Case-by-case analysis that utilizes the finer structural features of Ω, where available, may lead to sensible solutions. Nevertheless, quantiles and robust options are difficult when the underlying metric space has moderate to high dimension, as even for the simpler multivariate case the notion of multivariate quantiles remains far from settled (Chernozhukov *et al.*, 2017). An alternative perspective could be to use MDS co-ordinate summaries such as those proposed by Dai and Delicado to obtain representations in $\mathbb{R}^p$.

Random object time series. As already mentioned, our methods do not cover time series, where one has only one trajectory and time is discrete, whereas we consider samples of trajectories that are recorded in continuous time. This motivates us to consider random object time series methodology, where one investigates object-valued time series. These are of interest for many applications, whereas time series analysis to date has by and large been limited to multivariate high dimensional and functional data.

Incompletely observed trajectories. We assumed that the time courses of the random objects are fully observed; however, in principle, they are only observed on a grid, as Zemel mentioned, and therefore appropriate extensions are needed to cover this case. (Pre)smoothing of objects can be carried out with local Fréchet regression (Petersen and Müller, 2019) in analogy with presmoothing for traditional scalar-valued functional data that are observed on a discrete grid (Müller *et al.*, 2006).

In this vein, we may also need to deal with the issue that the observations made on grid points are contaminated with errors. Modelling errors for random objects is not straightforward since we cannot add zero-mean finite variance errors to random objects in the way that we would usually model error contamination for functional data. Error contamination may require a case-by-case analysis; typically one will assume that conditional Fréchet means are on target. For the error-contaminated observed random objects $\tilde{X}(t)$ at a grid point t a reasonable assumption is

$$\arg\min_{\omega\in\Omega} E[d^2\{\tilde{X}(t),\omega\}|X(t)] = X(t), \tag{27}$$

where such contaminations have been implemented for the case of distributions (Petersen and Müller, 2016). In some situations, the time grid on which observations are available may differ across subjects and in some cases it is sparse (Li and Hsing, 2010; Zhang and Wang, 2016, 2018).

Incompletely observed random objects. Another issue that requires further investigation is the fact that often objects are not fully observed but must be estimated from a random sample that is generated by the object, where one assumes that these within-object samples are independent from the functional random objects. This estimation step induces another error that will typically not satisfy assumption (27) but these errors will vanish as the within-object samples become larger and so under reasonable assumptions can be ignored in the asymptotic analysis; this kind of estimation error has been analysed in detail for the case of distributions (Panaretos and Zemel, 2016; Petersen and Müller, 2016), where the distributions are usually not known but must be estimated from samples.

Random objects indexed by sets other than (time) intervals. Spatially indexed objects are of interest for various applications as pointed out by Tavakoli. We wholeheartedly agree; this includes not only geographical domains (Tavakoli *et al.*, 2019), but also spheres (the Earth) and brain voxels or brain regions indexed by subsets of $\mathbb{R}^3$.

References in the discussion

Ahidar-Coutrix, A., Gouic, T. L. and Paris, Q. (2019) Convergence rates for empirical barycenters in metric spaces: curvature, convexity and extendible geodesics. *Probab. Theory Reltd Flds*, to be published.

Berg, C., Christensen, J. P. R. and Ressel, P. (1984) *Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions*. New York: Springer.

Bhansali, R. J. (1988) Consistent order determination for processes with infinite variance. *J. R. Statist. Soc.* B, **50**, 46–60.

Bhansali, R. J. (1997) Robustness of the autoregressive spectral estimate for linear processes with infinite variance. *J. Time Ser. Anal.*, **18**, 213–229.

Bigot, J. (2019) Statistical data analysis in the Wasserstein space. *Preprint arXiv:1907.08417*.

Boj, E., Caballé, A., Delicado, P., Esteve, A. and Fortiana, J. (2016) Global and local distance-based generalized linear models. *TEST*, **25**, 170–195.

Bolstad, B. M., Irizarry, R., Åstrand, M. and Speed, T. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.

Chen, K., Delicado, P. and Müller, H.-G. (2017) Modelling function-valued stochastic processes, with applications to fertility dynamics. *J. R. Statist. Soc.* B, **79**, 177–196.
Chernozhukov, V., Galichon, A., Hallin, M. and Henry, M. (2017) Monge–Kantorovich depths, quantiles, ranks and signs. *Ann. Statist.*, **45**, 223–256.
Chiou, J.-M. and Müller, H.-G. (2009) Modeling hazard rates as functional data for the analysis of cohort lifetables and mortality forecasting. *J. Am. Statist. Ass.*, **104**, 572–585.
Dai, X. and Müller, H.-G. (2018) Principal component analysis for functional data on Riemannian manifolds and spheres. *Ann. Statist.*, **46**, 3334–3361.
Delicado, P. (2011) Dimensionality reduction when data are density functions. *Computnl Statist. Data Anal.*, **55**, 401–420.
Dowson, D. and Landau, B. (1982) The Fréchet distance between multivariate normal distributions. *J. Multiv. Anal.*, **12**, 450–455.
Dryden, I., Hill, B., Wang, H. and Laughton, C. (2017) Covariance analysis for temporal data, with applications to DNA modelling. *Stat*, **6**, 218–230.
Dryden, I. L., Bai, L., Brignell, C. J. and Shen, L. (2009) Factored principal components analysis, with applications to face recognition. *Statist. Comput.*, **19**, 229–238.
Dubey, P. and Müller, H.-G. (2019) Fréchet analysis of variance for random objects. *Biometrika*, **106**, 805–821.
Dubin, A. and Müller, H. G. (2005) Dynamical correlation for multivariate longitudinal data. *J. Am. Statist. Ass.*, **100**, 872–881.
Dutilleul, P. (1999) The MLE algorithm for the matrix normal distribution. *J. Statist. Computn Simuln*, **64**, 105–123.
Elayouty, A., Scott, M., Miller, C. and Waldron, S. (2018) Functional principal component analysis for non-stationary dynamic time series. In *Proc. 33rd Int. Wrkshp Statistical Modelling, Bristol*, pp. 84–89.
Elayouty, A., Scott, M., Miller, C., Waldron, S. and Franco-Villoria, M. (2016) Challenges in modeling detailed and complex environmental data sets: a case study modeling the excess partial pressure of fluvial CO_2. *Environ. Ecol. Statist.*, **23**, 65–87.
Ghorbani, M., Cronie, O., Mateu, J. and Yu, K. (2019) Functional marked point processes: a natural structure to unify spatio-temporal frameworks and to analyse dependent functional data. *20th Wrkshp Stochastic Geometry, Stereology and Image Analysis, Sandbjerg Estate, June 2nd–7th*.
Ginestet, C. E., Li, J., Balachandran, P., Rosenberg, S. and Kolaczyk, E. D. (2017) Hypothesis testing for network data in functional neuroimaging. *Ann. Appl. Statist.*, **11**, 725–750.
Hörmann, S., Kidziński, Ł. and Hallin, M. (2015) Dynamic functional principal components. *J. R. Statist. Soc.* B, **77**, 319–348.
Hsing, T. and Eubank, R. (2015) *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. New York: Wiley.
Human Mortality Database (2019) Human mortality database. University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. (Available from `www.mortality.org`.)
Kim, K.-R., Dryden, I. L. and Le, H. (2019) Smoothing splines on Riemannian manifolds, with applications to 3D shape space. *Preprint arXiv:1801.04978*.
Kokoszka, P., Miao, H., Petersen, A. and Shang, H. L. (2019) Forecasting of density functions with an application to cross-sectional and intraday returns. *Int. J. Forecast.*, **35**, 1304–1317.
Li, Y. and Hsing, T. (2010) Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.*, **38**, 3321–3351.
Lin, Z. (2019) Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition. *SIAM J. Matrx Anal. Appl.*, **40**, 1353–1370.
Lu, X., Marron, J. S. and Haaland, P. (2014) Object-oriented data analysis of cell images. *J. Am. Statist. Ass.*, **109**, 548–559.
Lyons, R. (2013) Distance covariance in metric spaces. *Ann. Probab.*, **41**, 3284–3305.
Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979) *Multivariate Analysis*. Cambridge: Academic Press.
Marron, J. S. and Alonso, A. M. (2014) Overview of object oriented data analysis. *Biometr. J.*, **56**, 732–753.
Marron, J. S., Ramsay, J. O., Sangalli, L. M. and Srivastava, A. (2015) Functional data analysis of amplitude and phase variation. *Statist. Sci.*, **30**, 468–484.
Masarotto, V., Panaretos, V. M. and Zemel, Y. (2019) Procrustes metrices on covariance operators and optimal transportation of Gaussian processes. *Sankhya* A, **81**, 172–213.
Müller, R., Schuhmacher, D. and Mateu, J. (2019) Metrics and barycenters for point pattern data. Submitted to *Statist. Comput*.
Müller, H.-G., Stadtmüller, U. and Yao, F. (2006) Functional variance processes. *J. Am. Statist. Ass.*, **101**, 1007–1018.
Olkin, I. and Pukelsheim, F. (1982) The distance between two random vectors with given dispersion matrices. *Lin. Alg. Appl.*, **48**, 257–263.
Panaretos, V. M. and Zemel, Y. (2016) Amplitude and phase variation of point processes. *Ann. Statist.*, **44**, 771–812.

Panaretos, V. M. and Zemel, Y. (2019) Statistical aspects of Wasserstein distances. *A. Rev. Statist. Appl.*, **6**, 405–431.
Petersen, A. and Müller, H.-G. (2016) Functional data analysis for density functions by transformation to a Hilbert space. *Ann. Statist.*, **44**, 183–218.
Petersen, A. and Müller, H.-G. (2019) Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.*, **47**, 691–719.
Schoenberg, I. J. (1983) Metric spaces and positive definite functions. *Trans. Am. Math. Soc.*, **44**, 522–536.
Schölkopf, B., Smola, A. and Müller, K. R. (1998) Nonlinear component analysis as a kernel eigenvalue problem. *Neurl Computn*, **10**, 1299–1319.
Sejdinovic, D., Sriperumbudur, B., Gretton, A. and Fukumizu, K. (2013) Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.*, **41**, 2263–2291.
Severn, K. E. (2019) Manifold-valued data analysis of networks and shapes. *PhD Thesis*. University of Nottingham, Nottingham.
Severn, K. E., Dryden, I. L. and Preston, S. P. (2019) Manifold valued data analysis of samples of networks, with applications in corpus linguistics. *Preprint arXiv:1902.08290*.
Sriperumbudur, B. K., Fukumizu, K. and Lanckriet, G. R. G. (2011) University, characteristic kernels and RKHS embedding of measures. *J. Mach. Learn. Res.*, **12**, 2389–2410.
Székely, G. J. and Rizzo, M. L. (2009) Brownian distance covariance. *Ann. Appl. Statist.*, **3**, 1236–1265.
Székely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007) Measuring and testing dependence by correlation of distances. *Ann. Statist.*, **35**, 2769–2794.
Tang, R. and Müller, H.-G. (2008) Pairwise curve synchronization for functional data. *Biometrika*, **95**, 875–889.
Tavakoli, S., Pigoli, D., Aston, J. A. and Coleman, J. (2019) A spatial modeling approach for linguistic object data: analysing dialect sound variations across Great Britain. *J. Am. Statist. Ass.*, **114**, 1081–1096.
Wagner, H. and Kneip, A. (2019) Nonparametric registration to low-dimensional function spaces. *Computnl Statist. Data Anal.*, **138**, 49–63.
Wang, H. and Marron, J. S. (2007) Object oriented data analysis: sets of trees. *Ann. Statist.*, **35**, 1849–1873.
Zemel, Y. and Panaretos, V. M. (2019) Fréchet means and Procrustes analysis in Wasserstein space. *Bernoulli*, **25**, 932–976.
Zhang, X. and Wang, J.-L. (2016) From sparse to dense functional data and beyond. *Ann. Statist.*, **44**, 2281–2321.
Zhang, X. and Wang, J.-L. (2018) Optimal weighting schemes for longitudinal and functional data. *Statist. Probab. Lett.*, **138**, 165–170.
Zhang, Z. and Müller, H.-G. (2011) Functional density synchronization. *Computnl Statist. Data Anal.*, **55**, 2234–2249.