Econometrics and Statistics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

Modeling Probability Density Functions as Data Objects

Alexander Petersen[a,b], Chao Zhang[b], Piotr Kokoszka[c,*]

[a] Department of Statistics, Brigham Young University, Provo, UT 84602-0001, USA
[b] Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106-3110, USA
[c] Department of Statistics, Colorado State University, Fort Collins, CO 80523-1877, USA

ARTICLE INFO

Article history:
Received 15 October 2020
Revised 4 March 2021
Accepted 9 April 2021
Available online xxx

Keywords:
Object-oriented statistics
Probability density functions

ABSTRACT

Recent developments in the probabilistic and statistical analysis of probability density functions are reviewed. Density functions are treated as data objects for which suitable notions of the center of distribution and variability are discussed. Special attention is given to nonlinear methods that respect the constraints density functions must obey. Regression, time series and spatial models are discussed. The exposition is illustrated with data examples. A supplementary vignette contains expanded versions of data analyses with accompanying codes.

1. Introduction

Historically, statistical methodologies have utilized probability density functions as a means of modeling the generative process for observed data. The recovery of the generative distribution is the goal of statistical inference, allowing the analyst to assess uncertainty in estimates of model parameters and to produce forecasts and predictions for future observations. Hence, classical analyses in which the data atoms are random variables or vectors have treated the density function as a fixed, unknown quantity used to represent the heterogeneity in observations of distinct subjects or replicated experiments. However, with the increasing velocity and volume (two of the three 'v's of big data) of recorded data, it is increasingly the case that each observation in a data set is associated with its own probability distribution. In this case, one is interested in the heterogeneity amongst the observed (or, far more commonly, latent) sampled densities. Examples in which densities are the data atoms of interest include yearly income distributions, (Kneip and Utikal, 2001), zooplankton size structure in oceanography, (Nerini and Ghattas, 2007), population age and mortality distributions across different countries or regions (Hron et al., 2016; Bigot et al., 2017), distributions of functional connectivity patterns in the brain, (Petersen et al., 2016), and distributions of intra-day or cross-sectional financial returns, (Kokoszka et al., 2019), to name only a few. The shift from viewing pdfs as models of random observations to individual observations of a meta-generative process on the space of distributions is now hitting its stride, with new applications exerting pressure for novel methodologies. It should be remarked that this movement is related to nonparametric Bayesian models that also frequently involve a probability law on the space of distributions, but is distinct in its treatment of densities as the data objects.

Early approaches to the analysis of samples of probability density functions, including work on dimension reduction, (Kneip and Utikal, 2001), and classification, (Nerini and Ghattas, 2007), were influenced by contemporary developments in the field of functional data analysis, an established, though continually expanding, field of statistics; see, e.g., Ramsay and Silverman (2005); Ferraty and Vieu (2006); Horváth and Kokoszka (2012); Hsing and Eubank (2015); Wang et al. (2016);

* Corresponding author.
E-mail addresses: petersen@stat.byu.edu (A. Petersen), czhang@pstat.ucsb.edu (C. Zhang), Piotr.Kokoszka@colostate.edu (P. Kokoszka).

https://doi.org/10.1016/j.ecosta.2021.04.004
2452-3062/

Kokoszka and Reimherr (2017). However, progress was limited by the ubiquity of linear approaches to functional data, whereas the defining features of probability density functions are often violated by linear methods. In recent years, the useful strategy of data transformation has come to the rescue, allowing application of linear functional data methods to density samples; see Sections 2.1 and 2.2. Developments in the seemingly disjoint field of object-oriented data analysis, (Marron and Alonso, 2014; Patrangenaru and Ellingson, 2015), which were mostly designed for data on known manifolds and their underlying metrics, have also been adopted for density samples; see Section 2.3. In the case of densities, there are interesting connections between these approaches that will be explored in this review.

To begin a discussion of density data, one must have a notion of the sample space, meaning the set of values that the random data objects can assume. Denote by $\mathcal{D}$ a subset of probability density functions on the real line $\mathbb{R}$, i.e.,

$$\mathcal{D} \subset \{f : f \geq 0 \text{ and } \int_{\mathbb{R}} f(x)\mathrm{d}x = 1\}, \tag{1.1}$$

and let d be a generic metric on $\mathcal{D}$. Specific approaches discussed in the following rarely work for all densities and this is why a subset of the set of all densities is considered. Restrictions are usually placed on the allowable elements of $\mathcal{D}$ in order for d to be well defined. In this review, the discussion of the modeling of random densities or, more generally, random distributions, will be restricted to methods that model the densities nonparametrically, meaning that $\mathcal{D}$ cannot be reduced to a finite-dimensional space. The reason for this is that if $\mathcal{D}$ is a known parametric class, then classical multivariate analysis techniques can be used to analyze a sample of densities, treating the collection of parameters of each density as a random vector, even though the performance of such approaches must be investigated in specific problems. Furthermore, only passing treatment will be given to the naive approach of directly applying standard functional data technology to the densities themselves, for example if $\mathcal{D} \subset L^2[0,1]$, since such analyses are based on a Hilbert space structure that is extrinsic to the space $\mathcal{D}$. The review will instead focus on intrinsic methods, meaning that they incorporate the nonlinear geometric constraints inherent to $\mathcal{D}$, i.e. $f \geq 0$ and $\int f = 1$ for any $f \in \mathcal{D}$.

The review is structured as follows. Section 2 discusses different methods of representing densities as data, with different representations being amenable to different modes of analysis. Section 3 describes how these representations have been used to yield generalizations of common statistical methods for exploratory analysis, dimension reduction, and even regression. Included with the descriptions of these methods are illustrative analyses of real data sets; a supplementary vignette containing expanded versions of these analyses with accompanying codes is available at https://github.com/czhang-pstat/Density-Review. Section 4 outlines existing methodologies for samples of dependent densities. Finally, we conclude the paper with a discussion of future directions. Before proceeding with this outline, we wish to acknowledge the reality that this review is not completely comprehensive, but rather represents problems in which the authors have strong interest, some modest experience, and at least a modicum of expertise.

2. Representation Spaces for Densities

A unique feature of density-valued data compared to standard functional data is that a probability density function f is only one of the various functions that characterize the underlying distribution, in addition to the cumulative distribution function $F(\cdot) = \int_{-\infty}^{\cdot} f(x)\mathrm{d}x$ and quantile function $Q = F^{-1}$, among others. When performing exploratory analysis on a sample of distributions, or interpreting distribution-valued model outputs, visual inspection of densities is a common choice as these reveal differences in local features that are usually the focus of the problem at hand. However, the nonlinear constraints associated with densities make them difficult to handle from a modeling perspective. Hence, although distributional inputs and outputs are most usefully analyzed as densities, statistical modeling and computation is usually performed in an altogether different space in which the distributions are represented. In the remainder of this section, three examples of representation spaces and their interrelations will be described. The organization of these spaces within this review is designed to reflect the two schools of thought mentioned in the introduction, namely the functional and object-oriented approaches to analyzing density data, which are described in Sections 2.1 and 2.3, respectively. Interjected between these sections is a description of the Bayes space representation for densities, which serves as a bridge between the two approaches as it can actually be considered a special case of both. Moreover, various papers have already been published that develop methods under the Bayes space representation, so that it is arguably deserving of separate treatment.

2.1. Transformation Approach

When directly applied to density data, commonly used methods in functional data analysis, such as functional principal component analysis or functional linear regression, that are inherently designed for data in a Hilbert space can lead to difficulties in interpreting model outputs since $\mathcal{D}$ is nonlinear, (Kneip and Utikal, 2001; Delicado, 2011). In order to improve application of these highly useful methods to density samples, Petersen et al. (2016) proposed to perform a preliminary transformation prior to analysis. In general, suppose $\psi : \mathcal{D} \to \mathbb{H}$ is a transformation mapping densities into a Hilbert space of functions $\mathbb{H}$. The goal is to apply linear methods of functional data analysis to the functional variables post transformation, and to interpret the model outputs in the original density space $\mathcal{D}$. Hence, we refer to $\mathbb{H}$ as a representation space, which may depend on the transformation ψ. A necessary restriction on the map ψ is that it be invertible, loosely speaking. For example, the identity mapping $\psi(f) = f$ used in Kneip and Utikal (2001) or the popular method of kernel mean embedding,

(Muandet et al., 2016), are not valid transformations in this sense since most elements of the corresponding representation space cannot be mapped back to a density.

Petersen et al. (2016) defined two transformations that, under suitable constraints on $\mathcal{D}$, possess the necessary inverses. Of these two, the log quantile density (LQD) transformation has been found more useful in practical scenarios (Chen et al., 2019; Petersen et al., 2019a; Salazar et al., 2020). For a density $f \in \mathcal{D}$, let Q_f denote its quantile function. The LQD transformation is

$$\psi_{\text{LQD}}(f)(t) = -\log\{f \circ Q_f(t)\}, \quad t \in [0,1], \tag{2.1}$$

with the representation space being $L^2[0,1]$. Petersen et al. (2016) restricted the densities in $\mathcal{D}$ to have common support on [0,1], from which one can conclude that $\psi(\mathcal{D})$ is not a linear subspace of $L^2[0,1]$. To overcome this difficulty, a suitable inverse was defined as follows. Suppose $g \in L^2[0,1]$ satisfies $\theta_g := \int_0^1 \exp\{g(s)\}\mathrm{d}s < \infty$. Then set

$$\psi_{\text{LQD}}^{-1}(g)(x) = \theta_g \exp\{-g \circ F_g(x)\}, \quad F_g^{-1}(t) = \theta_g^{-1} \int_0^t \exp\{g(s)\}\mathrm{d}s. \tag{2.2}$$

This definition ensures that, for such g, its inverse will be a density with support [0,1]. In Kokoszka et al. (2019), the authors dealt with densities that did not share a common support, so that the above definition of the LQD transformation was not adequate, and a modified LQD transformation together with its inverse were defined for such situations. Intuitively, since one may perform a location shift to a distribution without altering its LQD, this modified transformation is a pair $(a, \psi_{\text{LQD}}(f))$, where a can be chosen either as a probability $F(x)$ for some fixed $x \in \mathbb{R}$, or else a quantile $F^{-1}(t)$ for fixed $t \in [0,1]$.

This transformation approach is agnostic to the metric d that a user may choose to quantify data variability or discrepancies between targets and their estimates, and is thus a general purpose method that makes the well developed toolkit of functional data analysis applicable to density data in a more coherent and faithful manner. The restrictions placed on $\mathcal{D}$ in (1.1) depend on the chosen transformation, and can be practical (e.g., to ensure that the map and its inverse are computationally feasible), theoretical (e.g., to ensure that the map is continuous in some sense, thus preserving certain statistical properties in the transformed space), or both. For specific examples of such restrictions with regard to the LQD transformation, see Petersen et al. (2016).

2.2. Bayes Spaces

Compositional data analysis consists of methodology designed for a random vector $X \in \mathbb{R}^p$ whose components are positive and satisfy a sum constraint, e.g., $\sum_{j=1}^p X_j = 1$, and thus live on a simplex, (Aitchison, 1986). The vast majority of methods for compositional data are based on the so-called Aitchison geometry, which establishes a linear structure for the simplex by transforming the components of X, exploiting the relative rather than absolute nature of the information that it represents. In Egozcue et al. (2006), it was observed that densities are essentially infinite-dimensional compositional data, and a pre-Hilbert space structure was developed for a restricted class of densities as a direct extension of the Aitchison geometry. This was later generalized to a full Hilbert space in Van den Boogaart et al. (2014) for a slightly larger class of functions. Specifically, let I be a closed interval, and define

$$B = \{f : f > 0, \ \int_I \{\log f(x)\}^2 \mathrm{d}x < \infty\}. \tag{2.3}$$

Note that B contains densities (and other positive functions) with square integrable logarithm. Elements $f, g \in B$ are said to be equivalent if there exists some $c > 0$ such that $f(x) = cg(x)$ almost everywhere. This relation reflects the notion that probability density functions contain only relative information, whence the connection to compositional data. With a slight abuse of notation, we use B to refer to the quotient space under this notion of equivalence, and the definitions that follow are easily seen to be invariant to the representative chosen from each equivalence class. Equip B with the addition and scalar multiplication operations

$$[f \oplus_B g](x) := f(x)g(x), \quad \alpha \odot f(x) := [f(x)]^\alpha, \quad \alpha \in \mathbb{R}, \tag{2.4}$$

and inner product

$$\langle f, g \rangle_B := \frac{1}{2|I|} \int_{I^2} \log\left\{\frac{f(x)}{f(y)}\right\} \log\left\{\frac{g(x)}{g(y)}\right\} \mathrm{d}x\mathrm{d}y. \tag{2.5}$$

Van den Boogaart et al. (2014) showed that this construction yields a Hilbert space on B, with corresponding norm $\|\cdot\|_B$ and metric d_B. Moreover, the restriction to densities on an interval is unnecessary, as Bayes space structures can be extended to domains of arbitrary dimension and shape. Although the above definition implicitly uses the uniform distribution on I as a reference measure, the development in Van den Boogaart et al. (2014) is more general, and Talská et al. (2020) have recently demonstrated the practical utility of alternative data-driven reference measures.

There exists an interesting connection between Bayes spaces and the transformation approach of the previous subsection. For any $f \in B$, the *centered log ratio* (clr) transformation

$$\psi_{\text{clr}}(f) := \log(f) - \frac{1}{|I|} \int_I \log\{f(x)\}\mathrm{d}x \tag{2.6}$$

satisfies $\psi_{\text{clr}}(f) \in L_0^2(I) := \{h \in L^2(I) : \int_I h(x)\mathrm{d}x = 0\}$, which is a Hilbert space with the ordinary functional notions of addition, scalar multiplication, and inner product. This map is a bijection, with

$$\psi_{\text{clr}}^{-1}(h)(x) = \exp\{h(x)\}, \quad h \in L_0^2(I). \tag{2.7}$$

It is then immediate that

$$\langle f, g\rangle_B = \int_I \psi_{\text{clr}(f)}(x)\psi_{\text{clr}}(g)(x)\mathrm{d}x,$$

so that ψ_{clr} is in fact an isomorphic isometry of Hilbert spaces B and $L_0^2(I)$. Thus, if we take $\mathcal{D}$ as the representatives of equivalence classes in B whose elements satisfy $\int_I f(x)\mathrm{d}x < \infty$, this yields another example of the general transformation approach with representation space $L_0^2(I)$. In contrast to the LQD transformation, for which the Hilbertian metric in the representation space $L^2[0,1]$ does not have a clear geometric interpretation in terms of densities, the metric d_B used to develop Bayes space methodologies is derived directly from the compositional nature of probability density functions.

2.3. Object Oriented Approach

The use of geometry to inform statistical models is central to the field of object oriented data analysis, (Marron and Alonso, 2014). Thus, the Bayes space approach constitutes and example of object-oriented analysis of densities, whereby the given metric directly informs the statistical model and its parameters. While the Bayes space induces a linear geometry that facilitates computation, other alternative metrics for analyzing probability distributions have been considered, including the Wasserstein, (Panaretos and Zemel, 2020), and Fisher-Rao, (Srivastava et al., 2007), metrics. The former is an optimal transport metric while the latter can be viewed as the geodesic distance between the square-roots of the densities, as opposed to the chord distance that yields Hellinger's distance, (Hellinger, 1909). The Wasserstein and Fisher-Rao metrics each correspond to a manifold structure on probability distributions. Definitions of these distances and their associated tangent space structures will be given in this section.

The Wasserstein metric is an optimal transport distance that measures the cost of transporting one distribution to another, and can be defined in quite general spaces, (Villani, 2003; Ambrosio et al., 2008). Its use in a variety of statistical and machine learning problems has risen to prominence in the last decade, as reviewed in Panaretos and Zemel (2019, 2020). For the purposes of this review that focuses on samples of random univariate distributions, define

$$\mathcal{W}_2 = \{\mu : \mu \text{ is a probability measure on } \mathbb{R} \text{ and } \int_{\mathbb{R}} x^2\mathrm{d}\mu(x) < \infty\}. \tag{2.8}$$

For any $\mu, \nu \in \mathcal{W}_2$, define $\Pi(\mu, \nu)$ to be the set of joint measures (called transport plans) on $\mathbb{R}^2$ with marginals μ and ν. The 2-Wasserstein, or simply Wasserstein, distance between these measures is

$$d_W(\mu, \nu) := \left[\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^2} (x-y)^2\mathrm{d}\pi(x,y)\right]^{1/2}. \tag{2.9}$$

Let F_μ and F_ν be the cumulative distribution functions of these measures, define the quantile function

$$F_\mu^{-1}(t) = \inf\{x \in \mathbb{R} : F_\mu(x) \geq t\},$$

and similarly F_ν^{-1}. If μ is absolutely continuous, it is known that unique optimal transport plan π^* that achieves the infimum in (2.9) is the joint distribution of the pair (X^*, Y^*), where $X^* \sim \mu$ and $Y^* = F_\nu^{-1} \circ F_\mu(X^*)$. The map $T_\mu^\nu = F_\nu^{-1} \circ F_\mu$ is known as the optimal transport map from μ to ν. Thus, when μ is absolutely continuous, we have

$$\begin{aligned} d_W(\mu, \nu) &= \left[\int_{\mathbb{R}^2} (x-y)^2\mathrm{d}\pi^*(x,y)\right]^{1/2} = \left[\int_{\mathbb{R}} (T_\mu^\nu(x) - x)^2\mathrm{d}\mu(x)\right]^{1/2} \\ &= \left[\int_0^1 \left\{F_\mu^{-1}(t) - F_\nu^{-1}(t)\right\}^2\mathrm{d}t\right]^{1/2}, \end{aligned} \tag{2.10}$$

where the last line follows using the change of variables $t = F_\mu^{-1}(x)$. In fact, this last expression for the Wasserstein distance in terms of quantile functions holds generally for measures in $\mathcal{W}_2$, whether or not μ is absolutely continuous; see Theorem 2.18 of Villani (2003). Clearly, if $\mathcal{D}$ contains only densities f for which $\int_{\mathbb{R}} x^2 f(x)\mathrm{d}x < \infty$, it can be considered a subclass of absolutely continuous distributions in $\mathcal{W}_2$. We may then write $d_W(f, g)$ as a shorthand for the Wasserstein metric between densities in $\mathcal{D}$.

The manifold structure of $\mathcal{W}_2$ begins with the notion of a tangent space at each absolutely continuous measure $\mu \in \mathcal{W}_2$, as defined in Equation (8.5.1) of Ambrosio et al. (2008). Tangent spaces for measures that are not absolutely continuous can be defined similarly with some additional notation. Letting id represent the identity map, the tangent space is

$$\text{Tan}_\mu = \overline{\left\{\lambda(T_\mu^\nu - \text{id}) : \lambda > 0,\ \nu \in \mathcal{W}_2\right\}}, \tag{2.11}$$

where the closure is in $L^2(\mu)$. Two essential related operations are the logarithmic and exponential maps that map between $\mathcal{W}_2$ and Tan_μ, which are defined as

$$\begin{aligned} \text{Log}_\mu^{\text{W}}(\nu) &= T_\mu^\nu - \text{id}, \quad \nu \in \mathcal{W}_2 \quad \text{and} \\ \text{Exp}_\mu^{\text{W}}(V)(A) &= \mu\big[(V + \text{id})^{-1}(A)\big], \quad V \in \text{Tan}_\mu, \end{aligned} \tag{2.12}$$

where A is any Borel set. A variety of other useful properties of $\mathcal{W}_2$, such as the geodesic structure and regularity of optimal transport maps, can be found in the cited texts. When using the Wasserstein geometry for statistical modeling of distributions, a common approach is to specify a priori a reference measure $\mu_\oplus$ (often a Fréchet mean; see Section 3.1 below). Conditional on this choice, models can be built in the tangent space $\text{Tan}_{\mu_\oplus}$ after application of the logarithmic map, using the Hilbertian metric of $L^2(\mu_\oplus)$. An equivalent approach is to specify models for the optimal transport map between $\mu_\oplus$ and the random distribution being modeled. It should be observed that, while the tangent space can be shown to be linear, the image of the logarithmic map is not, so the use of $\text{Tan}_{\mu_\oplus}$ as a representation space is not another instance of the transformation approach to a Hilbert space.

The Fisher-Rao metric began as a Riemannian structure for parametric models, but has also been facilitated for generic densities. Although its definition and structure can be found in a variety of sources, we refer the reader to Srivastava et al. (2007) which specified these in the context of analyzing a sample of densities. To follow the authors discussion in this work, take $\mathcal{D}$ to be the set of nonnegative densities with support contained in [0,1]. The Fisher-Rao distance between densities $f, g \in \mathcal{D}$ is

$$d_{FR}(f, g) := \arccos\left(\int_0^1 \sqrt{f(x)g(x)}\text{d}x\right). \tag{2.13}$$

This distance is perhaps best understood by observing that the square root of a density lies on the Hilbert unit sphere, i.e. $\int_0^1 (\sqrt{f(x)})^2 \text{d}x = 1$, so that $d_{FR}(f, g)$ measures the length of an arch connecting $\sqrt{f}$ and $\sqrt{g}$ along this sphere. In other words, it is the spherical geodesic distance between square root densities. The tangent space at $f \in \mathcal{D}$ can then be identified with the orthogonal complement of $\text{span}\{\sqrt{f}\}$ in $L^2[0, 1]$. Letting $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ denote the usual $L^2[0, 1]$ inner product and norm, the corresponding logarithmic and exponential maps at a given $f \in \mathcal{D}$ are

$$\begin{aligned} \text{Log}_f^{\text{FR}}(g) &= \frac{u}{\|u\|} d_{FR}(f, g), \quad u = \sqrt{g} - \langle \sqrt{f}, \sqrt{g} \rangle \sqrt{f}, \; g \in \mathcal{D} \quad \text{and} \\ \text{Exp}_f^{\text{FR}}(\nu) &= \cos(\|\nu\|)\sqrt{f} + \sin(\|\nu\|)\frac{\nu}{\|\nu\|}, \quad \nu \in L^2[0, 1]. \end{aligned} \tag{2.14}$$

Although the use of the Fisher-Rao geometry in modeling densities as the primary data objects appears to be limited to the work in Srivastava et al. (2007), it has been extensively used in problems involving registration of functional data, (Srivastava et al., 2011; Tucker et al., 2013; Marron et al., 2015).

3. Methods for Analyzing Samples of Densities

A random density is a measurable map $\mathfrak{f} : (\Omega, \mathcal{S}, P) \to (\mathcal{D}, \mathcal{B}_\mathcal{D})$, where $(\Omega, \mathcal{S}, P)$ is a probability space and $\mathcal{B}_\mathcal{D}$ contains the Borel sets on $(\mathcal{D}, d)$. We assume that a sample of iid random densities $\mathfrak{f}_i$ is available for analysis. A technical point that will be suppressed in the sequel is that the $\mathfrak{f}_i$ are rarely, if ever, observed. Rather, one observes collections of scalars $\{U_{ij}\}_{j=1}^{N_i}$, $i = 1, \ldots, n$, where $U_{ij} \sim \mathfrak{f}_i$. Recovery of the densities by nonparametric density estimation prior to their analysis is a necessary preprocessing step, and poses an additional theoretical challenge when investigating asymptotic properties of inferential procedures. Although some of the theoretical work related to methods described in this section has carefully examined the propogation of errors elicited by such preliminary smoothing (e.g., Panaretos and Zemel, 2016; Petersen et al., 2016; Bigot et al., 2018; Petersen and Müller, 2019; Chen et al., 2021+), these details will not be incorporated into the discussion.

Beginning with basic data summaries and exploratory analysis of densities, this section outlines the definition of population targets and their sample versions using notions that are intrinsic to the space $\mathcal{D}$ of densities that is under consideration, drawing on the various representation spaces presented in Section 2. These methods will be illustrated using real data sets. These illustrations are intended to compare and contrast the various methods rather than rank them, as well as to provide guidance as to which method one might prefer for a given data set. A practical guide to the use of these methods is given in the supplementary vignette available online, which contains the code for all analyses.

3.1. The Fréchet Mean: A Representative Density

Assuming

$$\mathbb{E}\left[\int_{\mathbb{R}} \mathfrak{f}^2(x)\text{d}x\right] < \infty, \tag{3.1}$$

one can treat the random density $\mathfrak{f}$ as a random element of the Hilbert space $L^2 = L^2(\mathbb{R}, dx)$, the space of functions on the real line square integrable with respect to the Lebesgue measure because then $\int_{\mathbb{R}} \mathfrak{f}^2(x)dx < \infty$ almost surely. We emphasize that just as for the other spaces of densities, this is an additional condition to $\int_{\mathbb{R}} \mathfrak{f}(x)dx < \infty$ almost surely. Treating densities as elements of L^2 corresponds to treating them as usual functional data objects. The expected value or Hilbertian mean $\mathbb{E}[X]$ of a random element X in a separable Hilbert space $\mathbb{H}$ is defined by the condition $\langle \mathbb{E}[X], y\rangle = \mathbb{E}[\langle X, y\rangle]$, which must hold for any $y \in \mathbb{H}$. This is a special case of the definition of the expected value in a separable Banach space by means of a Pettis integral, see e.g. Chapter 7 of Laha and Rohatgi (1979). In case of a random function $f \in L^2(\mathbb{R}, dx)$ satisfying $\mathbb{E}[\int_{\mathbb{R}} f^2(x)dx] < \infty$, it is easy to verify that the function $\mathbb{E}[f]$ is equal almost everywhere to the pointwise expected value of f, i.e. $(\mathbb{E}[f])(x) = \mathbb{E}[f(x)]$ for almost all $x \in \mathbb{R}$. In case of any separable Hilbert space, it is also easy to verify that $\mathbb{E}[X]$ is the element $g \in \mathbb{H}$ minimizing $E[\|X - g\|^2]$. The verification uses the property that for any *scalar* random variable $\langle X, y\rangle$, $\mathbb{E}[\langle X, y\rangle]$ is $t \in \mathbb{R}$ minimizing $\mathbb{E}[(\langle X, y\rangle - t)^2]$. In the context of a random density satisfying (3.1), we can thus write

$$\mathbb{E}_{L^2}[\mathfrak{f}] = \operatorname*{argmin}_{g \in L^2(\mathbb{R})} \mathbb{E}[\|\mathfrak{f} - g\|_{L^2}^2] = \operatorname*{argmin}_{g \in L^2(\mathbb{R})} \mathbb{E}[d_{L^2}^2(\mathfrak{f}, g)].$$

If the space of densities $\mathcal{D}$ is constrained not by (3.1) but by different conditions, like those discussed in Section 2, and if $\mathcal{D}$ is equipped with metric d, it is useful to adopt the following definition.

Definition 3.1. Let $\mathfrak{f}$ be a random density taking values in $\mathcal{D}$. The Fréchet mean set of $\mathfrak{f}$ with respect to d is

$$f_{\oplus}^{d} = \operatorname*{argmin}_{g \in \mathcal{D}} \mathbb{E}\big[d^2(\mathfrak{f}, g)\big]. \tag{3.2}$$

Similarly, the sample Fréchet mean set of $\mathfrak{f}_1, \ldots, \mathfrak{f}_n$ is

$$\hat{f}_{\oplus}^{d} = \operatorname*{argmin}_{g \in \mathcal{D}} \sum_{i=1}^{n} d^2(\mathfrak{f}_i, g). \tag{3.3}$$

Definition 3.1 leaves open the possibility that $f_{\oplus}^{d}$ and $\hat{f}_{\oplus}^{d}$ contain more than one element, or are even empty. When the Fréchet mean set has a single element, it will be referred to as the Fréchet mean and denoted also by $f_{\oplus}^{d}$, and similarly for the sample Fréchet mean. Next, these objects will be investigated for the various metrics discussed in Section 2.

Bayes Space. Define B by (2.3), viewed as a set of equivalence classes, and let d_B be the metric induced by the inner product (2.5). Here, $\mathcal{D}$ is identified as the equivalences classes in B whose elements have a finite integral, represented by a probability density function. Recall the clr transformation (2.6) and its inverse (2.7). Then, for $f, g \in \mathcal{D}$,

$$d_B^2(f, g) = \int_I [\psi_{\mathrm{clr}}(f(x) - \psi_{\mathrm{clr}}(g)(x)]^2 \mathrm{d}x = \|\psi_{\mathrm{clr}}(f) - \psi_{\mathrm{clr}}(g)\|_{L^2}^2.$$

Suppose that $\mathbb{E}\Big[\int_I \{\psi_{\mathrm{clr}}(\mathfrak{f})(x)\}^2 \mathrm{d}x\Big] < \infty$, so that the Hilbertian mean $h_{\oplus}$ of the random element $\psi_{\mathrm{clr}}(\mathfrak{f})$ exists and is unique. By convexity arguments, for any $h_1, h_2 \in \psi_{\mathrm{clr}}(\mathcal{D})$ satisfying $\int_I \exp\{h_j(x)\}\mathrm{d}x < \infty$, one has $\int_I \psi_{\mathrm{clr}}^{-1}([h_1(x) + h_2(x)]/2)\mathrm{d}x < \infty$, from which it follows that $\int_I \exp\{h_{\oplus}(x)\}\mathrm{d}x < \infty$. Hence, the Fréchet mean set of $\mathfrak{f}$ with resepect to d_B is the single element

$$f_{\oplus}^{d_B}(\cdot) = \frac{\psi_{\mathrm{clr}}^{-1}(h_{\oplus})(\cdot)}{\int_I \psi_{\mathrm{clr}}^{-1}(h_{\oplus})(x)\mathrm{d}x} = \frac{\exp\{h_{\oplus}(\cdot)\}}{\int_I \exp\{h_{\oplus}(x)\}\mathrm{d}(x)}. \tag{3.4}$$

The sample Fréchet mean is also unique by the same convexity arguments, taking the form

$$\hat{f}_{\oplus}^{d_B}(\cdot) = \frac{\exp\{\hat{h}_{\oplus}(\cdot)\}}{\int_I \exp\{\hat{h}_{\oplus}(x)\}\mathrm{d}x}, \quad \hat{h}_{\oplus} = \frac{1}{n}\sum_{i=1}^{n} \psi_{\mathrm{clr}}(\mathfrak{f}_i). \tag{3.5}$$

Wasserstein Metric. When working with the Wasserstein metric d_W in (2.10), we set $\mathcal{D}$ to be the set of probability density functions whose measures lie in $\mathcal{W}_2$, defined in (2.8). As the Wasserstein metric was defined in terms of measures, let μ (μ_i) be the random measure corresponding to the density $\mathfrak{f}$ ($\mathfrak{f}_i$). In analogy to (3.2) and (3.3), one can define a Fréchet mean set of μ consisting of measures, and similarly the sample Fréchet mean set. Proposition 3.2.3 of Panaretos and Zemel (2020) implies that this Fréchet mean set is nonempty for any random measure μ taking values in $\mathcal{W}_2$, while Proposition 3.2.7 of the same text demonstrates that this mean is unique once μ is assumed to be absolutely continuous with positive probability. Hence, since it is already assumed that μ arises from a random density $\mathfrak{f}$, one immediately obtains existence and uniqueness of a Fréchet mean measure, but this measure need not possess a density. However, in our case of probability densities on the real line, Theorem 5.5.2 of Panaretos and Zemel (2020) implies that, provided $\mathfrak{f}$ is bounded with positive probability, the Fréchet mean $f_{\oplus}^{d_W}$ exists as a (bounded) density. In light of (2.10), it is not surprising that this mean admits the closed form expression

$$f_{\oplus}^{d_W} = \big[F_{\oplus}^{d_W}\big]', \quad (F_{\oplus}^{d_W})^{-1}(t) = \mathbb{E}[\mathfrak{F}^{-1}(t)],\ \mathfrak{F}(\cdot) = \int_{-\infty}^{\cdot} \mathfrak{f}(x)\mathrm{d}x. \tag{3.6}$$

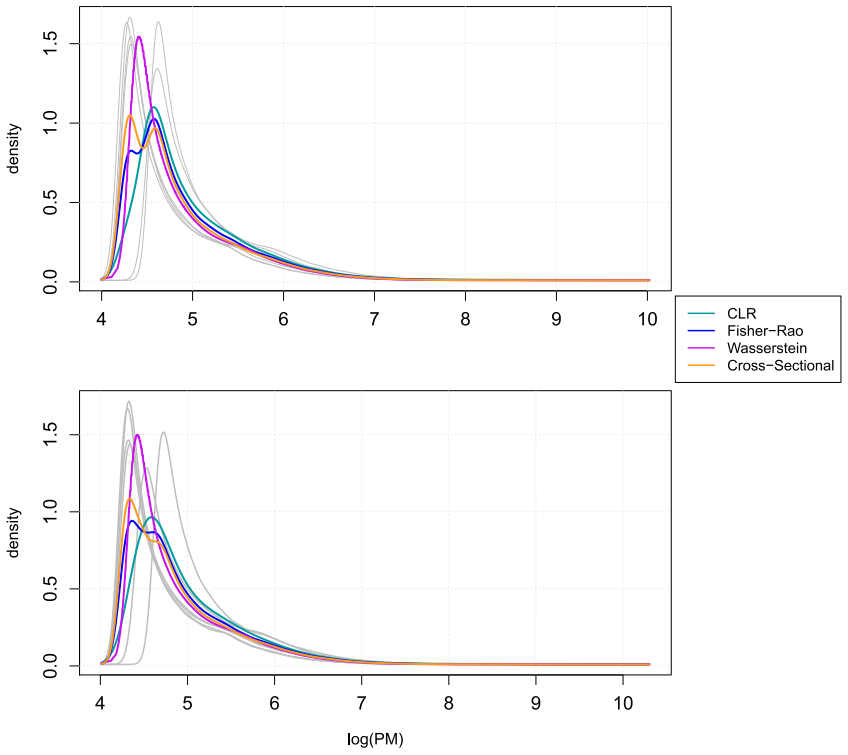


Fig. 1. Sample Fréchet mean densities of Log(PM) for the normal 2N (top) and the Ts1Cje (bottom) mice. The light grey-colored lines in the background are the raw densities. Labels for the metrics are as follows: 'CLR': d_B, 'Fisher-Rao': d_{FR}, 'Wasserstein': d_W, 'Cross-Section': d_{L^2} (i.e., the pointwise average of densities).

Here, $\mathfrak{F}^{-1}$ is the random quantile function corresponding to $\mathfrak{f}$. Put succinctly, under the stated regularity conditions, the unique Fréchet mean density can be obtained by first computing the usual Hilbertian mean of $\mathfrak{F}^{-1}$, and then converting this into a probability density.

The case of sample Fréchet means is analogous, with $\hat{f}_{\oplus}^{d_W}$ being a unique, bounded density as long as one of the $\mathfrak{f}_i$ is bounded; see Proposition 5.1 in Agueh and Carlier (2011). Its computation is also facilitated via quantile function as

$$\hat{f}_{\oplus}^{d_W} = \left[\hat{F}_{\oplus}^{d_W}\right]', \quad (\hat{F}_{\oplus}^{d_W})^{-1}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathfrak{F}_i^{-1}(t). \tag{3.7}$$

Panaretos and Zemel (2016) derived rates of convergence of the sample Fréchet mean to the population version under the Wasserstein metric, even incorporating the preliminary step of density estimation, while Bigot et al. (2018) later showed that these rates are minimax. When the sample sizes for each density are large enough, Panaretos and Zemel (2016) established a central limit theorem appropriate for the Wasserstein geometry.

Fisher-Rao Metric. Existence and uniqueness of population and sample Fréchet means under the metric d_{FR} in (2.13) have not been investigated. However, Srivastava et al. (2007) presented a shooting algorithm for computation of a sample Karcher mean density under this metric, corresponding to a stationary point of the objective in (3.3).

Data Illustration. As a brief illustration of the various sample Fréchet means discussed above, we consider a data example from Amano et al. (2004) (also analyzed in Zhang and Müller, 2011) in which global gene expression files were obtained from the brains of six Ts1Cje mice (used to model Down syndrome) and six normal (2N) mice using DNA microarrays. Figure 1 shows probability density functions obtained by kernel smoothing of samples of a quantity log(PM) obtained from each array, with the densities grouped as Ts1CJe/2N, with one density per mouse. Overlaying these densities are sample Fréchet means corresponding to various metrics.

The smoothed densities are each unimodal, with the primary difference between the densities in each group being the location of the mode, with the height of the mode a secondary source of variabiity especially in the Ts1Cje mice. For this data set, both the cross-sectional (i.e., ordinary functional) and Fisher-Rao means appear bimodal as they both attempt to average out pointwise behavior in the densities or square-root densities. The Fréchet mean derived from the Bayes space and Wasserstein metrics are more representative of the sample, as they are both unimodal. However, the height of the peak in the Wasserstein mean is more similar to those of the sample.

3.2. Exploring Variability and Dimension Reduction

Functional principal component analysis (FPCA), e.g., Ramsay and Silverman (2005); Kokoszka and Reimherr (2017), is a common tool in functional data analysis for performing exploratory analysis of the main sources of variability in the sample, Castro et al. (1986); Jones and Rice (1992), as well as dimension reduction. FPCA is based on the Karhunen-Loève decomposition of a random element Y of a Hilbert space, and the dimension reduction produced by truncating this expansion is optimal in terms of variance explained, where variance is defined in terms of the Hilbertian metric. However, for nonlinear spaces such as $\mathcal{D}$, linear methods of dimension reduction and exploratory analysis are inappropriate. For instance, functional modes of variation constructed from FPCA on densities cannot be guaranteed to remain bona fide densities, and the resulting dimensionality reduction is typically suboptimal if variance is quantified in terms of an intrinsic density metric rather than the usual L^2 metric used for functional data. Nevertheless, the computational simplicity and ease of interpretation of FPCA have led to several adaptations for density data that address the above concerns.

Transformation FPCA. Let ψ be a generic transformation mapping densities into a Hilbert space of functions $L^2(\mathcal{T})$, as outlined in Section 2.1, where $\mathcal{T} \subset \mathbb{R}$ is an interval; ψ_{LQD} in (2.1) and ψ_{clr} in (2.6) constitute specific examples. Petersen et al. (2016) proposed to apply standard FPCA to the transformed functional variable $Y = \psi(\dagger)$, while Hron et al. (2016) proposed in parallel the specific case of the transformation ψ_{clr}, which the authors termed Simplicial Principal Component Analysis due to its relation to compositional data analysis. For a generic transformation, one begins with the mean and covariance functions $\nu(t) = \mathbb{E}[Y(t)]$ and $G(s,t) = \mathrm{Cov}(Y(s), Y(t))$ $(s,t \in \mathcal{T})$ of Y, leading to the Karhunen-Loève decomposition, (Hsing and Eubank, 2015),

$$Y(t) = \nu(t) + \sum_{j=1}^{\infty} \xi_j \phi_j(t), \quad \xi_j = \int_{\mathcal{T}} [Y(t) - \nu(t)]\phi_j(t)\mathrm{d}t, \tag{3.8}$$

where the ϕ_j arise from the Mercer decomposition $G(s,t) = \sum_{j=1}^{\infty} \lambda_j \phi_j(s)\phi_j(t)$, and thus form an orthonormal basis of $L^2(\mathcal{T})$. The ϕ_j reflect the most dominant directions of variability in Y about its mean, and suggest the modes of variability

$$\nu(t) + \alpha\sqrt{\lambda_j}\phi_j(t), \quad \alpha \in \mathbb{R}, t \in \mathcal{T}, \tag{3.9}$$

as a device for visualizing this variability. Varying α allows one to visualize the effect the random score ξ_j in (3.8) has on this variability, and specifically represents the number of standard deviations ξ_j is away from its mean (zero) since $\mathrm{Var}(\xi_j) = \lambda_j$. These modes represent variability in the transformed space and can be difficult to interpret in terms of variability on the scale of densities, so Petersen et al. (2016) defined the transformation modes of variation

$$g_j(x,\alpha) = \psi^{-1}\left(\nu + \alpha\sqrt{\lambda_j}\phi_j\right)(x), \quad \alpha, x \in \mathbb{R}. \tag{3.10}$$

The advantage of performing FPCA in the transformed space, followed by application of the inverse transformation, is that the modes g_j will always correspond to densities regardless of the value of α used. Practically speaking, one should avoid excessive extrapolation of α ouside of the range of the data to prevent unnatural distortions in the modes.

Application of these ideas to a sample $f_i, 1 \le i \le n$, is straightforward. One first obtains the transformed functional observations $Y_i = \psi(f_i)$, followed by application of FPCA to this sample, leading to estimates $\hat{\nu}$, $\hat{\lambda}_j$, and $\hat{\phi}_j$. These are used to construct the sample transformation modes

$$\hat{g}_j(x,\alpha) = \psi^{-1}\left(\hat{\nu} + \alpha\sqrt{\hat{\lambda}_j}\hat{\phi}_j\right)(x), \quad \alpha, x \in \mathbb{R}. \tag{3.11}$$

Petersen et al. (2016) derived uniform convergence rates of the estimated modes to their population targets under structural assumptions on the transformation ψ. These rates also accounted for the preliminary step of density estimation.

Dimension reduction is achieved by constructing a truncated version of (3.8) using sample estimates. Using the first J functional principal component score estimates $\hat{\xi}_{ij} = \int_{\mathcal{T}} (Y_i(t) - \hat{\nu}(t))\hat{\phi}_j(t)\mathrm{d}t$, the J-dimensional representation of each density is

$$\hat{f}_{iJ}(x) = \psi^{-1}\left(\hat{\nu} + \sum_{j=1}^{J} \hat{\xi}_{ij}\hat{\phi}_j\right)(x), \quad x \in \mathbb{R}. \tag{3.12}$$

Tangent Space Log-FPCA. For metrics d under which $\mathcal{D}$ possesses a (pseudo) Riemannian structure, such as the Wasserstein and Fisher-Rao metrics, a computationally simple method is to first compute the sample Fréchet mean $\hat{f}^d_{\oplus}$, followed by mapping the densities to the tangent space at $\hat{f}^d_{\oplus}$, and applying ordinary FPCA in the tangent space, (Fletcher et al., 2004). Using Log and Exp as generic logarithmic and exponential maps, one computes $Y_i = \mathrm{Log}_{\hat{f}^d_{\oplus}}(f_i)$. Since these transformed functions reside in the tangent space at $\hat{f}^d_{\oplus}$, which is indeed a Hilbert space, one can apply FPCA directly to the Y_i as in the above transformation case. The j-th Log-FPCA sample mode of variation then becomes

$$\hat{g}_j(x,\alpha) = \mathrm{Exp}_{\hat{f}^d_{\oplus}}\left(\hat{\nu} + \alpha\sqrt{\hat{\lambda}_j}\hat{\phi}_j\right)(x), \quad \alpha, x \in \mathbb{R}. \tag{3.13}$$

Similarly, one can construct J-dimensional representations of the densities via

$$\hat{f}_{i,J}(x) = \mathrm{Exp}_{\hat{f}_\oplus^d}\left(\hat{v} + \sum_{j=1}^{J} \hat{\xi}_{ij}\hat{\phi}_j\right), \quad x \in \mathbb{R}. \tag{3.14}$$

Although this description is computationally identical to the transformation case, it should be noted that the logarithmic map is not a bijective map from $\mathcal{D}$ to the tangent space, and thus does not constitute a valid transformation in the sense of Section 2.1. As pointed out by Cazelles et al. (2018) in the case of the Wasserstein geometry, although the exponential map can be applied to any element formed in the tangent space to yield a valid density, it may yield distortions when this mode lies outside of the image of the logarithmic map, including densities that have vastly different supports than the observed f_i.

Wasserstein Geodesic PCA. In the case of the Wasserstein metric, the shortfalls of tangent space Log-FPCA prompted the development of principal *geodesic* analysis for densities by Bigot et al. (2017); Cazelles et al. (2018). Bigot et al. (2017) investigated the case of densities with support contained in a given interval of the real line, and defined the notion of geodesic subspaces of the Wasserstein manifold. These geodesic subspaces possess an intrinsic dimension, and replace the linear subspace structures used to define FPCA. Estimators of the principal geodesics from sampled densities were proposed, along with derivation of asymptotic statistical properties. Cazelles et al. (2018) extended geodesic FPCA for densities to the case of multidimensional support. Since computation of principal geodesic subspaces of more than one dimension can be quite expensive and do not generally yield a sequence of nested subspaces as the dimension is increased (as would naturally be a desirable feature for dimension reduction), Cazelles et al. (2018) proposed an alternative iterative approach yielding a sequence of 1-dimensional orthogonal geodesics. In either case, the principles of forming modes of variation and constructing finite-dimensional approximations of the observed densities are the same as those for Log-PCA. First, for a chosen dimension, one forms a principal geodesic surface, followed by projecting the Log-mapped densities onto this surface in the tangent space. Alternatively, using the iterative approach, the intersection of the image of $\mathcal{W}_2$ under the Log map at the Fréchet mean with the span of a sequence of orthogonal geodesics can be formed. Finally, the exponential map is applied to yield a mode of variation or finite-dimensional approximation in density space.

Quantifying Variance Explained. Given the many ways one can choose to explore variability and perform dimension reduction for densities, some objective assessment is helpful for comparing their performance for a given data set. Because modes of variation are used as a visual tool, a subjective assessment of their interpretability is typically used to determine a preferred method. However, in the case of dimension reduction, Petersen et al. (2016) proposed a metric-based measurement of variance explained to numerically compare the efficiency of different dimension reduction methods. Let d be a chosen metric for $\mathcal{D}$. In analogy to the usual variance of a scalar random variable and using the definition of the Fréchet mean in (3.2), the Fréchet variance of a random density $\mathfrak{f}$ is

$$V_\oplus = \mathbb{E}[d^2(\mathfrak{f}, f_\oplus^d)]. \tag{3.15}$$

Given a sample $\mathfrak{f}_i$, the sample Fréchet variance is defined analogously as

$$\hat{V}_\oplus = \frac{1}{n}\sum_{i=1}^{n} d^2(\mathfrak{f}_i, \hat{f}_\oplus^d). \tag{3.16}$$

If $\hat{f}_{i,J}$ are J-dimensional approximations computed from any given method, one can extract the fraction of variance explained as

$$\mathrm{FVE}_J^d = \frac{\hat{V}_J}{\hat{V}_\oplus}, \quad \hat{V}_J = \hat{V}_\oplus - \sum_{i=1}^{n} d^2(\mathfrak{f}_i, \hat{f}_{i,J}). \tag{3.17}$$

Methods that result in higher values of FVE_J^d for several J can be considered superior for a given data set.

Data Illustration. Mortality rates have long been objects of intense study in actuarial science and other fields. Data for regions of interest are often provided in the form of cross-sectional lifetables, whereby the number of people who died at a given age during a fixed year are recorded, where the total number of people is typically normalized to a fixed constant to allow for comparisons. The Human Mortality Database (www.mortality.org) currently provides such lifetables for 41 countries throughout the world, with records stretching back many decades. For a given country and year, one can compute a histogram of age-at-death, followed by smoothing to obtain a density. Following this procedure, Figure 2 plots the age-at-death density functions for $n = 40$ countries from the year 2008, conditional on reaching 20 years of age.

There is a large degree of similarity among the age-at-death distributions, making dimension reduction a very natural tool for their analysis. The most prominent sources of variability visible in this sample are the location and sharpness of the distributional modes.

Figures 3 and 4 plot the first and second modes of variation, respectively, for four different methods. The first two apply the transformation FPCA using ψ_{LQD} in (2.1) and ψ_{clr} in (2.6), the third applies Log-FPCA using the Fisher-Rao geometry, and the last applies geodesic FPCA as described in Cazelles et al. (2018). The first modes of variation show that the largest variability among the samples occurs in two age ranges: (40, 70) and (80-100), reflecting the structure highlighted in Figure 2. In particular, the first modes of variation captures the differences between countries in which deaths tended to occur at

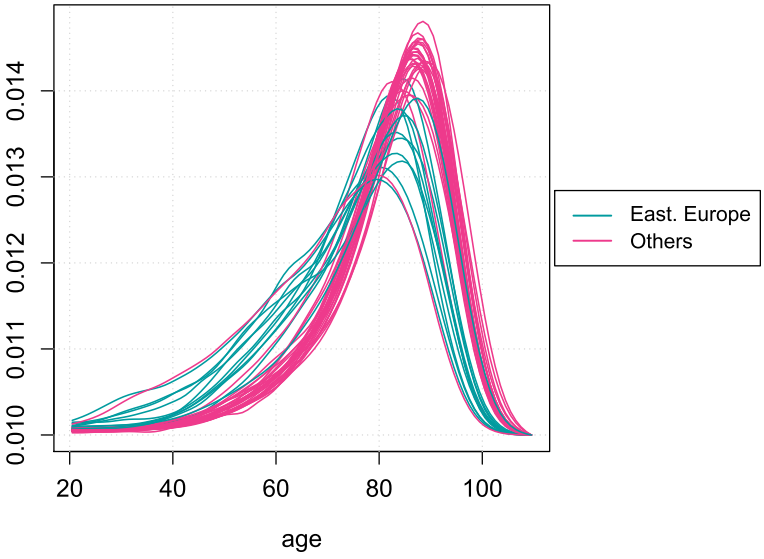


Fig. 2. Sample of density functions corresponding to the distribution of age-at-death for 40 countries in 2008. Densities were obtained by smoothing histograms formed by normalized lifetables. Data provided by the Human Mortality Database. Color corresponds to a binary variable indicating whether or not a country is located in Eastern Europe.

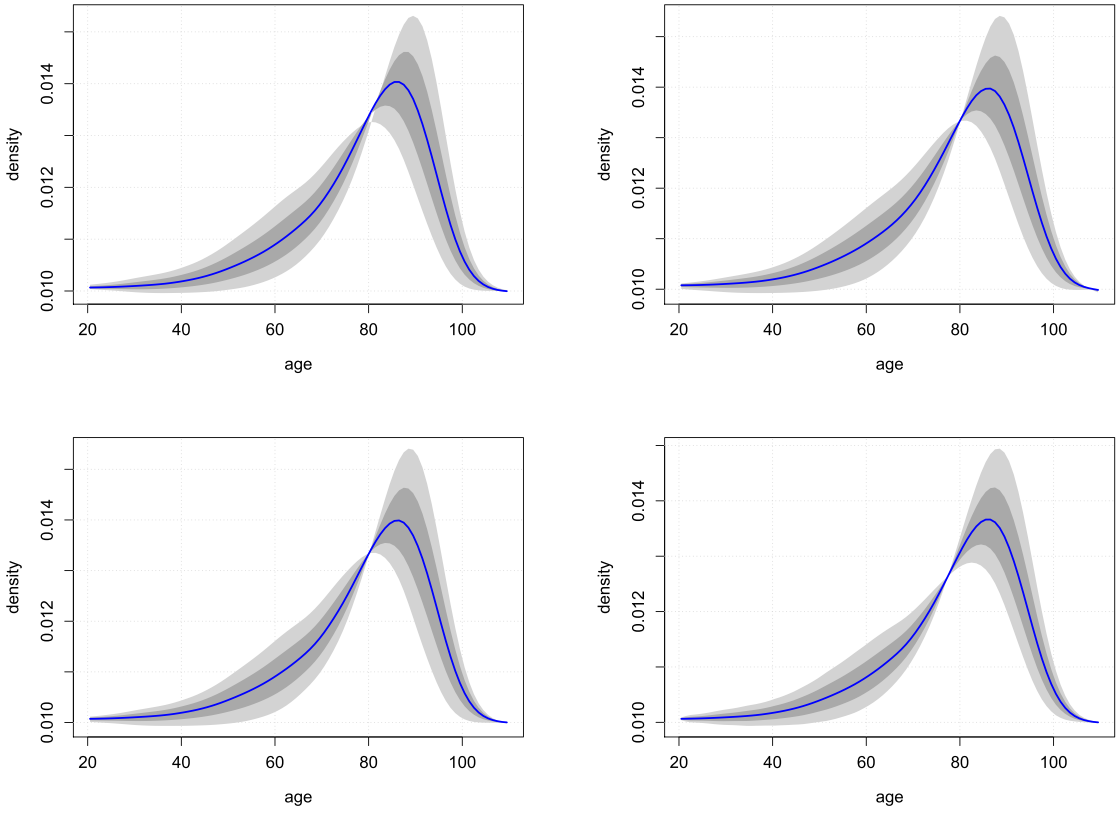


Fig. 3. First mode of variation extracted from the age-at-death distributions in Figure 2 using (top left) transformation FPCA with ψ_{LQD}, (top right) transformation FPCA with ψ_{clr}, (bottom left) tangent space FPCA under the Fisher-Rao geometry, and (bottom right) geodesic FPCA using the Wasserstein geometry. The blue line corresponds to $\alpha = 0$ in (3.11) and (3.13), while the boundaries of the dark and light shaded regions correspond to $\alpha = \pm 1, \pm 2$, respectively.

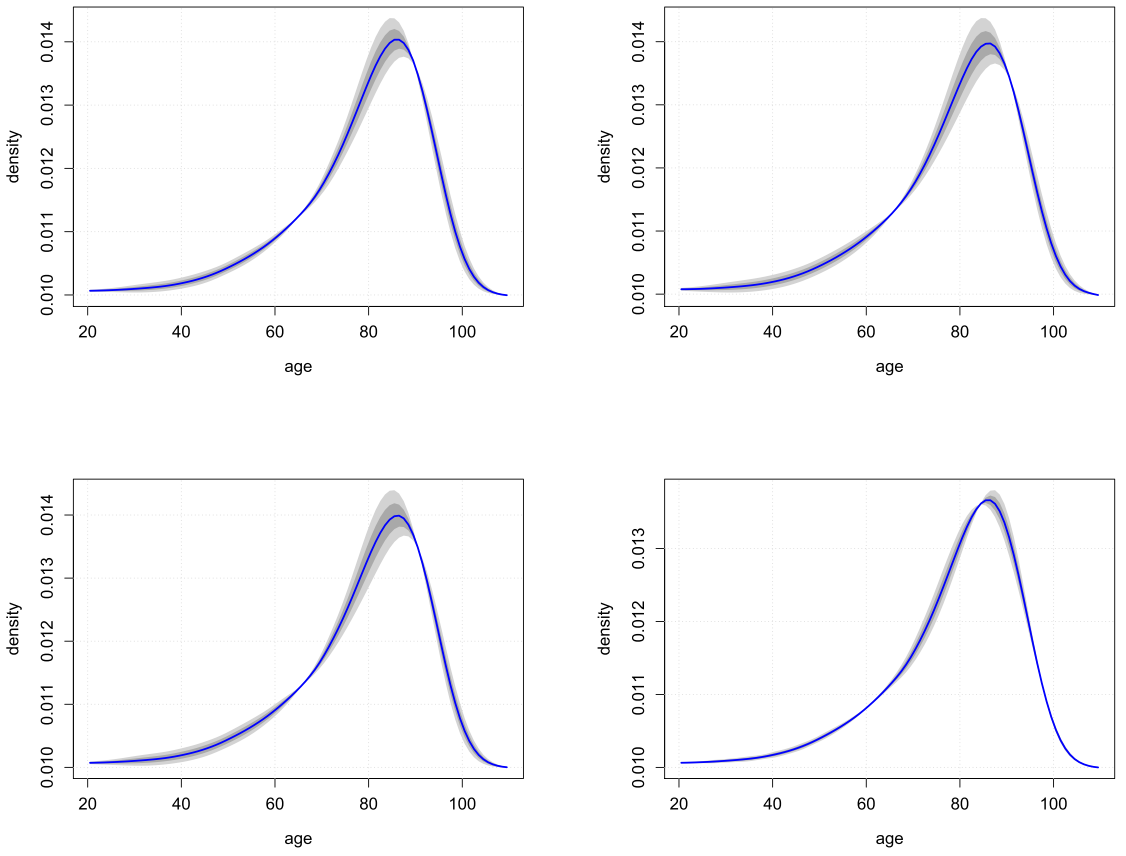


Fig. 4. Second mode of variation extracted from the age-at-death distributions in Figure 2 using (top left) transformation FPCA with ψ_{LQD}, (top right) transformation FPCA with ψ_{clr}, (bottom left) tangent space FPCA under the Fisher-Rao geometry, and (bottom right) geodesic FPCA using the Wasserstein geometry. The values of α used are the same as described in the caption of Figure 4.

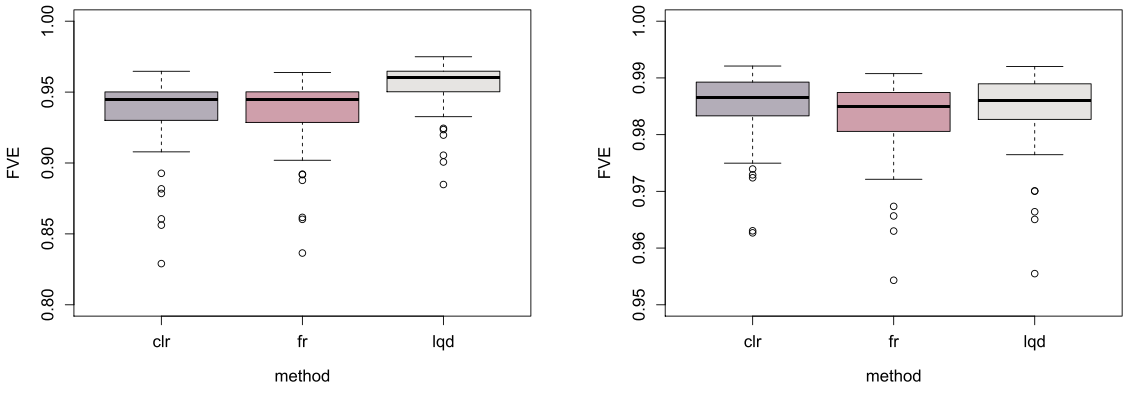
lower ages, but with high variability, and those with comparatively older ages at death and lower variability. One can also observe that the peak of the first modes of variation reflected by Geodesic PCA is slightly lower than the other methods. The second modes of variation are noticeably less prominent, and highlight variability around age 80 that was not captured by the first mode of variation. Geodesic PCA is again the outlier, as its second mode of variation seems to capture a similar pattern as that of the first mode, but contrasting the age ranges of (30, 60) and (70, 90).

As the first three methods give qualitatively the same message in terms of interpreting the variability in the sample of densities in the first two dimensions, we can use the FVE metrics described above to compare their relative efficiency in explaining the variability in the data. Figure 5 shows boxplots of FVE values for the first three dimension reduction methods as they relate to the Wasserstein metric d_W; FVE values for other metrics are given in the supplementary vignette. These boxplots were obtained by repeatedly dividing the data into training and testing sets, where the training set was used to estimate the components of the FPCA and the testing set was then used to compute the FVE. By reducing the sample of densities down to the first dimension, the FVE values are all above 80%, with the median for the LQD transformation exceeding 95%. Adding the second dimension results in all three methods nearly always achieving more than 95% of out-of-sample variance explained.

3.3. Density Response Regression

For many data sets that feature density functions, one also has available an assortment of covariates, which can be a mixture of categorical and numeric variables. In such situations, it is often of interest to build and estimate regression models that feature a random density $\mathfrak{f}$ as the response variable. A naive approach to this type of regression problem would be to assume a functional linear model, (Ramsay and Silverman, 2005; Faraway, 1997),

$$\mathfrak{f}(x) = \beta_0(x) + \sum_{j=1}^{p} U_{ij}\beta_j(x) + \epsilon_i(x), \tag{3.18}$$



(a) First dimension (b) First two dimensions

Fig. 5. Fraction of Variance Explained (FVE) values as in (3.17) under the Wasserstein metric $d = d_W$ for $J = 1$ (left) and $J = 2$ (right).

where U_{ij} are the covariates, ϵ_i are error processes, and β_j, $j = 0, \ldots, p$ are the functional parameters. However, the linear structure of this model ultimately renders it incoherent for representing random densities with their nonlinear constraints.

Instead, one may extend regression to metric spaces in the same way that the Fréchet mean generalizes ordinary expectations to such spaces. Consider the following definition.

Definition 3.2. Let $(U, \mathfrak{f})$ be a random element of $\mathbb{R}^p \times \mathcal{D}$, and d a metric on $\mathcal{D}$. The conditional Fréchet mean set of $\mathfrak{f}$ given $U = u$, with respect to d, is

$$f^d_\oplus(u) = \operatorname*{argmin}_{g \in \mathcal{D}} \mathbb{E}\big[d^2(\mathfrak{f}, g)|U = u\big]. \tag{3.19}$$

Since the goals of regression often involve inference beyond simple estimation, it is desirable to build models for these conditional Fréchet means, akin to parametric models for scalar response variables such as linear regression. Such models potentially allow for hypothesis testing for predictor effects as well as the construction of confidence bands for the conditional Fréchet mean densities. One such model was proposed by Petersen et al. (2019b) for a response in a generic metric space, termed Fréchet regression. Written in terms of the space $\mathcal{D}$ of densities, the Fréchet regression model is

$$f^d_\oplus(u) = \operatorname*{argmin}_{g \in \mathcal{D}} \mathbb{E}\big[s(U, u)d^2(\mathfrak{f}, g)\big], \quad s(U, u) = 1 + (U - \mathbb{E}[U])^\top \mathrm{Var}(U)^{-1}(u - \mathbb{E}[U]). \tag{3.20}$$

The intuition behind this model was that, if the random density response is replaced by a scalar random variable Y, and $\mathcal{D}$ is replaced by $\mathbb{R}$ and d by the absolute value metric, the model simplifies to

$$\mathbb{E}[Y|U = u] = \mathbb{E}[Y] + \mathrm{Cov}(Y, U)\mathrm{Var}(U)^{-1}(u - \mathbb{E}[U]),$$

and is thus a direct generalization of multiple linear regression. Indeed, the following regression models considered in the literature for density responses obey (3.20) for different metrics d.

Bayes space regression. Talská et al. (2018) proposed a functional response model for density responses by exploiting the compositional or Bayes space structure. Defining $\mathcal{D}$, d_B, and ψ_{clr} as in Section 2.2, the model is

$$\psi_{\mathrm{clr}}(\mathfrak{f}_i) = \beta_0 + \sum_{j=1}^{p} U_{ij}\beta_j + \epsilon_i, \tag{3.21}$$

where $\beta_j \in L^2_0(I)$ are the functional parameters and ϵ_i are zero-mean random elements of $L^2_0(I)$. Compared to (3.18), (3.21) imposes the linear model appropriately on the transformed densities that indeed constitute a linear space. Moreover, (3.21) and (3.20) are equivalent upon taking $d = d_B$ in (3.20). Talská et al. (2018) illustrated how to apply tools designed for the standard functional linear model to their transformed case, including B-spline representations that can handle the discrete nature of the observed data in practical scenarios when the $\mathfrak{f}_i$ are not directly observed. Asymptotic properties and theoretically justified inferential procedures remain an open problem for this model. We also mention that, as ψ_{clr} is a special case of the general transformation approach of Petersen et al. (2016), one can build similar regression models for any suitable transformation ψ, including ψ_{LQD}. Specifically, replacing ψ_{clr} in (3.21) with a generic transformation ψ that maps to a Hilbert space $\mathbb{H}$ (with norm $\|\cdot\|_{\mathbb{H}}$), this generalized version of (3.21) becomes equivalent to (3.20) upon taking $d(f, g) = \|\psi(f) - \psi(g)\|_{\mathbb{H}}$.

Wasserstein Regression. Petersen et al. (2021) applied the Wasserstein metric d_W to model (3.20), along with algorithms for computing sample estimates $\hat{f}^{d_W}_\oplus(u)$. Unlike the Bayes space regression, this model does not possess and additive linear

Table 1
Fréchet regression summary

	F-statistic	*p*-value (truncation)	*p*-value (Satterthwaite)
gdp	0.046	0.605	0.607
inf	4.708	0.002	< 0.001
uem	1.602	0.014	0.008
Global F-statistic (by Satterthwaite method): 49.213 on 3.328 DF, p-value: 1.965e-10			

structure like (3.21) due to the nonlinear geometry of $\mathcal{W}_2$. Instead, Petersen et al. (2021) considered the optimal transport errors $T_i = \mathfrak{F}_i^{-1} \circ F_{\oplus}^{d_W}(U_i)$, where $\mathfrak{F}_i$ is the random quantile function corresponding to $\mathfrak{f}_i$, and $F_{\oplus}^{d_W}(u)$ is the distribution function of $f_{\oplus}^{d_W}(u)$. In other words, T_i is the optimal transport map from the conditional Fréchet mean density to the observed one, and is thus the appropriate notion of error under the Wasserstein geometry.

By imposing distributional assumptions on the maps T_i, Petersen et al. (2021) developed inferential procedures. For instance, consider the null hypothesis of no effect, in which the conditional Fréchet means $f_{\oplus}^{d_W}(u)$ all coincide with the marginal Fréchet mean $f_{\oplus}^{d_W}$, i.e., $\mathcal{H}_0 : f_{\oplus}^{d_W}(u) \equiv f_{\oplus}^{d_W}$. Setting $\hat{\mathfrak{f}}_i = \hat{f}_{\oplus}^{d_W}(U_i)$ to be the fitted values and $\hat{f}_{\oplus}^{d_W}$ the sample Fréchet mean, the test statistic for this global null hypothesis is

$$F_G^* = \sum_{i=1}^{n} d_W^2(\hat{\mathfrak{f}}_i, \hat{f}_{\oplus}^{d_W}), \tag{3.22}$$

mimicking the sum of squares due to regression from multiple linear regression models. It was shown that, conditional on the observed predictors, F_G^* converges weakly (for almost all sequences of predictors) to an infinite mixture of independent χ_p^2 random variables, where the mixture weights are related to the common covariance kernel of the random optimal transport processes T_i. A similar statistic was investigated for hypotheses of partial effects, including individual covariate effects. Asymptotic consistency and power analyses of these tests were also established.

Two types of confidence bands were investigated for the conditional Fréchet means. The first, as motivated by the Wasserstein metric, involves the optimal transport map $\hat{V}_u$ from the target $f_{\oplus}^{d_W}(\mathrm{u})$ to its estimate. Weak convergence of $\sqrt{n}(\hat{V}_u - \mathrm{id})$ to a Gaussian process limit was established, and the resulting confidence band was shown to result in a subset of distributions that can be interpreted as a bracket in the stochastic ordering of distributions. The second approach resulted in a traditional confidence band of the form

$$\hat{f}_{\oplus}^{d_W}(u, x) \pm q_\alpha \mathrm{SE}(\hat{f}_{\oplus}^{d_W}(u, x)),$$

where q_α and $\mathrm{SE}(\hat{f}_{\oplus}^{d_W}(u)$ are the quantile and standard deviation estimates arising from the Gaussian limiting process of $\sqrt{n}(\hat{f}_{\oplus}^{d_W}(u) - f_{\oplus}^{d_W}(u))$ in the space of bounded functions with the uniform metric. Hence, these bands are simultaneous in the functional argument x of the density, but not across different values u of the predictor.

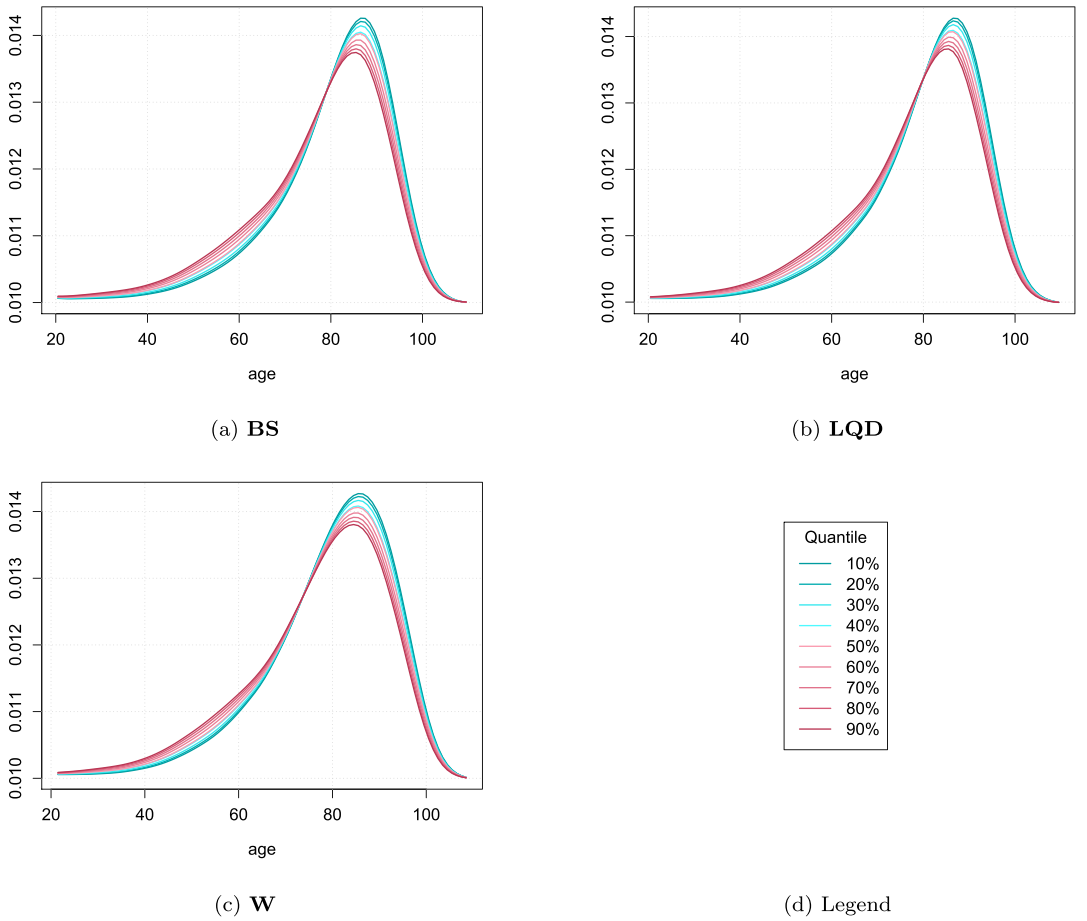
Smoothing Methods. Instead of developing a parametric-type model for the conditional Fréchet means, smoothing approaches have been investigated. For instance, Petersen et al. (2019b) demonstrated a version of local Fréchet regression for density response data with a single predictor, similar to local linear regression for scalar responses. When several predictors are present, Han et al. (2020) proposed to use the transformation approach of Petersen et al. (2016) to estimate a functional additive model.

Data Illustration. We will illustrate density response regression using the mortality data (see Figure 2) with three economic indicators as predictors. Since the age-at-death distributions correspond to the year 2008, the economic indicators were taken from 1998 to measure their effect on mortality after one decade. U_{i1} and U_{i2} are the percentage change in gross domestic product (gdp) and inflation rate (inf) from the previous year, and U_{i3} is the unemployment rate (uem). These predictors were used to fit three versions of the Fréchet regression model in (3.20), corresponding to Bayes space regression (**BS**), the functional response model of Faraway (1997) applied to the LQD-transformed densities (**LQD**), and regression under the Wasserstein metric (**W**).

As a starting point, Table 1 reports the *p*-values from the fit of model **W**, using the methods of Petersen et al. (2021). As the reference distributions for the statistics correspond to infinite mixtures of χ^2 random variables, the *p*-values were computed in two different ways; first, by truncating to a finite mixture and obtaining the reference distribution by simulation using estimated mixture coefficients and, second, by using Satterthwaite's apprxoimation (Satterthwaite, 1941). Inflation and unemployment appear to be the important predictors from this fit.

The effect of the fitted regression models are best interpreted through effects plots, whereby all but one of the predictors are set to fixed values, and the estimated conditional Fréchet mean densities are visualized over a grid of values for a chosen focal predictor. For instance, Figure 6 demonstrates the effects of the unemployment rate on the fitted mean densities when gdp and inf are set to their sample averages. The different models agree that lower levels of unemployment cause an increase in longevity of the population.

Another advantage of the Fréchet regression model is that, like multiple linear regression, categorical information can be easily incorporated. For instance, as seen in Figure 2, countries in Eastern Europe (EE) tended to have distributions



(a) **BS** (b) **LQD**

(c) **W** (d) Legend

Fig. 6. Effects plots for Fréchet regression of age-at-death densities in Figure 2 using GDP, inflation, and unemployment as predictors. Shown are the estimated conditional mean densities under the three models with GDP and inflation set to their sample averages, and for varying rates of unemployment as indicated by the colored legend.

indicative of shorter life spans. Adding this binary indicator into the model, inference under model **W** indicates that is highly significant with a p-value of 0.002. A more informative comparison between Eastern European and other countries is demonstrated in Figure 7, which shows an effects plot for this indicator with gdp, inf, and uem set to their sample averages. A simultaneous confidence band is also shown for each fitted conditional Fréchet mean density, demonstrating significantly different risks of death during ages 40-70 and again after reaching 80 years of age. The notably wider band for EE countries is due to the fact that fewer than a quarter of the sampled countries are located there.

To compare different models for a single data set, a generalization of the coefficient of determination was suggested by Petersen et al. (2019b). Similar to the metric-based FVE in (3.17), let d be a metric chosen for assessment, and $\hat{f}_{\oplus}^{d}$ the corresponding sample Fréchet mean. For a generic regression model, one obtains fitted densities $\hat{\mathfrak{f}}_i$, from which one can compute the Fréchet coefficient of determination

$$\hat{R}^2_{\oplus} = 1 - \frac{\sum_{i=1}^n d^2(\mathfrak{f}_i, \hat{\mathfrak{f}}_i)}{\sum_{i=1}^n d^2(\mathfrak{f}_i, \hat{f}_{\oplus}^{d})}. \tag{3.23}$$

Just as the Fréchet regression model simplifies to multiple linear regression when the response is a scalar random variable, $\hat{R}^2_{\oplus}$ also simplifies to the usual coefficient of determination in this setting. Table 2 gives the values of $\hat{R}^2_{\oplus}$ for the three fitted models (with and without the indicator variable for EE countries) using $d = d_W$, and demonstrates both the general agreement of the three models for this relatively homogeneous data set and the importance of the EE indicator in all models.

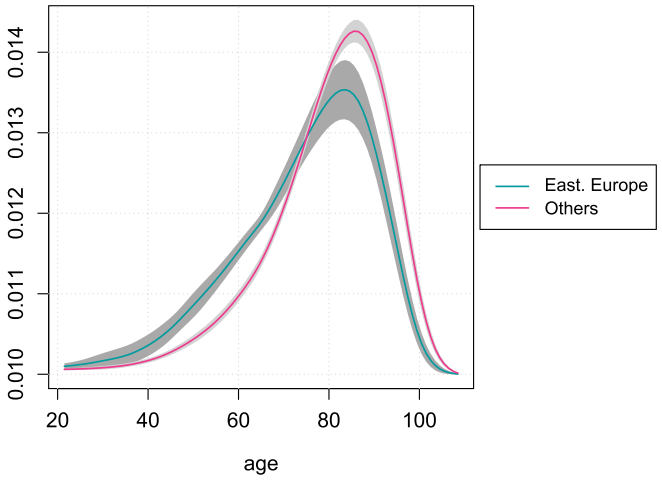


Fig. 7. Effect plot for Fréchet regression of age-at-death densities in Figure 2 using GDP, inflation, unemployment, and an indicator for Eastern European (EE) countries as predictors. Shown are the estimated conditional mean densities for EE and other countries under model **W**, with GDP, inflation, and unemployment set to their sample averages.

Table 2
Fréchet coefficients of determination under the Wasserstein metric

	Model		
	BS	**LQD**	**W**
Without EE	0.539	0.536	0.539
With EE	0.803	0.805	0.805

4. Analysis of Dependent Densities

In this section, we discuss densities that are either temporally or spatially dependent, i.e. they form either a time series or a spatial field of densities. We begin with density time series and then proceed to discuss fairly extensive work on spatial random fields of densities.

Density time series A time series model, $\{X_t, t \in \mathbb{Z}\}$, is a sequence of random objects exhibiting some form of dependence. In traditional time series analysis, the X_t are either scalars or vectors. Many excellent textbooks cover this setting, e.g., Brockwell and Davis (1991), Lütkepohl (2006) and Shumway and Stoffer (2018). Density time series form a subclass of functional time series. A functional time series consists of X_t in a function space, which is most commonly the Hilbert space $L^2(\mathcal{T})$ of square integrable functions on some domain $\mathcal{T}$, but X_t in a Banach space, for example the space of continuous functions on a compact interval, have also been studied. There is now considerable work on functional time series. Pioneering work on linear models models in this class is presented in Bosq (2000). More recent developments are presented in Horváth and Kokoszka (2012) and Kokoszka and Reimherr (2017). This is a growing field. We will not review the most recent work, but focus on the density time series.

An observed density time series is a finite sequence of densities, $f_1, f_2, \ldots, f_n$. A density time series model is a random sequence $\{\mathfrak{f}_t, t \in \mathbb{Z}\}$ with the dependence between the random densities $\mathfrak{f}_t$ quantified in some way. The most common approach of time series analysis is to transform the observations is such a way that the transformed data can be assumed to be invariant to time shifts in a certain way. For such stationary observations, a stochastic model can be proposed, which can be estimated by exploiting the fact that invariance to time shifts offers a form of replication of the stochastic structure. To put the development that follows in a suitable context, we present the following definition.

Definition 4.1. A sequence $\{X_t\}$ of elements of a separable Hilbert space is said to be stationary if the following conditions hold: (i) $\mathbb{E}[\|X_t\|^2] < \infty$, (ii) $\mathbb{E}[X_t]$ does not depend on t, and (iii) the autocovariance operators defined by $\mathbb{E}\big[\langle (X_t - \mu), x\rangle (X_{t+h} - \mu)\big]$ do not depend on t ($\mu = \mathbb{E}[X_0]$). The sequence $\{X_t\}$ is said to be strictly stationary if the distribution of $(X_{t+1}, X_{t+2}, \ldots X_{t+h})$ does not depend on t.

Stationary sequences are often called weakly stationary. If $\mathbb{E}[\|X_t\|^2] < \infty$, strict stationarity implies weak stationarity. Definition 4.1 could, in principle, be applied to square integrable densities, i.e. those $\mathbb{E}[\int \mathfrak{f}^2(u)du] < \infty$, but approaches based on the ideas explained in the iid context in previous sections are generally more useful. At the most fundamental level,

densities do not have to be square integrable. To define a stationary sequence of densities, we proceed as follows. Let $f_{\oplus,t}$ be the Fréchet mean of $\mathfrak{f}_t$ under the metric d_W as defined in (3.2), and $T_t = F_t^{-1} \circ F_{\oplus,t}$ be the optimal transport from $f_{\oplus,t}$ to the random density $\mathfrak{f}_t$. It is convenient to introduce the notation $\mathfrak{f}_t \ominus f_{\oplus,t} = \mathrm{Log}_{f_{\oplus,t}}(f_t) = T_t - \mathrm{id}$. The "difference" $f_t \ominus f_{\oplus,t}$ is analogous to $X_t - \mathbb{E}[X_t]$. The Wasserstein covariance kernel is defined by

$$\begin{aligned} \mathcal{C}_{t,t'}(u,v) &= \mathrm{Cov}((\mathfrak{f}_t \ominus f_{\oplus,t})(u),\ (\mathfrak{f}_{t'} \ominus f_{\oplus,t'})(v)) \\ &= \mathrm{Cov}(T_t(u) - u, T_{t'}(v) - v). \end{aligned} \tag{4.1}$$

Since $\int_{\mathbb{R}} \mathbb{E}[(\mathfrak{f}_t \ominus f_{\oplus,t}(u))^2] f_{\oplus,t}(u)du < \infty$, $\mathbb{E}[(\mathfrak{f}_t \ominus f_{\oplus,t}(u))^2] < \infty$ for almost all u in the support of $f_{\oplus,t}$. This means that the kernels $\mathcal{C}_{t,t'}(u,v)$ are defined for almost all $(u,v) \in \mathrm{supp}(f_{\oplus,t}) \times \mathrm{supp}(f_{\oplus,t'})$. We can now state the following definition.

Definition 4.2. A density time series $\{\mathfrak{f}_t, t \in \mathbb{Z}\}$ is said to be (second-order or weakly) stationary if (i) $f_{\oplus,t} = f_\oplus$ for all $t \in \mathbb{Z}$, (ii) $E\left[d_W^2(\mathfrak{f}_t, f_{\oplus,t})\right] < \infty$, (iii) For any $t, h \in \mathbb{Z}$, and almost all $u, v \in \mathrm{supp}(f_\oplus)$, $\mathcal{C}_{t,t+h}(u,v)$ does not depend on t.

As explained in Zhang et al. (2021), if $\{\mathfrak{f}_t, t \in \mathbb{Z}\}$ satisfies Definition 4.2, then $V_t = \mathrm{Log}_{f_\oplus}(\mathfrak{f}_t)$ is stationary according to Definition 4.1 with the separable Hilbert space in the latter being the tangent space $\mathcal{T}_{f_\oplus}$. Condition (ii) in Definition 4.2 is actually implied by condition (i) because the existence of the Wasserstein mean implies that the Wasserstein variance is finite. For this reason, strict stationarity in the sense of the following definition implies stationarity in the sense of Definition 4.2, provided the Wasserstein mean exists.

Definition 4.3. A density time series $\{\mathfrak{f}_t, t \in \mathbb{Z}\}$ is said to be strictly stationary if the joint distribution on $\mathcal{D}^h$ of $(\mathfrak{f}_{t+1}, \mathfrak{f}_{t+2}, \ldots, \mathfrak{f}_{t+h})$ does not depend on t.

A large part of time series analysis is devoted to constructing models for stationary time series, methods of their estimation, and prediction based on these models. In the realm of density time series, work on prediction methods is relatively well developed, as discussed below, but only limited work on model development exists. Zhang et al. (2021) developed a Wasserstein autoregressive (WAR) model, which we now outline in the simplest case of order 1 autoregression with a scalar coefficient. The starting point is a density $f_\oplus$, which serves as the Wasserstein mean of a stationary density time series model. The AR(1) equations are postulated to hold in the tangent space $\mathcal{T}_{f_\oplus}$ for a functional time series $\{V_t, t \in \mathbb{Z}\}$ that generates the random measures as $\mu_t = \mathrm{Exp}_{f_\oplus}(V_t)$. It is thus postulated that $V_t = \beta V_{t-1} + \epsilon_t$, where $\{\epsilon_t\}$ is an iid error sequence in $\mathcal{T}_{f_\oplus}$ and β is a real number. The existence and properties of AR(1) $\{V_t\}$ in $\mathcal{T}_{f_\oplus}$ follow from well–known arguments. In particular a stationary solution exists if $|\beta| < 1$. The challenge is to find conditions such that, almost surely, the measures $\mu_t = \mathrm{Exp}_{f_\oplus}(V_t)$ possess densities $\mathfrak{f}_t$ and satisfy $V_t = \mathrm{Log}_{f_\oplus}(\mu_t)$. Notice that the second condition is not automatic since the exponential and logarithmic maps are not precise inverses of one another. Estimation and prediction algorithms also require observed densities as input. Zhang et al. (2021) present sufficient conditions for these properties to hold. Most interestingly, unlike the scalar (or Hilbert space) AR models, conditions on the errors cannot be separated from the conditions on the autoregressive coefficients. For example, set $\epsilon_t(u) = \eta_t + \delta_t u$, where $\{\eta_t\}$ and $\{\delta_t\}$ are independent sequences of iid finite variance, zero mean scalar random variables. One can show that if $|\delta_t| < 1 - |\beta|$, then the errors ϵ_t are permissible. In general, sufficient conditions on the errors involve bounds on the derivatives $\epsilon_t'(u)$. Zhang et al. (2021) developed estimation and prediction algorithms for their WAR(p) models, including the selection of the order p. They also considered models with operator autoregressive coefficients for prediction, but these models performed worse than the models with scalar coefficients. A WAR(1) model for more general measures, rather than those admitting densities, was developed independently by Chen et al. (2021+).

In essence, the theory of Zhang et al. (2021) and Chen et al. (2021) uses the Log map to lift densities (or measures) to the tangent space and formulates a time series model in the tangent space. Seo (2017) and Seo and Beare (2019) use a mapping from the Bayes space as follows. Consider the time series of densities $\{\mathfrak{f}_t, t \in \mathbb{Z}\}$ as a stochastic sequence in the Bayes space, iid ϵ_t in this space and a sequence of bounded linear operators A_k mapping the Bayes space into itself. These objects can be mapped to $L^2(du)$ via the clr map introduced in Section 2.2. Denote their images, respectively, by $\tilde{f}_t$, $\tilde{\epsilon}_k$ and $\tilde{A}_k$. Then one can define a linear process in $L^2(du)$ via $\tilde{f}_t = \sum_{k\geq 1} \tilde{A}_k(\tilde{\epsilon}_{t-k})$. The densities are said to follow an AR(p) model if for operators Φ_j satisfying suitable conditions $\tilde{f}_t = \sum_{j=1}^p \tilde{\Phi}_j(\tilde{f}_{t-j}) + \tilde{\epsilon}_t$. This approach thus lifts densities to $L^2(du)$ via the clr map and formulates a model in $L^2(du)$. These papers focus on cointegration analysis. Another econometric application, unit root testing, is considered by Chang et al. (2016) who treat densities as elements of $L^2(du)$.

Practical methodology for prediction of density time series was developed before any fundamental theory had been available. Perhaps the first approach was put forward by Wang (2012). The idea is to fit the same parametric model to each density, so that $\mathfrak{f}_t(u) = f(u; \boldsymbol{\theta}_t)$, where $\boldsymbol{\theta}_t$ is a low dimensional parameter vector. For example, one can assume that $f(u; \boldsymbol{\theta}_t)$ is a skewed t density, in which case $\dim(\boldsymbol{\theta}_t) = 4$. The vector time series of estimated parameters, say $\hat{\boldsymbol{\theta}}_t$, can be predicted in a number of ways, for example using vector autoregression. This method is appealing and works reasonably well, but is not competitive against more recent approaches, at least on the data sets considered in Kokoszka et al. (2019) and Zhang et al. (2021), who compared many methods on several financial density time series. The following list can serve as a reference to the currently available prediction methods applicable to density time series.

ST *Skewed t Distribution.* This is the method of Wang (2012) described above. (The skewed t distribution is suitable for specific financial densities, and can be replaces by another parametric family.)

HC *Dynamic Functional Principal Component Regression.* This is the method of Horta and Ziegelmann (2018). Essentially it applies FPCA with a specific kernel, then forecasts the scores with a vector autoregressive (VAR) model. Predictions are produced by reconstructing densities with predicted scores. Negative predictions are replaced by zero and the reconstructed densities are standardized.

CoDa *Compositional Data Analysis.* This is a method proposed by Kokoszka et al. (2019) who removed the constraints on f_t by applying a centered log-ratio transformation. The forecast is produced by first applying FPCA to the output of these transformations, then fitting a time series model to the coefficient vectors.

LQD *Log Quantile Density Transformation.* This approach is based on the work of Petersen et al. (2016) and modified by Kokoszka et al. (2019). It transforms the density f_t to a Hilbert space where multiple FDA tools can be applied to forecast the transformed density, then applies the inverse transformation to get the forecast density back. Specifically, a modified log quantile density (LQD) transformation was applied to get the density forecasts.

WAR *Scalar coefficients WAR(p) model.* This method was proposed by Zhang et al. (2021). It constructs predictions in the tangent space via

$$\widehat{V}_{t+1}(u) = \sum_{j=1}^{p} \hat{\phi}_j V_{t+1-j}(u),$$

where the $\hat{\phi}_j$ are estimated scalar autoregressive coefficients. The V_{t+1-j} are obtained from the densities f_{t+1-j} via the Log map and the predicted $\hat{f}_{t+1}$ from $\widehat{V}_{t+1}$ via the Exp map.

FF-WAR *Fully functional WAR(p) model.* This is a more general variant considered by Zhang et al. (2021) and, in the case $p = 1$, indpendently by Chen et al. (2021+). It constructs predictions in the tangent space via

$$\widehat{V}_{t+1}(u) = \sum_{j=1}^{p} \int_{\mathbb{R}} \hat{\phi}_j(u, v) V_{t+1-j}(v) dv,$$

where the $\hat{\phi}_j(\cdot, \cdot)$ are estimated kernels of the autoregressive operator coefficients. The predicted $\hat{f}_{t+1}$ is constructed as in the WAR method.

Spatial fields of densities We begin with kriging, a term used to describe any method of predicting a value of an observation at a location for which relevant data are not available using data at locations at which they are available. Kriging, with its origins in mining, basically gave rise to the field of spatial statistics. Kriging of scalar and multivariate data is discussed in many textbooks, e.g., Schabenberger and Gotway (2005) and Wackernagel (2003). Over the last decade, kriging of functional data has received attention; Chapter 9 of Kokoszka and Reimherr (2017) provides an introduction and many references. Densities indexed by spatial locations are functional spatial data subject to nonlinear constraints.

Gouet et al. (2015) propose methodology for kriging in general metric spaces that are *uniquely geodesic*. This means that for any two points, there is a unique geodesic curve connecting them. Gouet et al. (2015) state basic definitions and refer to more extensive mathematical treatments. The Wasserstein space $(\mathcal{W}_2, d_W)$ is uniquely geodesic with the optimal transport map being the unique geodesic curve. In fact, $(\mathcal{W}_2, d_W)$ is an *Hadamard space*, which is a type of uniquely geodesic metric space required by the methodology of Gouet et al. (2015). We present their ideas in the special case of the space $(\mathcal{W}_2, d_W)$. Suppose $\{f(\boldsymbol{s}), \boldsymbol{s} \in \mathbb{R}^d\}$ is a field of densities viewed as random elements of $\mathcal{W}_2$. Define the variogram by

$$\gamma_W(\boldsymbol{s}, \boldsymbol{t}) = \mathbb{E}[d_W^2(\mathfrak{f}(\boldsymbol{s}), \mathfrak{f}(\boldsymbol{t}))] - d_W^2(f_{\oplus,\boldsymbol{s}}, f_{\oplus,\boldsymbol{t}}), \quad \boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^d.$$

For a scalar field $\{Z(\boldsymbol{s}), \boldsymbol{s} \in \mathbb{R}^d\}$, the variogram is defined by

$$\gamma(\boldsymbol{s}, \boldsymbol{t}) = \mathrm{Var}[Z(\boldsymbol{s}) - Z(\boldsymbol{t})] = \mathbb{E}[(Z(\boldsymbol{s}) - Z(\boldsymbol{t}))^2] - \{\mathbb{E}[Z(\boldsymbol{s})] - \mathbb{E}[Z(\boldsymbol{t})]\}^2.$$

In traditional spatial statistics, $\mathrm{Var}[Z(\boldsymbol{s}) - Z(\boldsymbol{t})]$ is a more natural expression because it indicates that the variogram captures the variability of the field. Since one cannot subtract densities, γ_W is defined in terms of expectations. Stationarity of a scalar random field is defined analogously to Definition 4.1 by requiring that $\mathbb{E}[Z(\boldsymbol{t})]$ and $\mathrm{Cov}(Z(\boldsymbol{t} + \boldsymbol{h}), Z(\boldsymbol{t}))$ do not depend on $\boldsymbol{t}$. In spatial statistics, one often works with *intrinsically stationary* fields, which are defined by the requirements that $\mathbb{E}[Z(\boldsymbol{t})]$ and $\mathrm{Var}[Z(\boldsymbol{t} + \boldsymbol{h}) - Z(\boldsymbol{t})]$ do not depend on $\boldsymbol{t}$. It is easy to verify that every stationary field is intrinsically stationary, and one can construct examples of fields that are intrinsically stationary, but not stationary. Since $\mathrm{Var}[Z(\boldsymbol{t} + \boldsymbol{h}) - Z(\boldsymbol{t})] = \gamma(\boldsymbol{t} + \boldsymbol{h}, \boldsymbol{t})$, the variogram is a natural tool for estimating the second order structure of intrinsically stationary fields, and actually it is also used to estimate the covariances $\mathrm{Cov}(Z(\boldsymbol{t} + \boldsymbol{h}), Z(\boldsymbol{t}))$ of a stationary field. Gouet et al. (2015) use the following definition.

Definition 4.4. The density random field $\{\mathfrak{f}(\boldsymbol{s}), \boldsymbol{s} \in \mathbb{R}^d\}$ is intrinsically stationary if the following conditions hold: (i) There is $f_\oplus$ such that for each $\boldsymbol{s}$, $f_{\oplus,\boldsymbol{s}}^{d_W} = f_\oplus$, (ii) $\gamma_W(\boldsymbol{s} + \boldsymbol{h}, \boldsymbol{t} + \boldsymbol{h})$ does not depend on $\boldsymbol{h}$.

In the scalar case, if $\gamma(\boldsymbol{s} + \boldsymbol{h}, \boldsymbol{t} + \boldsymbol{h})$ does not depend on $\boldsymbol{h}$, setting $\boldsymbol{s} = \boldsymbol{0}$, we see that $\gamma(\boldsymbol{h}, \boldsymbol{h} + \boldsymbol{t})$ does not depend on $\boldsymbol{h}$, so Definition 4.4 implies intrinsic stationarity in the usual sense. Using

$$\gamma(\boldsymbol{s}, \boldsymbol{t}) = \mathrm{Var}[Z(\boldsymbol{s})] + \mathrm{Var}[Z(\boldsymbol{t})] - 2\mathrm{Cov}(Z(\boldsymbol{s}), Z(\boldsymbol{t})),$$

it is easy to verify that stationarity implies intrinsic stationarity in the sense of Definition 4.4. The relationships between these three notions in the case of density random fields remain to be investigated.

Under Definition 4.4, Gouet et al. (2015) propose the following kriging approach. Suppose $f(\boldsymbol{s}_0)$ is to be predicted using $f(\boldsymbol{s}_k)$, $= 1, 2, \ldots, n$. The predictor $\hat{f}(\boldsymbol{s}_0)$ is defined as the projection of $f(\boldsymbol{s}_0)$ onto a convex span of the $f(\boldsymbol{s}_1), \ldots f(\boldsymbol{s}_n)$, subject to the constraint that the Wasserstein-Fréchet mean of $\hat{f}(\boldsymbol{s}_0)$ be equal to $f_\oplus$. Since $f(\boldsymbol{s}_0)$ is unknown, additional computation and estimation steps are needed. Among several approaches the authors consider, a reasonable strategy is to map the densities to the tangent space $\mathcal{T}_{f_\oplus}$ via the Log map, perform kriging in $\mathcal{T}_{f_\oplus}$, and recover the predictor via the Exp map. Since $\mathcal{T}_{f_\oplus}$ is a Hilbert space of functions, kriging can be performed without major difficulties, see e.g. Section 9.3 of Kokoszka and Reimherr (2017). This is the same strategy as employed by Zhang et al. (2021), but details need to be worked out as the Exp map mathematically results in a distribution, not a density. In practice, this does not appear to be a problem. Gouet et al. (2015) discuss several other approaches. Mathematical relationships between them remain to be explored and their feasibility and performance on realistic spacial data sets of densities remains to be evaluated, as the authors emphasize.

Motivated by kriging the distribution of particle (grain) sizes in a hydrology application, Menafoglio et al. (2014) consider a random field of densities, $f(\boldsymbol{s})$, where $\boldsymbol{s}$ is in a spatial domain. All densities are assumed to be defined on a common compact interval, which is suitable for the application. It is further assumed that the densities are in the Bayes space and form a stationary field in that space. Stationarity is defined according to the following definition.

Definition 4.5. The density random field $\{\mathfrak{f}(\boldsymbol{s}), \boldsymbol{s} \in \mathbb{R}^d\}$ is stationary in the Bayes space if the following conditions hold: (i) $f_{\oplus,\boldsymbol{s}}^{d_B} =: f_\oplus$ exists and does not depend on $\boldsymbol{s}$, (ii) $\mathbb{E}[\|\mathfrak{f}(\boldsymbol{s})\|_B^2] < \infty$ for each $\boldsymbol{s}$, (iii) $\mathbb{E}[\langle(\mathfrak{f}(\boldsymbol{s}+\boldsymbol{h}) \ominus f_\oplus), (\mathfrak{f}(\boldsymbol{s}) \ominus \mu_B)\rangle_B]$ does not depend on $\boldsymbol{s}$.

If the observed densities satisfy Definition 4.5, the kriging problem reduces to a kriging problem in the Bayes space, which is a Hilbert space. Specific formulas and estimation of quantities appearing in them are provided in Menafoglio et al. (2014), who also note that their approach can be implement in a non-stationary setting as well. A detailed application to 60 density curves constructed along a borehole within an aquifer is presented.

Menafoglio et al. (2016a) extend this approach to a setting where soil properties may vary with locations. The class to which soil at a given location belongs is used as an additional predictor. Using these approaches Menafoglio et al. (2016b) develop an algorithm for simulating particle size density curves. Menafoglio et al. (2018b) use the embedding of densities into Bayes space to develop an algorithm for monitoring density functions arising in an application to pore size distributions in alloys. Menafoglio et al. (2018a) propose methodology for dealing with spatial domains of a complex shape and local properties.

To summarize, we list the kriging methods discussed above.

GKP *Geodesic kriging predictor.* This is the set of methods outlined in Gouet et al. (2015) that use projections on convex spans of observed densities (or other functions).

FCK *Functional compositional kriging.* This is the method of Menafoglio et al. (2014). Kriging is performed in the Bayes space exploiting its Hilbert space structure. This method has several variants.

The FCK approach has been extended to more general data objects. Menafoglio and Secchi (2017) provide a review of object-oriented spatial statistics, while Menafoglio and Secchi (2019) discuss challenges and opportunities in this field.

5. Outlook and Future Directions

This review has focused on essentially two approaches to the analysis of density samples. In the first, one performs a preliminary (invertible) transformation on the densities to map them to a functional Hilbert space, after which any methodology from functional data analysis (FDA) can be applied. Given the vast and continually growing body of literature on methodological developments in FDA, it is clear that the illustrations given in this article merely scratch the surface of what the combined transformation and FDA approach for densities can offer the analyst. In particular, a large number of parametric and semi-parametric models with their theoretically justified tools for hypothesis testing and uncertainty assessment make the FDA approach an appealing one. On the other hand, the geometric approach allows one to break the bonds of linear analyses, often leading to more intuitive modeling and interpretable results. Once again, this review represents only the tip of the iceberg in this direction. For example, distance-based methods such as the distance covariance, (Székely et al., 2007; Lyons, 2013), the energy distance, (Székely and Rizzo, 2013), and a host of clustering algorithms all depend only on the concept of distance between data objects, rendering them suitable for the analysis of density samples. Distance-based statistics are, for the most part, inherently nonparametric, and thus provide a flexible alternative to the comparatively more structured models encountered in FDA.

It is possible that an area of great potential growth lies at the intersection of these two approaches for density analysis. Some of the methodologies in this article, such as the Fréchet regression models under the Bayes space or Wasserstein geometries in Section 3.3 or the Wasserstein autoregressive models in Section 4, attest to this, as they demonstrate how classical parametric modeling tools can be adapted for use with non-Hilbertian metrics. Future challenges in this line include the involvement of high-dimensional covariates and densities with complex and multidimensional supports.

Another set of open questions involves the effects of performing preliminary density estimation, both in an asymptotic and finite-sample sense. These details were omitted from this review so as to keep the focus on modeling and population targets, but they are a practical reality. Some of the methodologies presented herein have at least partially developed theoretical analyses that incorporate the errors resulting from density estimation, but these are mostly restricted to rates of

convergence or lower bounds on the sample sizes available from each density so as to make density estimation asymptotically negligible. A related and ongoing discussion in the field of FDA deals with the labeling of functional data samples as being "densely" or "sparsely" observed, or somewhere in between. The division between these regimes has recently been clarified in Zhang and Wang (2016), providing guidance for how one might approach this theoretical issue for density samples. Still, the case of density samples seems more difficult at the outset, as they are observed only through a sample, and not via discrete and noisy measurements made directly on the function as is the common assumption in FDA.

Looking beyond univariate densities, one often encounters multivariate distributions, so that the sample space $\mathcal{D}$ contains densities with support being a subset of $\mathbb{R}^d$, $d > 1$. Extension of the models and methods discussed in the review to this setting is nontrivial at best. For instance, the LQD transformation in (2.1) has no obvious generalization to the multivariate setting due to the absence of quantile functions. The Bayes space representation, on the other hand, provides a sound theoretical base for multivariate densities, although its practical implementation and utility has only been given limited, if any, consideration. The object-oriented approach using the Wasserstein metric has perhaps seen the most investigation for multivariate densities. The mathematical groundwork establishing the requisite geometric properties seems to be well in place, (Villani, 2003; Ambrosio et al., 2008; Panaretos and Zemel, 2020), with little work so for in the direction of statistical modeling for random multivariate densities. A notable exception is found in Zemel and Panaretos (2019), where an algorithm for computation of the sample Fréchet mean was developed in analogy to the Procrustes algorithm for shape spaces; the authors further showed asymptotic consistency of the sample mean, extending similar results derived in Panaretos and Zemel (2016) in the case $d = 1$. An apparent difficulty in this algorithm, and one that will continue to pose a stumbling block for future methodologies, is the lack of a closed-form expression for the logarithmic map when $d > 1$.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ecosta.2021.04.004

References

Agueh, M., Carlier, G., 2011. Barycenters in the Wasserstein space. SIAM Journal on Mathematical Analysis 43 (2), 904–924.
Aitchison, J., 1986. The Statistical Analysis of Compositional Data. Chapman & Hall, Ltd..
Amano, K., Sago, H., Uchikawa, C., Suzuki, T., Kotliarova, S.E., Nukina, N., Epstein, C.J., Yamakawa, K., 2004. Dosage-dependent over-expression of genes in the trisomic region of Ts1Cje mouse model for Down syndrome. Human Molecular Genetics 13 (13), 1333–1340.
Ambrosio, L., Gigli, N., Savaré, G., 2008. Gradient Flows in Metric Spaces and in the Spaces of Probability Measures. Springer Science & Business Media.
Bigot, J., Gouet, R., Klein, T., López, A., 2017. Geodesic PCA in the Wasserstein space by convex PCA. Annales de l'Institut Henri Poincaré B: Probability and Statistics 53, 1–26.
Bigot, J., Gouet, R., Klein, T., Lopez, A., et al., 2018. Upper and lower risk bounds for estimating the Wasserstein barycenter of random measures on the real line. Electronic Journal of Statistics 12 (2), 2253–2289.
Van den Boogaart, K.G., Egozcue, J.J., Pawlowsky-Glahn, V., 2014. Bayes Hilbert spaces. Australian & New Zealand Journal of Statistics 56 (2), 171–194.
Bosq, D., 2000. Linear Processes in Function Spaces. Springer.
Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods. Springer.
Castro, P.E., Lawton, W.H., Sylvestre, E.A., 1986. Principal modes of variation for processes with continuous sample curves. Technometrics 28, 329–337.
Cazelles, E., Seguy, V., Bigot, J., Cuturi, M., Papadakis, N., 2018. Geodesic PCA versus Log-PCA of histograms in the Wasserstein space. SIAM Journal on Scientific Computing 40 (2), B429–B456.
Chang, Y., Kim, C.S., Park, J.Y., 2016. Nonstationarity in time series of state densities. Journal of Econometrics 192, 152–167.
Chen, Y., Lin, Z., Müller, H.-G., 2021+. Wasserstein regression, arXiv: 2006.09660.
Chen, Z., Bao, Y., Li, H., Spencer Jr, B.F., 2019. LQD-RKHS-based distribution-to-distribution regression methodology for restoring the probability distributions of missing SHM data. Mechanical Systems and Signal Processing 121, 655–674.
Delicado, P., 2011. Dimensionality reduction when data are density functions. Computational Statistics and Data Analysis 55, 401–420.
Egozcue, J.J., Diaz-Barrero, J.L., Pawlowsky-Glahn, V., 2006. Hilbert space of probability density functions based on Aitchison geometry. Acta Mathematica Sinica 22, 1175–1182.
Faraway, J.J., 1997. Regression analysis for a functional response. Technometrics 39 (3), 254–261.
Ferraty, F., Vieu, P., 2006. Nonparametric Functional Data Analysis.. Springer, New York, New York.
Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S., 2004. Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Transactions on Medical Imaging 23 (8), 995–1005.
Gouet, R., Lopez, A., Ortiz, J.M., 2015. Geodesic kriging in the Wasserstein space. Proceedings of the 17th annual conference of the international association for mathematical geosciences. researchgate.net/publication/304157672
Han, K., Müller, H.-G., Park, B.U., 2020. Additive functional regression for densities as responses. Journal of the American Statistical Association 115 (530), 997–1010.
Hellinger, E., 1909. Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen.. Journal für die reine und angewandte Mathematik (Crelles Journal) 1909 (136), 210–271.
Horta, E., Ziegelmann, F., 2018. Dynamics of financial returns densities: A functional approach applied to the Bovespa intraday index. International Journal of Forecasting 34, 75–88.
Horváth, L., Kokoszka, P., 2012. Inference for Functional Data with Applications. Springer.
Hron, K., Menafoglio, A., Templ, M., Hruzova, K., Filzmoser, P., 2016. Simplicial principal component analysis for density functions in Bayes spaces. MOX-report 25, 2014.
Hsing, T., Eubank, R., 2015. Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators. John Wiley & Sons.
Jones, M.C., Rice, J.A., 1992. Displaying the important features of large collections of similar curves. The American Statistician 46, 140–145.
Kneip, A., Utikal, K.J., 2001. Inference for density families using functional principal component analysis. Journal of the American Statistical Association 96 (454), 519–542.
Kokoszka, P., Miao, H., Petersen, A., Shang, H.L., 2019. Forecasting of density functions with an application to cross-sectional and intraday returns. International Journal of Forecasting 35 (4), 1304–1317.
Kokoszka, P., Reimherr, M., 2017. Introduction to Functional Data Analysis. CRC Press.
Laha, R.G., Rohatgi, V.K., 1979. Probability Theory. John Wiley & Sons.

Lütkepohl, H., 2006. New Introduction to Multiple Time Series Analysis. Springer.
Lyons, R., 2013. Distance covariance in metric spaces. The Annals of Probability 41 (5), 3284–3305.
Marron, J.S., Alonso, A.M., 2014. Overview of object oriented data analysis. Biometrical Journal 56 (5), 732–753.
Marron, J.S., Ramsay, J.O., Sangalli, L.M., Srivastava, A., 2015. Functional data analysis of amplitude and phase variation. Statistical Science 468–484.
Menafoglio, A., Gaetani, G., Secchi, P., 2018. Random domain decompositions for object-oriented kriging over complex domains. Stochastic Environmental Research and Risk Assessment 32, 3421–3437.
Menafoglio, A., Grasso, M., Secchi, P., Colosimo, B., 2018. Profile monitoring of probability density functions via simplicial functional PCA with application to image data. Technometrics 60, 497–510.
Menafoglio, A., Guadagnini, A., Secchi, P., 2014. A kriging approach based on Aitchison geometry for the characterization of particle-size curves in heterogeneous aquifers. Stochastic environmental research and risk assessment 28, 1835–1851.
Menafoglio, A., Secchi, P., 2017. Statistical analysis of complex and spatially dependent data: A review of object oriented spatial statistics. European Journal of Operational Research 258, 401–410.
Menafoglio, A., Secchi, P., 2019. O2S2: A new venue for computational geostatistics. Applied Computing and Geosciences 2. doi:10.1016/j.acags.2019.100007.
Menafoglio, A., Secchi, P., Guadagnini, A., 2016a. A class-kriging predictor for functional compositions with application to particle-size curves in heterogeneous aquifers. Mathematical Geosciences 48, 463–485.
Menafoglio, A., Secchi, P., Guadagnini, A., 2016b. Stochastic simulation of soil particle-size curves in heterogeneous aquifer systems through a Bayes space approach. Water Resources Research 52, 5708–5726.
Muandet, K., Fukumizu, K., Sriperumbudur, B., Schölkopf, B., 2016. Kernel mean embedding of distributions: A review and beyond. arXiv: 1605.09522.
Nerini, D., Ghattas, B., 2007. Classifying densities using functional regression trees: Applications in oceanology. Computational Statistics and Data Analysis 51, 4984–4993.
Panaretos, V.M., Zemel, Y., 2016. Amplitude and phase variation of point processes. The Annals of Statistics 44 (2), 771–812.
Panaretos, V.M., Zemel, Y., 2019. Statistical aspects of Wasserstein distances. Annual review of statistics and its application 6, 405–431.
Panaretos, V.M., Zemel, Y., 2020. An Invitation to Statistics in Wasserstein Space. Springer Nature.
Patrangenaru, V., Ellingson, L., 2015. Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis. CRC Press.
Petersen, A., Chen, C.-J., Müller, H.-G., 2019. Quantifying and visualizing intraregional connectivity in resting-state functional magnetic resonance imaging with correlation densities. Brain connectivity 9 (1), 37–47.
Petersen, A., Liu, X., Divani, A.A., 2021. Wasserstein F-tests and confidence bands for the Fréchet regression of density response curves. The Annals of Statistics 49 (1), 590–611.
Petersen, A., Müller, H.-G., 2019. Wasserstein covariance for multiple random densities. Biometrika 106 (2), 339–351.
Petersen, A., Müller, H.-G., et al., 2016. Functional data analysis for density functions by transformation to a Hilbert space. The Annals of Statistics 44 (1), 183–218.
Petersen, A., Müller, H.-G., et al., 2019. Fréchet regression for random objects with Euclidean predictors. The Annals of Statistics 47 (2), 691–719.
Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis. Springer Series in Statistics, second Springer, New York.
Salazar, P., Di Napoli, M., Jafari, M., Jafarli, A., Ziai, W., Petersen, A., Mayer, S.A., Bershad, E.M., Damani, R., Divani, A.A., 2020. Exploration of multiparameter hematoma 3D image analysis for predicting outcome after intracerebral hemorrhage. Neurocritical care 32 (2), 539–549.
Satterthwaite, F.E., 1941. Synthesis of variance. Psychometrika 6 (5), 309–316.
Schabenberger, O., Gotway, C.A., 2005. Statistical Methods for Spatial Data Analysis. CRC Press.
Seo, W.-K., 2017. Cointegrated density-valued linear processes. arXiv: 1710.07792.
Seo, W.-K., Beare, B.K., 2019. Cointegrated linear processes in Bayes Hilbert space. Statistics and Probability Letters 147, 90–95.
Shumway, R., Stoffer, S., 2018. Time Series Analysis and Its Applications. Springer.
Srivastava, A., Jermyn, I., Joshi, S., 2007. Riemannian analysis of probability density functions with applications in vision. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 1–8.
Srivastava, A., Wu, W., Kurtek, S., Klassen, E., Marron, J. S., 2011. Registration of functional data using Fisher-Rao metric. arXiv: 1103.3817.
Székely, G.J., Rizzo, M.L., 2013. Energy statistics: A class of statistics based on distances. Journal of statistical planning and inference 143 (8), 1249–1272.
Székely, G.J., Rizzo, M.L., Bakirov, N.K., et al., 2007. Measuring and testing dependence by correlation of distances. The Annals of Statistics 35 (6), 2769–2794.
Talská, R., Menafoglio, A., Hron, K., Egozcue, J.J., Palarea-Albaladejo, J., 2020. Weighting the domain of probability densities in functional data analysis. Stat 9 (1), e283.
Talská, R., Menafoglio, A., Machalová, J., Hron, K., Fišerová, E., 2018. Compositional regression with functional response. Computational Statistics & Data Analysis 123, 66–85.
Tucker, J.D., Wu, W., Srivastava, A., 2013. Generative models for functional data using phase and amplitude separation. Computational Statistics & Data Analysis 61, 50–66.
Villani, C., 2003. Topics in Optimal Transportation, Number: 58. American Mathematical Soc..
Wackernagel, H., 2003. Multivariate Geostatistics. Springer.
Wang, J., 2012. A state space model approach to functional time series and time series driven by differential equations. Rutgers University-Graduate School-New Brunswick Ph.D. thesis.
Wang, J.-L., Chiou, J.-M., Müller, H.-G., 2016. Functional data analysis. Annual Review of Statistics and its Application 3, 257–295.
Zemel, Y., Panaretos, V.M., 2019. Fréchet means and procrustes analysis in wasserstein space. Bernoulli 25 (2), 932–976.
Zhang, C., Kokoszka, P., Petersen, A., 2021. Wasserstein autoregressive models for density time series, Journal of Time Series Analysis, doi:10.111/jtsa.12590.
Zhang, X., Wang, J.-L., 2016. From sparse to dense functional data and beyond. The Annals of Statistics 44 (5), 2281–2321.
Zhang, Z., Müller, H.-G., 2011. Functional density synchronization. Computational Statistics and Data Analysis 55, 2234–2249.