Simplicial principal component analysis for density functions in Bayes spaces

K. Hron [a,b,*], A. Menafoglio [c], M. Templ [d,e], K. Hrůzová [a,b], P. Filzmoser [d]

[a] *Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacký University, 17. listopadu 12, 771 46 Olomouc, Czech Republic*

[b] *Department of Geoinformatics, Faculty of Science, Palacký University, 17. listopadu 50, 77146 Olomouc, Czech Republic*

[c] *MOX - Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy*

[d] *Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Wiedner Hauptstraße 8-10, 1040 Vienna, Austria*

[e] *Department of Methodology, Statistics Austria, Guglgasse 13, 1110 Vienna, Austria*

ARTICLE INFO

ABSTRACT

Probability density functions are frequently used to characterize the distributional properties of large-scale database systems. As functional compositions, densities primarily carry relative information. As such, standard methods of functional data analysis (FDA) are not appropriate for their statistical processing. The specific features of density functions are accounted for in Bayes spaces, which result from the generalization to the infinite dimensional setting of the Aitchison geometry for compositional data. The aim is to build up a concise methodology for functional principal component analysis of densities. A simplicial functional principal component analysis (SFPCA) is proposed, based on the geometry of the Bayes space $\mathcal{B}^2$ of functional compositions. SFPCA is performed by exploiting the centred log-ratio transform, an isometric isomorphism between $\mathcal{B}^2$ and L^2 which enables one to resort to standard FDA tools. The advantages of the proposed approach with respect to existing techniques are demonstrated using simulated data and a real-world example of population pyramids in Upper Austria.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Nowadays, an increasing number of studies are based on complex data, such as curves, surfaces or images. As a direct consequence, the importance of Functional Data Analysis (FDA, e.g., Ramsay and Silverman, 2002, and references therein) has recently strongly increased. In recent years, a large body of literature has been developed in this field (e.g., Ramsay and Silverman, 2005; Horváth and Kokoszka, 2012, and references therein), however, still little attention has been paid to the problem of dealing with functional data that are probability density functions (Delicado, 2007, 2011; Nerini and Ghattas, 2007; Zhang and Müller, 2011; Menafoglio et al., 2014a,b). Even though it might seem that density functions are just a special case of functional data – with a constant-integral-constraint equal to one – standard FDA methods appear to be inappropriate for their treatment, as they do not consider the particular constrained nature of the data. This problem is well known in

* Corresponding author at: Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacký University, 17. listopadu 12, 771 46 Olomouc, Czech Republic. Tel.: +420 585 634 605.

E-mail addresses: hronk@seznam.cz (K. Hron), alessandra.menafoglio@polimi.it (A. Menafoglio), templ@statistik.tuwien.ac.at (M. Templ), klara.hruzova@gmail.com (K. Hrůzová), p.filzmoser@tuwien.ac.at (P. Filzmoser).

the finite dimensional setting, where specific techniques have been worked out to deal with compositional data (e.g., Aitchison, 1986; Pawlowsky-Glahn and Egozcue, 2001; Egozcue and Pawlowsky-Glahn, 2006; Egozcue, 2009; Pawlowsky-Glahn and Buccianti, 2011, and references therein), i.e., multivariate data carrying only relative information, usually represented in proportions or percentages. Those techniques are mainly based on a geometric perspective grounded on the Aitchison geometry in the simplex, which properly incorporates the compositional nature of the data. In this context, probability density functions have recently been interpreted as functional compositional data, i.e., functional data carrying only relative information. To handle this kind of data, the Aitchison geometry has been recently extended to the so called Bayes spaces: a Hilbert space structure for σ-finite measures, including probability measures, has been worked out in Van den Boogaart et al. (2014), based on the pioneering work of Egozcue et al. (2006) and the subsequent developments of Van den Boogaart et al. (2010) and Egozcue et al. (2013). The name Bayes spaces comes from the primary purpose of the approach, which is to assign a simple algebraic interpretation for the basic notions of mathematical statistics (e.g., the Bayes theorem as a paradigm of information acquisition; Van den Boogaart et al., 2010). The idea of Bayes spaces was first exploited in Delicado (2011) for the statistical analysis of density functions in the context of dimensionality reduction through multidimensional scaling. Very recently, the Hilbert space structure of probability density functions with a compact support has been used by Menafoglio et al. (2014a,b) to elaborate a kriging methodology for probability density functions.

The aim of this work is to take a step forward in the direction of functional principal component analysis in Bayes spaces, moving from the work of Delicado (2011). In particular, we shall geometrically solve the problem of functional principal component analysis (FPCA) in the Bayes Hilbert space of probability density functions. The sample space of compositional data – i.e., the infinite dimensional simplex – motivates the name *simplicial* functional principal component analysis (SFPCA). Furthermore, we will propose the use of the centred log-ratio transform (clr, Van den Boogaart et al., 2014; Menafoglio et al., 2014a) for its practical implementation. We remark that, unlike the non-linear transformations which are commonly used for probability density functions (e.g., the logarithmic transformation), the centred log-ratio transformation is an isometric isomorphism between the Bayes space of probability density functions and the space L^2 of square-integrable real measurable functions. From an application viewpoint this is extremely important, as it allows to solve the problem of FPCA with the usual L^2 geometry, while accounting for the specific features of Bayes spaces.

The remaining part of the paper is organized as follows. Section 2 introduces the Bayes space of probability density functions as well as the clr-transformation which will be used for their processing. FPCA is recalled in Section 3 for L^2 data. The extension of FPCA to Bayes spaces is proposed in Section 4, and is further explored in Section 5 through a simulation study. In Section 6, we apply the developed methodology to a real case study dealing with age distributions in Upper Austria. Section 7 eventually concludes the work.

2. Density functions as elements of a Bayes space

The theory of Bayes spaces (Egozcue et al., 2006; Van den Boogaart et al., 2010, 2014; Egozcue et al., 2013) has been introduced as a generalization to density functions of the Aitchison geometry. This is commonly used for compositional data, i.e., multivariate observations carrying only relative information (e.g., Aitchison, 1986; Pawlowsky-Glahn and Buccianti, 2011, and references therein), which are usually collected in the form of constrained data summing up to a constant, usually set to 1 or 100, in case of proportions or percentages, respectively. Any probability density function f can be considered as a compositional vector with infinitely many parts (Egozcue et al., 2006): as such, it inherits the key features of compositions (e.g., Egozcue, 2009).

We denote by μ an absolutely continuous measure with respect to the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, with compact support $I \subset \mathbb{R}$ and density f. To keep the notation simple, hereafter we refer to the properties of μ through the properties of f, even though Van den Boogaart et al. (2010); Egozcue et al. (2013); Van den Boogaart et al. (2014) develop the theory of Bayes spaces in a complete generality. We say that two density functions f, g are equivalent if they are proportional, and we denote this by $f =_{\mathcal{B}} g$. As such, the integral constraint $\int_I f(x)\, dx = 1$ of a probability density function singles out a representative within an equivalence class of functional compositions that provide the same set of information. Indeed, any other representative $\tilde{f}$ within the same class (and characterized by a constraint $\int_I \tilde{f}(x)\, dx = c$ for $c > 0$) carries the same relative information on the contribution of any Borel subset of $\mathbb{R}$ to the measure of the support. In this setting – as noted by Egozcue et al. (2013) – the probability of a given event has not a meaning *per se*, but should be compared with the probability of the entire sample space, which is conventionally set to 1, but could be equivalently set to another positive constant c. This property is known as *scale invariance*.

A second important feature of functional compositions is the *relative scale* property: the relative increase of a probability over a Borel set from 0.05 to 0.1 (2 multiple) differs from the increase 0.5–0.55 (1.1 multiple), although the absolute differences are the same in both cases. This property reflects the relative nature of functional compositions, and further motivates the use of the log-ratio approach – already extensively employed in compositional data analysis – to deal with density functions.

In fact, both the scale invariance and the relative scale properties are completely ignored when considering probability density functions just like unconstrained functional data. In particular, the usual notions of sum and product by a constant appear inappropriate when applied to compositions, since the space of functional compositions endowed with those operations is not a vector space (e.g., the point-wise sum of two compositions is not necessarily a composition). Instead, the geometry of the Bayes Hilbert space of Van den Boogaart et al. (2014), which is described below, enables one to capture

and properly incorporate these properties. In the following, we restrict our attention to density functions with compact support, as in Delicado (2011). Both theoretical and practical reasons motivate this choice. Indeed, when the support is the whole real line, the Lebesgue measure cannot be used as reference probability measure, leading to highly technical issues. Moreover, in most real datasets, finite values for the inferior and superior extremes of the support can be determined without a substantial loss of generality.

We call $\mathcal{B}^2(I)$ the Bayes space of (equivalence classes of) positive functional compositions f on I with square-integrable logarithm. In particular, we here consider continuous (hence bounded) functional compositions on the compact support I. Hereafter, the representative of an equivalence class will be its element integrating to 1. Moreover, the symbol I will denote an interval $[a, b]$, but any subset of $\mathbb{R}$ with finite measure could be dealt with analogously. Given two absolutely integrable density functions $f, g \in \mathcal{B}^2(I)$ and a real number $\alpha \in \mathbb{R}$ we indicate with $f \oplus g$ and $\alpha \odot f$ the perturbation and powering operations, respectively, defined as (Egozcue et al., 2006; Van den Boogaart et al., 2014):

$$(f \oplus g)(t) = \frac{f(t)g(t)}{\int_I f(s)g(s)\,ds}, \qquad (\alpha \odot f)(t) = \frac{f(t)^\alpha}{\int_I f(s)^\alpha\,ds}, \quad t \in I.$$

The resulting functions are readily seen to be probability density functions. Egozcue et al. (2006) prove that $\mathcal{B}^2(I)$ endowed with the operations $(\oplus, \odot)$ is a vector space. Note that the neutral elements of perturbation and powering are $e(t) = 1/\eta$, with $\eta = b - a$ (i.e., the uniform density), and 1, respectively. Moreover, the difference between two elements $f, g \in \mathcal{B}^2(I)$, denoted by $f \ominus g$, is obtained as perturbation of f with the reciprocal of g, i.e., $(f \ominus g)(t) = (f \oplus [(-1) \odot g])(t)$, $t \in I$.

The following example aims at illustrating the effect of the perturbation and powering operations in $\mathcal{B}^2(I)$, as opposed to standard operations of sum and product by a constant in $L^2(I)$. We consider the restriction to $I = [-5, 5]$ of the Gaussian densities $f =_{\mathcal{B}} \exp\{-t^2/2\}$ and $g =_{\mathcal{B}} \exp\{-(t-m)^2/(2s^2)\}$, with $m = 1$ and $s^2 = 2$. Hereafter in this example – particularly in Fig. 1 – we will always consider as representatives the elements which honour the unit integral constraint over the compact support $I = [-5, 5]$. The left panel of Fig. 1 represents the perturbation of f by g ($f \oplus g$), as well as the sum in L^2 of f and g ($f + g$). We first notice that the L^2 sum of f and g does not result in a probability density function, and thus would not be appropriate as an operation in $\mathcal{B}^2(I)$. Instead, the perturbation of f by g results into a density function. The latter can be explicitly obtained as the Gaussian density

$$\begin{aligned}
f \oplus g &=_{\mathcal{B}} \exp\{-t^2/2\} \cdot \exp\{-(t-m)^2/(2s^2)\} \\
&= \exp\{-(t-m_1)^2/(2s_1^2)\} \cdot \exp\{-m^2/(2s^2+2)\} =_{\mathcal{B}} \exp\{-(t-m_1)^2/(2s_1^2)\},
\end{aligned}$$

with $m_1 = 1/3$ and $s_1^2 = 2/3$. We note that the Gaussian family is closed with respect to the perturbation operation (Van den Boogaart et al., 2010). This is not surprising if one recalls that the perturbation operation $\oplus$ can be interpreted as a Bayesian updating of information – and $\ominus$ as a cancellation of information – (Egozcue et al., 2013), and that the Gaussian family is conjugate with itself. Notice that $f \oplus g$ is more concentrated than f and shifted towards g: this is the consequence of adding to f the information content in g (the opposite argument applies as well). The right panel of Fig. 1 displays the result of the powering operation $\alpha \odot f$ in $\mathcal{B}^2(I)$ as opposed to the multiplication $\alpha \cdot f$ in $L^2(I)$, for the same f considered before and the scalar $\alpha = 2$. We first notice that $\alpha \cdot f$ is not a density function, and, as an element of $\mathcal{B}^2(I)$, it belongs to the same equivalence class of f itself. In fact, the multiplication operation applied to elements in $\mathcal{B}^2(I)$ simply changes the representative within the equivalence class: as such, it is not a sensible operation in this setting. Instead, the powering of f by $\alpha = 2$ has the effect of increasing the concentration of f around its mean (i.e., it decreases the variance of f by a factor 2). In the Bayesian framework, this is interpreted as the increase of information which is obtained by incrementing the "evidence" in f by the "evidence" in f itself.

To endow $\mathcal{B}^2(I)$ with a Hilbert space structure, Egozcue et al. (2006) define the inner product

$$\langle f, g \rangle_{\mathcal{B}} = \frac{1}{2\eta} \int_I \int_I \ln\frac{f(t)}{f(s)} \ln\frac{g(t)}{g(s)}\,dt\,ds, \quad f, g \in \mathcal{B}^2(I), \tag{1}$$

with $\eta = b - a$, which induces the following norm,

$$\|f\|_{\mathcal{B}} = \left[\frac{1}{2\eta} \int_I \int_I \ln^2\frac{f(t)}{f(s)}\,dt\,ds\right]^{1/2}.$$

The space $\mathcal{B}^2(I)$, endowed with the inner product (1), is proved to be a separable Hilbert space in Egozcue et al. (2006). As such, it is isomorphic to the Hilbert space $L^2(I)$ of (equivalence classes of) square-integrable real functions on I. An isometric isomorphism between $\mathcal{B}^2(I)$ and $L^2(I)$ is defined by the *centred log-ratio* (clr) transformation (Van den Boogaart et al., 2014; Menafoglio et al., 2014a), which is defined, for $f \in \mathcal{B}^2(I)$, as

$$\mathrm{clr}(f)(t) = f_c(t) = \ln f(t) - \frac{1}{\eta} \int_I \ln f(s)\,ds. \tag{2}$$

We remark that such an isometry allows to compute operations and inner products among the elements in $\mathcal{B}^2(I)$ in terms of their counterpart in $L^2(I)$ among the clr-transforms, i.e.

$$\mathrm{clr}(f \oplus g)(t) = f_c(t) + g_c(t), \qquad \mathrm{clr}(\alpha \odot f)(t) = \alpha \cdot f_c(t), \qquad \langle f, g \rangle_{\mathcal{B}} = \langle f_c, g_c \rangle_2 = \int_I f_c(t)g_c(t)\,dt.$$
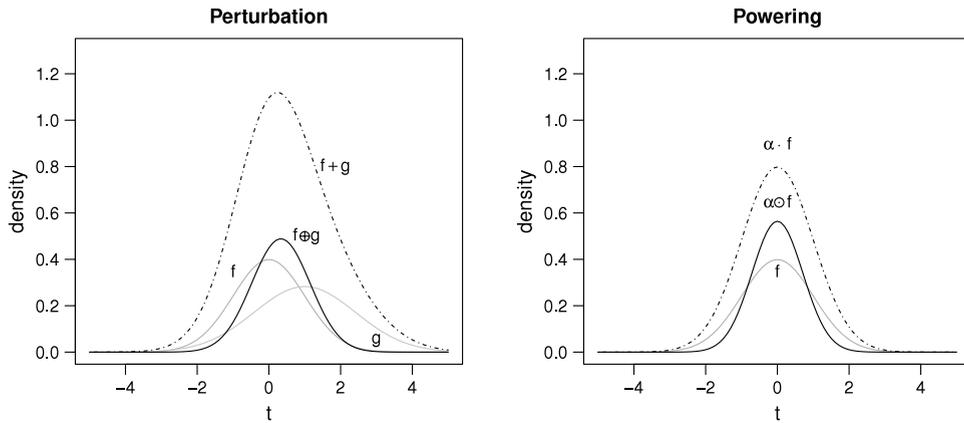
Fig. 1. Example of perturbation and powering in $\mathcal{B}^2(I)$, compared to the standard operations in $L^2(I)$. Left: Perturbation $f \oplus g$ (solid black line) of two Gaussian densities f, g restricted to $I = [-5, 5]$ (grey lines), and the sum $f + g$ in the space $L^2(I)$ (dot-dashed line). Right: Powering of a Gaussian density f restricted to $I = [-5, 5]$ (grey line) by $\alpha = 2$, $\alpha \odot f$ (solid black line), and the counterpart $\alpha \cdot f$ in L^2 (dot-dashed line).

However, we notice that, by construction, one has the constraint

$$\int_I \text{clr}(f)(t)dt = \int_I \ln f(t)\, dt - \int_I \frac{1}{\eta} \int_I \ln f(s)\, ds\, dt = 0.$$

This additional condition needs to be taken into account for computation and analysis on clr-transformed density functions, as we shall show in Section 4.

3. Principal component analysis for functional data

Principal component analysis (PCA) is a widely used multivariate statistical technique aiming to capture the main modes of variability of the data by means of a small number of linear combinations of the original variables. In the functional context, the same aim is reached by Functional Principal Component Analysis (FPCA). Here, we briefly recall FPCA, referring the reader, e.g., to Ramsay and Silverman (2002, Chapter 8), Horváth and Kokoszka (2012, Chapter 3) and Shang (2014), for further details on this topic.

Let us consider a functional random sample $X_1, \ldots, X_N$ in $L^2(I)$, and indicate with $\langle x, y \rangle_2 = \int_I x(t)y(t)dt$ the inner product between two elements x, y in $L^2(I)$ and with $\|x\|_2 = (\int_I |x(t)|^2 dt)^{1/2}$ the induced norm. For ease of notation and without loss of generality, we assume the samples to be centred. FPCA firstly looks for the main mode of variability, i.e., for the element ξ_1 in $L^2(I)$ – called first functional principal component (FPC) – maximizing over $\xi \in L^2(I)$

$$\frac{1}{N} \sum_{i=1}^N \langle X_i, \xi \rangle_2^2 \text{ subject to } \|\xi\|_2 = 1. \tag{3}$$

The remaining FPCs, $\{\xi_j\}_{j\geq 2}$, capture the remaining modes of variability subject to be mutually orthogonal, and are thus obtained by solving problem (3) with the additional orthogonality constraint $\langle \xi_k, \xi_j \rangle_2 = 0, k < j$.

Analogously to the multivariate case, the FPCs $\{\xi_j\}_{j\geq 1}$ coincide with the eigenfunctions of the sample covariance operator $V : L^2(I) \rightarrow L^2(I)$ (e.g., Horváth and Kokoszka, 2012, p. 26), acting on $x \in L^2(I)$ as

$$Vx = \frac{1}{N} \sum_{i=1}^N \langle X_i, x \rangle_2 X_i,$$

or, equivalently, as

$$Vx = \int_I v(\cdot, t)x(t)dt,$$

the kernel $v : I \times I \rightarrow \mathbb{R}$ being the sample covariance function

$$v(s, t) = \frac{1}{N} \sum_{i=1}^N x_i(s)x_i(t), \quad s, t \in I.$$

Therefore, the jth FPC ξ_j and the associated scores $\Psi_{ij} = \langle X_i, \xi_j \rangle_2$, $i = 1, \ldots, N$, are obtained by solving the eigenvalue equation

$$V\xi_j = \rho_j \xi_j, \tag{4}$$

where ρ_j denotes the jth eigenvalue, with $\rho_1 \geq \rho_2 \geq \cdots$. As in multivariate PCA, for each j, the term $\rho_j / \sum_j \rho_j$ is associated with the proportion of total variability explained by the FPC ξ_j.

Several computational methods can be utilized to solve Eq. (4) (e.g., Ramsay and Silverman, 2005; Jones and Rice, 1992; Kneip and Utikal, 2001, and references therein). Ramsay and Silverman (2005, Chapter 8.4) suggest to express each datum X_i, $i = 1, \ldots, N$, as a linear combination of K known basis functions $\phi_1, \ldots, \phi_K$ and to solve the eigenproblem (4) through an appropriate matrix coefficient. Indeed, suppose that each datum X_i, $i = 1, \ldots, N$, admits the basis expansion

$$X_i(\cdot) = \sum_{k=1}^{K} c_{ik} \phi_k(\cdot), \tag{5}$$

or, in matrix notation, $\mathbf{X}(\cdot) = \mathbf{C}\boldsymbol{\phi}(\cdot)$, with $\mathbf{C} = (c_{ik}) \in \mathbb{R}^{N,K}$, $\mathbf{X}(\cdot) = (X_i(\cdot))$, and $\boldsymbol{\phi}(\cdot) = (\phi_i(\cdot))$. Then the variance–covariance function takes the form $v(s, t) = N^{-1}\boldsymbol{\phi}(s)'\mathbf{C}'\mathbf{C}\boldsymbol{\phi}(t)$, $s, t \in I$. Suppose further that the eigenfunction ξ_j, $j \geq 1$, admits the expansion $\xi_j(\cdot) = \sum_{k=1}^{K} b_{jk}\phi_k(\cdot)$, or in matrix notation $\xi_j(\cdot) = \boldsymbol{\phi}(\cdot)'\mathbf{b}_j$. This yields $V\xi_j(\cdot) = \boldsymbol{\phi}(\cdot)'N^{-1}\mathbf{C}'\mathbf{CMb}_j$, where $\mathbf{M}_{kl} = \langle \phi_k, \phi_l \rangle_2$. Therefore the eigenvalue equation (4) reduces to

$$N^{-1}\mathbf{C}'\mathbf{CMb}_j = \rho_i \mathbf{b}_j, \tag{6}$$

and $\mathbf{b}_j$ is obtained as solution of the linear system (6). Note that in case of basis orthonormality $\mathbf{M} = \mathbf{I}$ the FPCA problem reduces to standard multivariate PCA of the coefficient matrix $\mathbf{C}$. Otherwise, Ramsay and Silverman (2005, Chapter 8.4) show that problem (6) is equivalent to the eigenproblem

$$\frac{1}{N}\mathbf{M}^{1/2}\mathbf{C}'\mathbf{CM}^{1/2}\mathbf{u}_j = \rho_i \mathbf{u}_j$$

with $\mathbf{u}_j = \mathbf{M}^{1/2}\mathbf{b}_j$, i.e., FPCA reduces to a multivariate PCA of the transformed coefficient matrix $\mathbf{CM}^{1/2}$ followed by the transformation $\mathbf{b} = \mathbf{M}^{-1/2}\mathbf{u}$.

4. Simplicial functional principal component analysis

As functional compositions, probability density functions are featured by specific properties, such as the scale invariance and relative scale properties. The latter would be neglected if one applied the functional principal component analysis described in Section 3 to density functions. Aim of this section is to derive a simplicial version of FPCA, named SFPCA, by following the same scheme that led to the formulation of FPCs in Section 3, but in agreement with the Bayes Hilbert space geometry introduced in Section 2.

Let $\widetilde{X}_1, \ldots, \widetilde{X}_N$ be a sample in $\mathcal{B}^2(I)$, and denote by $X_1, \ldots, X_N$ the corresponding centred sample, i.e., for $i = 1, \ldots, N$, $X_i = \widetilde{X}_i \ominus \overline{X}$, where $\overline{X}$ stands for the sample mean $\overline{X} = \frac{1}{N} \odot \bigoplus_{i=1}^{N} \widetilde{X}_i$. We consider the problem of finding the simplicial functional principal components (SFPCs) in $\mathcal{B}^2(I)$, i.e., the elements $\{\zeta_j\}_{j \geq 1}$, $\zeta_j \in \mathcal{B}^2(I)$, maximizing the following objective function over $\zeta \in \mathcal{B}^2(I)$:

$$\frac{1}{N}\sum_{i=1}^{N}\langle X_i, \zeta \rangle_{\mathcal{B}}^2 \text{ subject to } \|\zeta\|_{\mathcal{B}} = 1; \qquad \langle \zeta, \zeta_k \rangle_{\mathcal{B}} = 0, \quad k < j, \tag{7}$$

where the orthogonality condition $\langle \zeta, \zeta_k \rangle_{\mathcal{B}} = 0$, for $k < j$, holds only for $j \geq 2$.

$\mathcal{B}^2(I)$ being a separable Hilbert space, problem (7) is well posed (Horváth and Kokoszka, 2012, Theorem 3.2, p. 38). Thus, the solution of problem (7) exists and is unique. Indeed, analogously to the $L^2(I)$ case previously discussed, the jth SFPC solves the eigenvalue equation

$$V\zeta_j = \lambda_j \odot \zeta_j, \tag{8}$$

(λ_j, ζ_j) being the jth eigenpairs of the sample covariance operator $V : \mathcal{B}^2(I) \to \mathcal{B}^2(I)$, acting on $x \in \mathcal{B}^2(I)$ as

$$Vx = \frac{1}{N} \odot \bigoplus_{i=1}^{N} \langle X_i, x \rangle_B \odot X_i.$$

In order to proceed with (7) in practice, we apply the isometric isomorphism between $\mathcal{B}^2(I)$ and $L^2(I)$ defined by the clr-transform (2) that allows to rewrite problem (7) as a maximization of the term

$$\frac{1}{N}\sum_{i=1}^{N}\langle \text{clr}(X_i), \text{clr}(\zeta) \rangle_2^2 \text{ subject to } \|\text{clr}(\zeta)\|_2 = 1; \qquad \langle \text{clr}(\zeta), \text{clr}(\zeta_k) \rangle_2 = 0, \quad k < j$$

over $\zeta \in \mathcal{B}^2(I)$. On this basis, for $j \geq 1$ the maximization problem (7) can be equivalently restated as finding $\nu \in L^2$ which maximizes

$$\frac{1}{N}\sum_{i=1}^{N}\langle \text{clr}(X_i), \nu \rangle_2^2 \text{ subject to } \|\nu\|_2 = 1; \qquad \langle \nu, \nu_k \rangle_2 = 0, \quad k < j; \qquad \int_I \nu = 0, \tag{9}$$

where the orthogonality constraint is meaningful only for $j \geq 2$ and the zero-integral constraint incorporates the corresponding clr-transform property.

We now show that (9) is solved by the eigenfunctions $\{\xi_j\}_{j \geq 1}$ of the sample covariance operator $V_{\text{clr}} : L^2(I) \to L^2(I)$ of the transformed sample $\text{clr}(X_1), \ldots, \text{clr}(X_N)$, acting on $x \in L^2(I)$ as

$$V_{\text{clr}} x = \frac{1}{N} \sum_{i=1}^{N} \langle \text{clr}(X_i), x \rangle_2 \, \text{clr}(X_i).$$

We first notice that the eigenfunctions $\{\xi_j\}_{j \geq 1}$ would have solved problem (9), if it had been stated without the zero-integral condition $\int_I \nu = 0$, since in that case (9) would have been equivalent to (3) (with the orthogonality constraints for $j \geq 2$). Therefore, to prove that $\nu = \xi_j$ maximizes (9) it suffices to show that ξ_j fulfils also the constraint $\int_I \xi_j = 0$, for all $j \geq 1$. To this end, we note that the zero-integral property of the clr-transformed sample $\text{clr}(X_1), \ldots, \text{clr}(X_N)$ implies that V_{clr} admits a zero eigenvalue with associated eigenfunction $\xi_0 \equiv 1/\sqrt{b-a}$:

$$V_{\text{clr}} \xi_0 = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sqrt{b-a}} \left[\int_I \text{clr}(X_i) \right] \text{clr}(X_i) \equiv 0.$$

Since the eigenfunctions $\{\xi_j\}$ corresponding to the remaining non-null eigenvalues $\{\rho_j\}$ are to be orthogonal to the eigenfunction ξ_0, the ξ_j's need to satisfy the zero-integral condition $\int_I \xi_j = 0$, as $\langle \xi_j, \xi_0 \rangle_2 = 1/\sqrt{b-a} \int_I \xi_j$. Another way to see this is to notice that (a) the image of the sample covariance operator V_{clr} is the span of the clr-transformed observations, (and the constant function $\xi_0 \equiv \frac{1}{\sqrt{b-a}}$ belongs to its kernel), and (b) the eigenfunctions corresponding to the non-null eigenvalues form a basis of the image of V_{clr}. As such, each eigenfunction ξ_j can be written as a unique linear combination of the functions $\text{clr}(X_1), \ldots, \text{clr}(X_n)$. Therefore, the zero-integral condition is fulfilled since it holds, by construction, for each of the functions $\text{clr}(X_i)$, $i = 1, \ldots, N$. Thus, problem (7) can be restated in terms of clr-transforms as (9) and the SFPCs can be obtained by transforming the eigenfunctions $\{\xi_j\}_{j \geq 1}$ associated to the non-null eigenvalues $\{\rho_j\}_{j \geq 1}$ of V_{clr} through the inverse of the function clr, namely $\zeta_j = \text{clr}^{-1}(\xi_j) =_{\mathcal{B}} \exp(\xi_j)$, with $j \geq 1$. Note that, as in classical PCA, the eigenfunctions ξ_j are determined up to sign changes. Accordingly, the SFPCs are determined up to a powering by ± 1 (i.e., if ζ_j solves problem (7), $-1 \odot \zeta_j$ is a solution as well).

To compute the eigenfunctions ξ_j we resort to a method based on a B-spline basis expansion. Following Machalová et al. (2015), we consider for $\text{clr}(X_1), \ldots, \text{clr}(X_N)$ and $\xi_j, j \geq 1$, a B-spline basis fulfilling the zero-integral constraint,

$$\text{clr}(X_i)(\cdot) = \sum_{k=1}^{K} c_{ik} \phi_k(\cdot), \qquad \xi_j(\cdot) = \sum_{k=1}^{K} b_{jk} \phi_k(\cdot). \tag{10}$$

To compute the B-splines coefficients the usual parametrization of smoothing splines applies, and the additional constraint is incorporated in the estimation algorithm as described in Machalová et al. (2015). We refer to the Appendix for further details. Hence, with the same arguments used in Section 3, $\mathbf{b}_j = (b_{jk})$ is obtained as solution of the eigenproblem

$$N^{-1} \mathbf{C}' \mathbf{CMb}_j = \rho_i \mathbf{b}_j,$$

with analogue orthogonality arguments as those previously introduced, the zero integral constraint is inherently kept in the PCA algorithm, and thus does not need to be explicitly imposed.

Alternative computational approaches could be employed as well to perform SFPCA. For instance, one may employ an expansion based on a (truncated) polynomial basis, such as Lagrange, Legendre or Hermite bases. In these settings, the zero-integral constraint should be still accounted for, e.g., by leaving the constant function of the polynomial basis out of the fitting procedure (e.g., Van den Boogaart et al., 2014). Other approaches to data preprocessing in Bayes spaces have been proposed in the literature, e.g., based on Bernstein polynomials (e.g., Menafoglio et al., 2014a) or kernel density estimation (e.g., Delicado, 2011). However, the latter preprocessing procedures do not allow to express eigenproblem (8) in terms of a multivariate eigenproblem, and would thus require to find the eigen-decomposition of V_{clr} numerically.

Finally, note that an equivalence exists between metric multidimensional scaling and principal component analysis, if these are based on the same distance (Johnson and Wichern, 2002, Chapter 12.6). On this basis, one could perform dimensionality reduction by employing MDS based on the Aitchison distance (Delicado, 2011), and then working with the coordinates of a reduced space—which play the same role as the scores in SFPCA. The viewpoint of MDS is however different, as it looks for a low dimensional representation of the data preserving the distance among the original observations. In this sense, the focus of MDS is on the distance between observations, as opposed to SFPCA whose focus is on the observations themselves. In fact, the availability of a comprehensive framework for SFPCA allows to employ a rich set of techniques to explore its results. For instance, for the purpose of dimensionality reduction, the choice of the number of SFPCs to be retained can follow the same strategies as those used in FPCA: one may fix a threshold in the amount of variability explained by the retained SFPCs, or look for an elbow in the scree plot. Even the interpretation of the results of SFPCA may follow the main lines used in the $L^2(I)$ case, since the SFPCs represent the main modes of variability of the observations around the global mean function, but in the space $\mathcal{B}^2(I)$ endowed with its own geometry. Useful tools to visualize and interpret the results of SFPCs are then the scores plan graph and the representation of the mean function perturbed by the jth SFPC ξ_j

(a) Density functions.

(b) Clr-transforms.

(c) Covariance function in L^2.

(d) Covariance function in $\mathcal{B}^2$.

Fig. 2. Gaussian densities on $I = [-5, 5]$ with $\mu = 0$ and $\sigma_i = \exp(-1 + (i - 1)/10)$ for $i = 1, \ldots, 21$.

powered by an appropriate constant. In the framework of Bayes spaces, the latter is interpreted as the distribution obtained when updating/cancelling the information carried by the mean density with the evidence within the *j*th mode of variability (possibly increased by a proper factor to amplify its effect). In the following we will make an extensive use of this kind of representation, which in fact corresponds to the graphs of the mean $\pm$ the FPCs multiplied by a constant, advocated by Ramsay and Silverman (2005, Chapter 8.3) in the $L^2(I)$ case.

5. Simulation study

In this section, we explore the main features of the Bayes Hilbert space geometry and the subsequent SFPCA through a simulation study based on density functions within the exponential family. Emphasis will be given to the effect of utilizing a compositional approach to the analysis of densities, with particular regard to the impact of the relative scale and scale invariance properties.

5.1. The effect of clr-transformation

We here examine the effect of processing density functions using clr-transformation (2) in case of two common distribution families, censored according to fixed compact support I. Similarly as the artificial data considered by Delicado (2011) and the densities introduced in Section 2, we first use a set of Gaussian densities, with mean $\mu = 0$ and standard deviation $\sigma_i = \exp(-1 + (i - 1)/10)$ for $i = 1, \ldots, 21$ (Fig. 2(a)) on the interval $I = [-5, 5]$, i.e.,

$$f(t; \sigma_i) =_{\mathcal{B}} \exp\left\{-\frac{t^2}{2\sigma_i^2}\right\}, \quad t \in I. \tag{11}$$

Recall that, for each $i = 1, \ldots, 21$, the kernel in (11) uniquely determines the $\mathcal{B}^2(I)$-equivalence class to which $f(t; \sigma_i)$ belongs, without need of specifying the normalizing constant (i.e., scale invariance). This reflects on clr-transformation (2), which is invariant to the normalizing constant. The clr-transforms of densities (11) are obtained as

$$f_c(t; \sigma_i) = -\frac{t^2}{2\sigma_i^2} + \frac{25}{6\sigma_i^2}, \quad t \in I. \tag{12}$$

These are displayed, on grey scale, in Fig. 2(b). Due to the zero-integral constraint, all the clr-transforms $f_c(t; \sigma_i)$, $i = 1, \ldots, 21$ have to cross the x-axis. In the present case, all the $f_c(t; \sigma_i)$ equal zero at $t = \pm\sqrt{25/3}$, as obtained by solving, for $i = 1, \ldots, 21, f_c(t; \sigma_i) = 0$ and using (12). Hence, at $t = \pm\sqrt{25/3}$ the variability of the densities, in the Aitchison sense, is zero. This is also reflected by the covariance function in $\mathcal{B}^2(I)$, which is reported in Fig. 2(d). Note that the existence of the

(a) Density functions.

(b) Clr-transforms.

(c) Covariance function in $L^2(I)$.

(d) Covariance function in $\mathcal{B}^2(I)$.

(e) A zoom into the top-left corner of the covariance function in $\mathcal{B}^2(I)$.
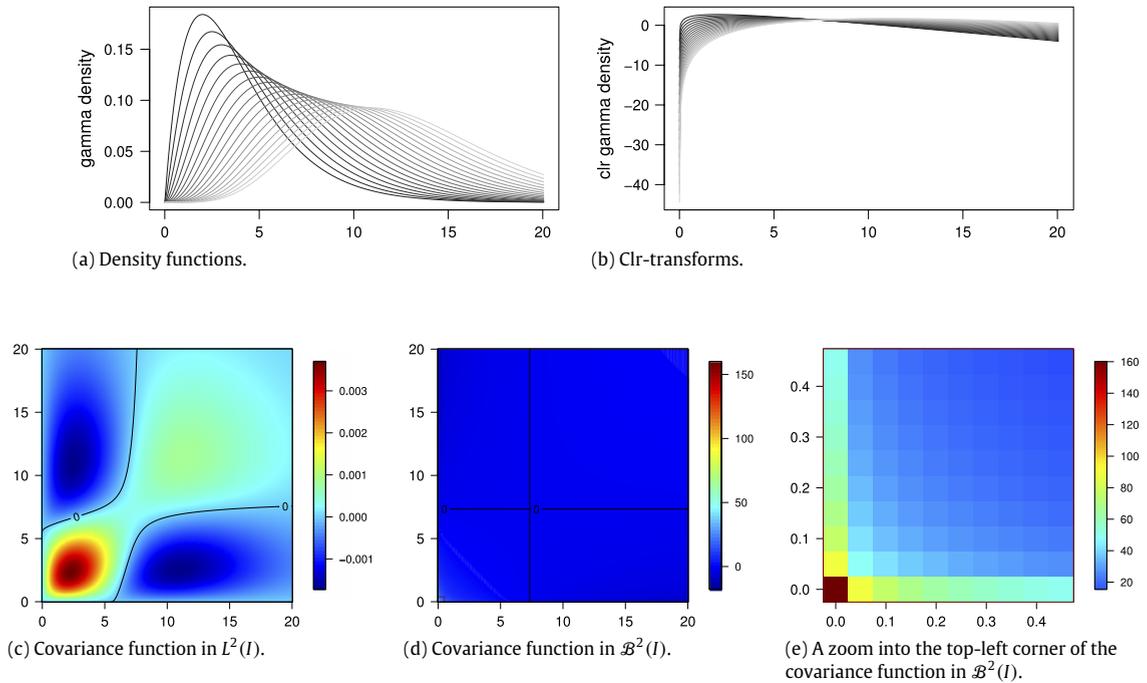
**Fig. 3.** Gamma densities on $I = [e^{-7}, e^3]$ with scale parameter $\theta = 2$, and shape parameters $\kappa = 2 + (i - 1)/4$ for $i = 1, \ldots, 19$.

points of zero $\mathcal{B}^2(I)$-variability depends on the functional form of the considered densities and may not appear in general. For instance, points of zero $\mathcal{B}^2(I)$-variability would not be obtained in case of Gaussian densities affected by variability in both the mean and the standard deviation.

The fact that $\mathcal{B}^2(I)$ honours the relative scale property of densities is apparent when observing Fig. 2(a) and (b). Indeed, when looking at the densities in the original space, major differences seem to be exhibited among the concave parts of the curves. Instead, on a relative scale, the variability of the densities is more significant within the tails of the distribution. The latter are key to characterize the densities, though having nearly zero absolute contributions. This is well reflected by the clr-transforms in Fig. 2(b), which exhibit high variability near the boundary of the support (Fig. 2(d)). In this regard, we remark that the clr-transforms are elements of $L^2(I)$, as a result of an isometric transformation from the Bayes space $\mathcal{B}^2(I)$ to $L^2(I)$. Hence, their variability can be evaluated in the clr-space (Fig. 2(b)) via the usual L^2 optics (e.g., a high variability is present near the boundaries of the domain).

To see the effect of clr-transformation on right-skewed densities, we employ a set of gamma densities with scale parameter $\theta = 2$ and shape parameter $\kappa_i = 2 + (i - 1)/4$ for $i = 0, \ldots, 19$ and $I = [e^{-7}, e^3]$, i.e.

$$f(t; \kappa_i) =_{\mathcal{B}} t^{\kappa_i - 1} \exp \left\{ -\frac{t}{2} \right\}, \quad t \in I.$$

These are displayed, together with their clr-transforms, in Fig. 3(a)–(b). As before, the relative scale property drives the variability when this is measured in the sense of $\mathcal{B}^2(I)$ (Fig. 3(d)–(e)). Indeed, the contribution to the variability provided by the portion of the support associated to small probabilities (i.e., the left tails) is more pronounced than that associated with the peaks of the densities. This is also apparent when comparing the covariance functions estimated in $L^2(I)$ and in $\mathcal{B}^2(I)$ (Fig. 3(c) and (d)–(e), respectively). Accordingly, one can expect that the effect of the relative scale property will be highly influential to SFPCA, whose aim is precisely to capture the relative variability as primary source of information in density functions. In particular, unlike FPCA, the first principal direction of variability in $\mathcal{B}^2(I)$ is likely to be "attracted" by the variability within the regions where densities assume small values, because these regions have a substantial impact when evaluating log-ratios between contributions of subdomains (i.e., the basis upon which the Bayes Hilbert geometry is built). To appreciate these differences, in the next subsection, we shall study the effect of analysing a dataset of densities though a relative scale (as in $\mathcal{B}^2(I)$), as opposed to the absolute scale employed when considering the L^2 geometry.

5.2. SFPCA for density functions in the exponential family

In this subsection we perform SFPCA of sets of densities and compare the results with those of FPCA. Furthermore, we shall give an interpretation to the clr-transform in terms of the transformation with general validity as advocated by Delicado (2011) to perform the principal component analysis of compactly supported density functions.

To this end, we focus on the SFPCA of two sets of density functions belonging to the extended exponential family. We recall that a *k-parametric extended exponential family* on I, $\text{Exp}_{\mathcal{B}(I)}(g, \boldsymbol{T}, \boldsymbol{\vartheta})$ (Van den Boogaart et al., 2010) is a collection of densities

$$f(t, \boldsymbol{\alpha}) =_{\mathcal{B}} g(t) \cdot \exp\left\{\sum_{j=1}^{k} \vartheta_j(\boldsymbol{\alpha}) T_j(t)\right\}, \quad t \in I,$$

where $\boldsymbol{\alpha}$ denotes the k-dimensional vector of parameters in a k-dimensional parameter space A, while $g : I \to \mathbb{R}$, $\vartheta_j : A \to \mathbb{R}$ and $T_j : I \to \mathbb{R}$, $j = 1, \ldots, k$, are Borel-measurable functions. Van den Boogaart et al. (2010) prove that an extended exponential family on I is a finite dimensional affine subspace of the Bayes space $\mathcal{B}^2(I)$. In this regard, we note that a density in $\text{Exp}_{\mathcal{B}(I)}(g, \boldsymbol{T}, \boldsymbol{\vartheta})$ can be expressed as a linear combination in $\mathcal{B}^2(I)$:

$$f(t, \boldsymbol{\alpha}) =_{\mathcal{B}} g(t) \oplus \bigoplus_{j=1}^{k} \left[\vartheta_j(\boldsymbol{\alpha}) \odot \exp\{T_j(t)\}\right], \quad t \in I,$$

and the corresponding clr-transform is obtained as:

$$f_c(t, \boldsymbol{\alpha}) = g_c(t) + \sum_{j=1}^{k} \left[\vartheta_j(\boldsymbol{\alpha}) \cdot \text{clr}(\exp\{T_j(t)\})\right], \quad t \in I. \tag{13}$$

In particular, if k is minimal for $\text{Exp}_{\mathcal{B}(I)}(g, \boldsymbol{T}, \boldsymbol{\vartheta})$ (e.g., Witting, 1985), then $\text{Exp}_{\mathcal{B}(I)}(g, \boldsymbol{T}, \boldsymbol{\vartheta})$ is called *strictly k-parametric* and the dimension of the corresponding affine space is precisely k. We note that, by virtue of (13), for a strictly k-parametric exponential family with $k_0 \leq k$ uncertain parameters, the SFPCA estimates an orthonormal basis of the corresponding k-dimensional affine space in $\mathcal{B}^2(I)$, which is associated to $k_0 \leq k$ non-zero eigenvalues.

For the purpose of the simulation study, we here consider the following sets of densities:

- Set 1: $n = 21$ densities with Gaussian kernel $N(0, \sigma_i)$, with $\sigma_i = \exp(-1 + (i - 1)/10)$ for $i = 1, \ldots, 21$, and domain $I = [-5, 5]$;
- Set 2: $n = 100$ densities with kernel Gamma $\Gamma(\theta_i, \kappa_j)$, with $\theta_i = 1/9 + (i-1)/9$ and $\kappa_j = 2 + (j-1)/4$ for $i, j = 1, \ldots, 10$, and domain $I = [e^{-7}, e^3]$.

Note that a Gaussian density $N(0, \sigma^2)$ restricted on I belongs to a 1-parametric extended exponential family, with $\alpha = \sigma$, $\vartheta_1(\alpha) = 1/\sigma^2$, and $T_j(t) = -t^2$. Instead, a Gamma distribution $\Gamma(\theta, \kappa)$ on I belongs to a 2-parametric extended exponential family with $\boldsymbol{\alpha} = (\theta, \kappa)$, $\vartheta_1(\boldsymbol{\alpha}) = \theta$, $\vartheta_2(\boldsymbol{\alpha}) = \kappa$, $T_1(t) = -t$, and $T_2(t) = \ln(t)$, for $t \in I$. As such, we expect that a sensible dimensionality reduction method will single out the dimension $k = 1$ and $k = 2$ of the densities in Set 1 and in Set 2.

We first focus on the SFPCA of Set 1, which is the same set of densities analysed in Section 5.1. A similar set of densities is considered in Delicado (2011). This author applies the transformation $\Psi_N(f)(t) \equiv \frac{\partial}{\partial t} \log(f(t))$ – firstly proposed by Ramsay and Silverman (2002) – to linearize the problem and consequently apply standard FPCA to the transformed data. Nevertheless, Delicado (2011) remarks that the use of the transformation $\Psi_N(f)$ is only justified in the presence of (nearly) Gaussian densities, with consequent strong limitations in its applicability. Accordingly, he proposes the use of multidimensional scaling (MDS) as a general dimensionality reduction method for density functions, with particular regard to the MDS based on the compositional distance.

Fig. 4 reports the results of SFPCA on Set 1. The first SFPC – displayed in Fig. 4(c) – captures the entire variability of the dataset and is precisely interpreted in terms of mass concentration. Indeed, the positive scores along the first SFPC are associated with the highest standard deviations of Set 1 and vice versa (Fig. 4(b); here the indices $i = 1, \ldots, 21$ refer to the standard deviation σ_i of the corresponding density). Such an interpretation can be easily derived from the plot of the mean density perturbed by $\oplus/\ominus$ the first SFPC weighted according to the corresponding standard deviation – i.e., $\sqrt{\lambda_1}, \lambda_1$ appearing in (8) – that is depicted in Fig. 4(e). These results are in agreement with those of Delicado (2011), obtained via the transformation $\Psi_N(f)$. Nevertheless, the SFPCA (a) allows to ease the interpretation of the results through the use of standard graphical displays and (b) is based on a transformation with general validity, namely the clr-transform.

As noted, SFPCA correctly identifies the dimensionality of the affine space of $\mathcal{B}(I)$ corresponding to the considered extended exponential family. This is empirically confirmed by the results on Set 1, and further corroborated by the SFPCA on Set 2 (Fig. 5). In this case – which could not have been analysed via the transformation Ψ_N – only two non-zero SFPCs are obtained, which together capture the entire variability of the dataset. As expected, both the SFPCs tend to be driven by the left tail of the distribution (Fig. 5(c)–(d)), due to the relative scale property of densities, which is accounted for in the Bayes space $\mathcal{B}^2(I)$.

In fact, very different results would have been obtained via standard FPCA in $L^2(I)$, as shown in Fig. 6. Here, the dimensionality reduction suggested by the scree-plot (Fig. 6(a)) does not respect the dimensionality of the problem, and the non-linear relationship between the first two FPCs is apparent. Note that the first two FPCs seek to characterize areas with high absolute value of the densities (i.e., their peaks, see Fig. 6(e)–(f) and (h)), and provide poorer approximations in the regions associated with small probability. This is a typical effect of analysing densities on an absolute scale, and marks a key difference between SFPCA and FPCA. Further, the approximation of the densities via the first two FPCs is inappropriate,

(a) Explained variance.

(b) Scores along SFPC$_1$ and SFPC$_2$.

(c) SFPC$_1$ (100% of variability).

(d) SFPC$_2$ (0% of variability).

(e) Mean $\pm\sqrt{\lambda_1}$ SFPC$_1$.

(f) Mean $\pm\sqrt{\lambda_2}$ SFPC$_2$.

(g) clr(Mean) $\pm\sqrt{\lambda_1}$ clr(SFPC$_1$).

(h) clr(Mean) $\pm\sqrt{\lambda_2}$ clr(SFPC$_2$).

(i) Original densities.

(j) Approximated densities (via SFPC$_1$).
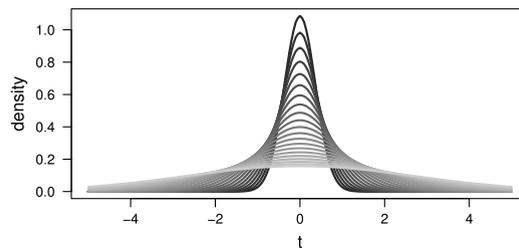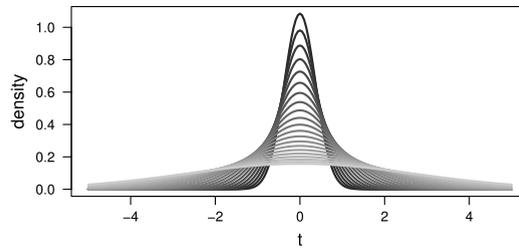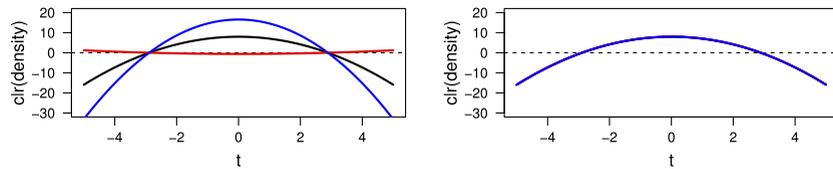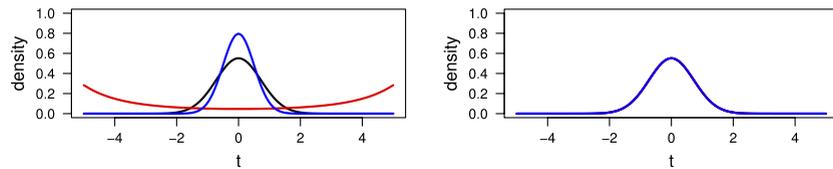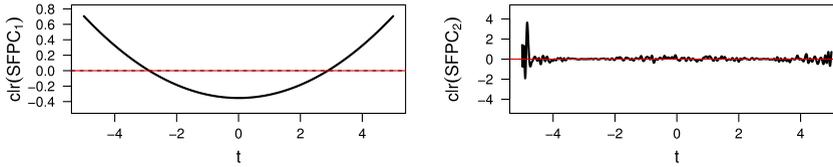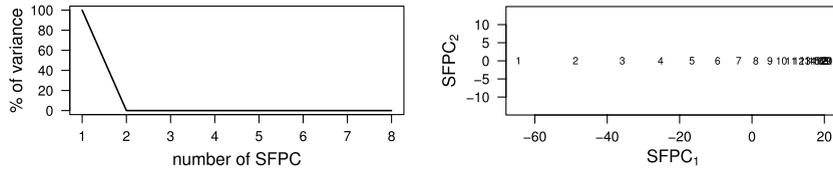
Fig. 4. SFPCA of Gaussian densities with $\mu = 0$ and $\sigma_i = \exp(-1 + (i-1)/10)$ for $i = 1, \dots, 21$. In panels (e) and (f): the solid black curve indicates the mean function, the red curve indicates the mean $\ominus$ the SFPC, the blue curve indicates the mean $\oplus$ the SFPC. In panels (g) and (h): the solid black curve indicates the clr-mean function, the red curve indicates the clr-mean minus the SFPC, the blue curve indicates the clr-mean plus the SFPC. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(a) Explained variance.

(b) Scores along $SFPC_1$ and $SFPC_2$.

(c) $SFPC_1$ (93.4% of variability).

(d) $SFPC_2$ (6.6% of variability).

(e) Mean $\pm\sqrt{\lambda_1}$ $SFPC_1$.

(f) Mean $\pm\sqrt{\lambda_2}$ $SFPC_2$.

(g) clr(Mean) $\pm\sqrt{\lambda_1}$ clr($SFPC_1$).

(h) clr(Mean) $\pm\sqrt{\lambda_2}$ clr($SFPC_2$).

(i) Original densities.

(j) Approximated densities (via $SFPC_1$ and $SFPC_2$).

Fig. 5. SFPCA of Gamma densities with $\theta = 1/9 + (i-1)/9$ and $k = 2 + (j-1)/4$ for $i, j = 1, \ldots, 10$. In panels (e) and (f): the solid black curve indicates the mean function, the red curve indicates the mean $\ominus$ the SFPC, the blue curve indicates the mean $\oplus$ the SFPC. In panels (g) and (h): the solid black curve indicates the clr-mean function, the red curve indicates the clr-mean minus the SFPC, the blue curve indicates the clr-mean plus the SFPC. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(a) Explained variance.

(b) Scores along FPC$_1$ and FPC$_2$.

(c) FPC$_1$ (62.1% of variability).

(d) FPC$_2$ (28.0% of variability).

(e) Mean $\pm\sqrt{\rho_1}$ FPC$_1$.

(f) Mean $\pm\sqrt{\rho_2}$ FPC$_2$.

(g) Original densities.

(h) Approximated densities (via FPC$_1$ and FPC$_2$).

Fig. 6. FPCA of Gamma densities with $\theta = 1/9 + (i-1)/9$ and $k = 2 + (j-1)/4$ for $i, j = 1, \ldots, 10$. In panels (e) and (f): the solid black curve indicates the mean function, the red curve indicates the mean minus the FPC, the blue curve indicates the mean plus the SFPC. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

as it exhibits negative values (Fig. 6(h)). Instead, not only the densities approximation via SFPCs is guaranteed to produce proper density functions, but exactly reproduces the original densities when the number of retained components respects the dimensionality of the problem (Fig. 5(g)).

Fig. 7. Upper Austria and its districts.

We finally note that the presented simulation study focuses on the case of densities from the exponential family of distributions. These densities are widely used in applications and have been thoroughly analysed from the perspective of Bayes spaces (e.g., Van den Boogaart et al., 2010). Nevertheless, a more comprehensive picture of the effects of employing the Bayes space methodology for analysing densities could be obtained through an extension of the present simulation study to a larger family of densities. In this sense, the analysis of the real dataset presented in the next section can be considered as an illustration of SFPCA in a more general setting.

6. Analysis of population age distributions in the Upper Austria regions

In this section we perform the SFPCA of a real dataset dealing with the population age distributions in Upper Austria. Out of the nine states constituting Austria, this is the fourth-largest Austrian state in terms of land area and third-largest by population. Upper Austria is formed by 15 political districts, displayed in Fig. 7. The dataset we consider collects the age distributions of men and women living in $N = 57$ municipalities of Upper Austria. In the literature this kind of data is often referred to as *population pyramid*; a similar dataset – but referring to different countries in the world for the year 2000 – has been considered in Delicado (2011) in the context of dimensionality reduction with particular emphasis on graphical displays. The aim of the current study is to characterize the available population age densities, and perform a dimensionality reduction according to the geometry of the Bayes space $\mathcal{B}^2(I)$ as opposed to the usual $L^2(I)$ geometry. In particular, we here focus on the impact and the advantages of incorporating the basic properties of functional compositions when performing the statistical analysis of any dataset of densities, not necessarily in the exponential family.

For the purpose of the present study, the possible spatial dependence among the observations will not be considered. Instead, the geographical information will be considered for the interpretation of the scores. We refer to Menafoglio et al. (2013, 2014a,b) for a geostatistical approach to account for the spatial dependence in the presence of Hilbert data, and particularly functional compositions in $\mathcal{B}^2(I)$.

The raw data have been smoothed by using the procedure proposed in Machalová et al. (2015), and recalled in the Appendix: the discrete clr-transforms of raw densities (Egozcue and Pawlowsky-Glahn, 2006) have been projected on a B-spline basis with support $I = [0, 100]$ and five equally spaced knots in the years (0, 25, 50, 75, 100), with constraints to fulfil the zero-integral condition. We remark that the choice of the number and the position of the knots can follow the same strategies that are used in FDA. Here, the number of knots has been set in order to obtain a good fit to the raw data and to avoid undesired artefacts at the boundaries of the domain. Fig. 8(a) represents two instances of raw data, together with the corresponding smoothed curves, in the original as well as in the clr-space; the whole sample of smoothed densities is displayed in Fig. 8(b), coloured according to gender information. As observed in the simulation study, the clr-transform seems to highlight the variability near the right boundary of the support, which is associated with the smallest values in the density functions.

The smoothed data have been embedded into the space $\mathcal{B}^2(I)$ of functional compositions, and the methodology devised in Section 4 has been applied, resorting to the clr-transform (2) to make computations. Here, we illustrate the results of the SFPCA applied separately to the male and female sub-datasets, summarized in Fig. 9. Fig. 9(c) and (d) reports the first two SFPCs for males and females, clr-transformed to ease their interpretation. Fig. 9(c) evidences that for both groups, the first clr-SFPC contrasts the right tail of the distribution (i.e., the incidence of the old population) against its left part. This is a clear

(a) Example of smoothing raw data via B-splines.


(b) Smoothed data.
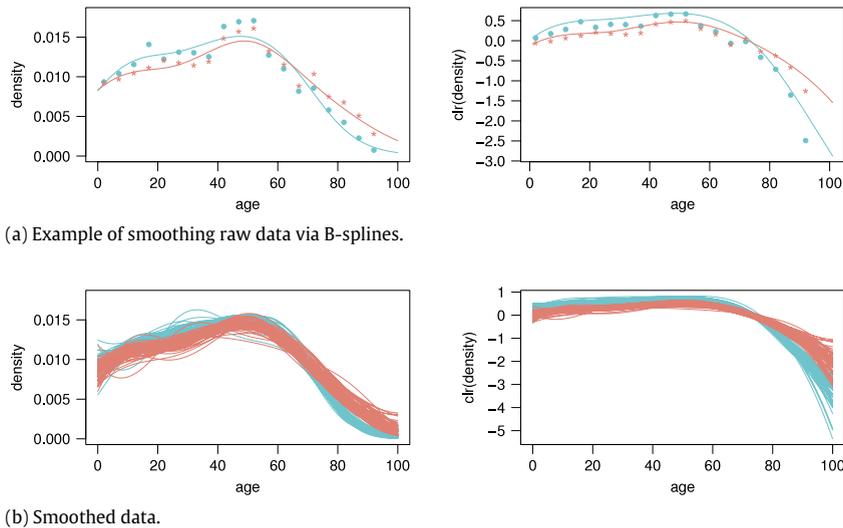
Fig. 8. Population age densities in Upper Austria and their clr-transforms.

consequence of the relative scale property of densities – captured by the clr-transformation – that highlights the variability within the regions of small values of the densities. From a practical perspective, this is important and readily interpretable, as the information content related to the oldest class of population is associated to the life expectancy, and thus to the living conditions.

Fig. 9(e) and (g) displays, for females and males, the plot of the mean perturbed by the first SFPC powered by $\pm$ twice the standard deviation along the corresponding direction (i.e., $\overline{X} \oplus / \ominus 2\sqrt{\lambda_1} \odot \zeta_1$). Recall that this is interpreted as updating/cancelling the information carried by the mean density by twice the evidence within the first mode of variability. Having fixed the sign of the clr-eigenfunction as in Fig. 9(c), high scores along the first SFPCs associate with a high incidence of the old population on the overall number of inhabitants; conversely, low scores associate with a high relative contribution of the youth to the overall population (Fig. 9(b)). We note that, even though the clr-eigenfunctions of the two sub-populations appear pretty similar, they cross the zero level for different ages, namely at about 75 years in the females group and 84 years in the male group. Therefore, the "old" class identified by the first SFPC appears to start earlier in the female than in the male subgroup.

The second SFPCs, displayed in Fig. 9(d), still characterize the variability of the right part of the distribution. Indeed, the main contribution to both the second clr-SFPCs is provided by the contrast between the 69–95 years-old population (associated with low scores in Fig. 9(b)) and the remaining part of the population (associated with high scores in Fig. 9(b)), with particular emphasis to the left and right boundaries of the densities support. We note particularly that high scores along the second SFPC – where the sign of the eigenfunction is set as in Fig. 9(d) – associate with heavy tails and vice versa. However, we note that some uncertainty could affect the second estimated SFPC, due to the absence of data for ages lower than 2 years or higher than 92 years as these values formed interval representatives in the aggregated dataset.

To compare the results obtained in $\mathcal{B}^2(I)$ with those that would have been obtained in $L^2(I)$, we refer to Figs. 10 and 11. Fig. 10 compares the covariance operators estimated for the two groups in $\mathcal{B}^2(I)$ and $L^2(I)$. These are quite different, even though in both cases the covariance structures of the female and male subgroups appear pretty similar. Fig. 10(a) confirms that in $\mathcal{B}^2(I)$ the major contribution to the variability is provided by the right tails of the distributions. Instead, Fig. 10(b) evidences that the L^2 geometry, looking at the data on an absolute scale, ascribes more variability to the left part of the support. This directly reflects on the corresponding principal components, displayed in Fig. 11(c)–(d). Indeed, even though the first FPC of female subpopulation (Fig. 11(c)) is interpreted similarly as the corresponding first SFPC, it attributes much higher variability to the left part of the support. Instead, the male first FPC appears hard to interpret, and seems to characterize the position of the mode of the density and the mass concentration around it. The second female FPC seems to resemble the first male FPC; no straightforward interpretation is suggested by the graphical inspection of the second male FPC.

The metric used to measure the variability readily reflects on the dimensionality reduction. Indeed, the scree-plot relative to the FPCA (Fig. 11(a)) suggests the reduction to three or five FPCs, as opposed to its SFPCA counterpart (Fig. 9(a)) which suggests a more parsimonious representation based on two SFPCs instead. Indeed, the first two SFPCs explain the 88.7% of the variability for the female subpopulation, and 92.6% for the male one. The approximation of the densities obtained by projecting the smoothed data on the space generated by the corresponding first two SFPCs is shown in Fig. 9(j) (colours correspond to the gender information). The approximation obtained by retaining the same number $K = 2$ of FPCs – explaining 69.1% and 67.9% of the L^2 variability for female and male subpopulations, respectively – is displayed in Fig. 11(j). We notice that the densities approximated via SFPCA tend to precisely reproduce the pattern visible in the right part of the domain, while they seem rather smoothed within the left part. Instead, the approximation via FPCA seems to better reproduce the

(a) Explained variance.

(b) Scores along $SFPC_1$ and $SFPC_2$.

(c) $SFPC_1$ (f: 77.75%; m: 82.51% of variability).

(d) $SFPC_2$ (f: 10.95%; m: 10.06% of variability).

(e) f: Mean $\pm 2\sqrt{\lambda_1}$ $SFPC_1$.

(f) f: Mean $\pm 2\sqrt{\lambda_2}$ $SFPC_2$.

(g) m: Mean $\pm 2\sqrt{\lambda_1}$ $SFPC_1$.

(h) m: Mean $\pm 2\sqrt{\lambda_2}$ $SFPC_2$.

(i) Original densities.

(j) Approximated densities (via $SFPC_1$ and $SFPC_2$).
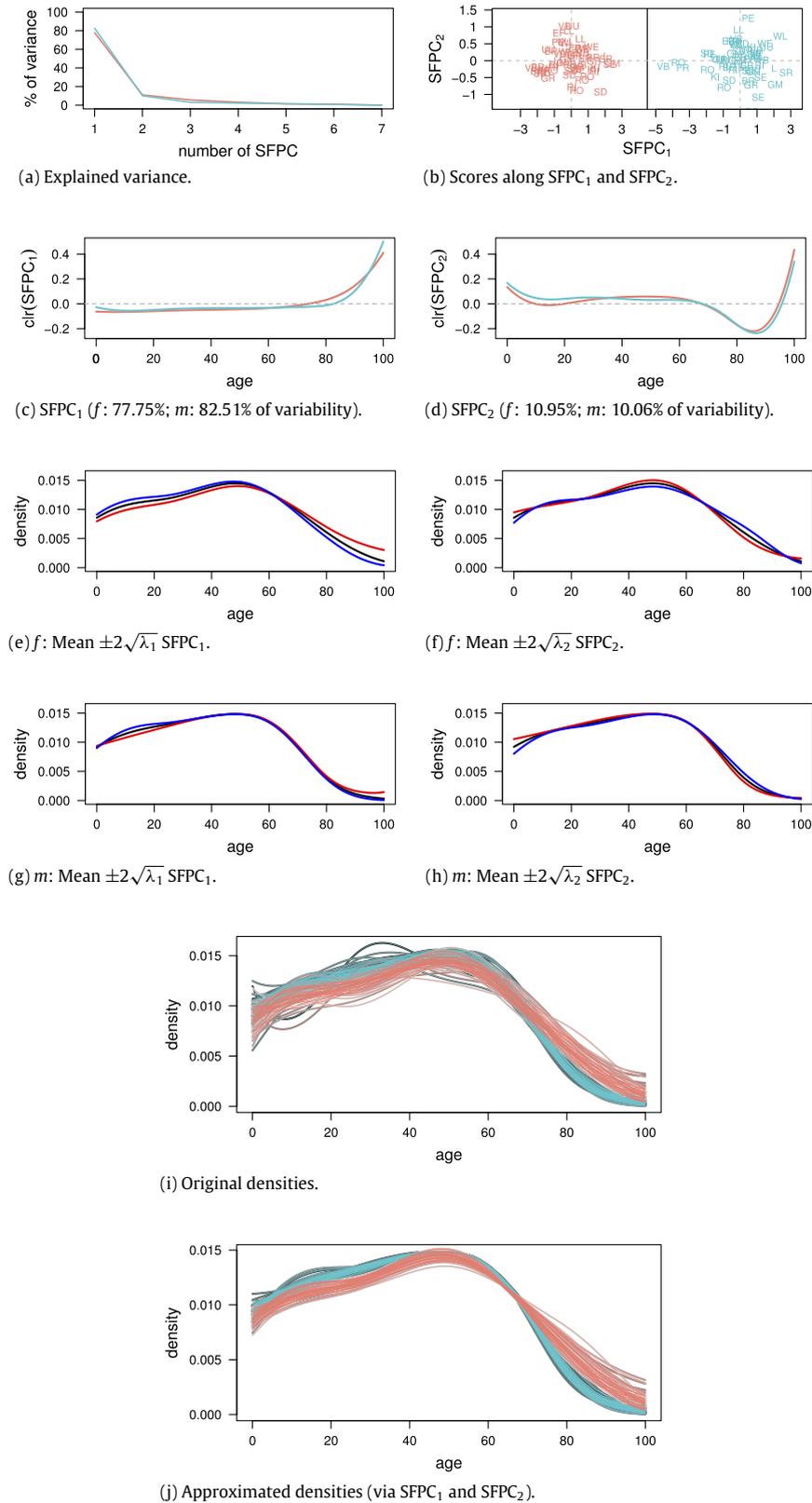
Fig. 9. SFPCA of population age densities. In panels (e)–(h): the black curve indicates the mean function, the red curve indicates the mean $\ominus$ twice the SFPC, the blue curve indicates the mean $\oplus$ twice the SFPC. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(a) $\mathcal{B}^2$ sample covariance operators for female (left) and male (right) subpopulations.

(b) L^2 sample covariance operators for female (left) and male (right) subpopulations.
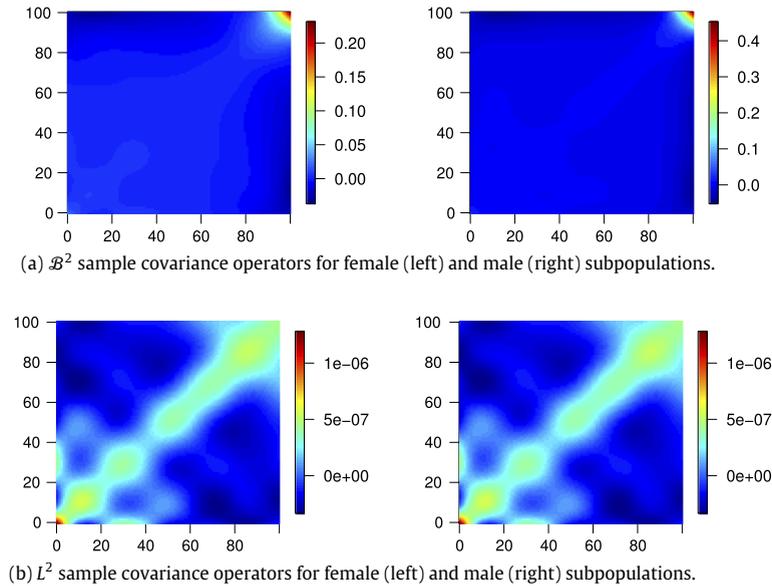
Fig. 10. Sample covariance operators estimated in $\mathcal{B}^2(I)$ and $L^2(I)$ from the smoothed data.

data variability. This is due to the problem of evaluating densities on a relative basis, when they are represented within the space that one is used to evaluate on an absolute scale. Indeed, the dimensionality reduction operated via SFPCA seeks to resemble the relative contributions of each sub-domain to the measure of the support (i.e., it reflects the relative scale property of a functional composition). As such, care needs to be taken when evaluating the quality of the approximation by graphical comparison of Fig. 9(i) and (j). For this reason, we recommend the joint use of the representation in the original and in the clr-space, the latter being reported in Fig. 12. Here, the approximation obtained via SFPCA (Fig. 12(b)) appears much more accurate than that of FPCA (Fig. 12(c)).

For spatial interpretation we focus on SFPCA scores. Fig. 9(b) shows the scores relative to the first two SFPCs, with the signs fixed as in Fig. 9(c)–(d) and coloured according to the gender information. The geographical interpretation is based on two representations of scores: (a) the score maps – in Fig. 13 – and (b) the LISA maps (Local Indicators of Spatial Association, Anselin, 1995) based on Local Morgan's index—in Fig. 14. The Moran's index yields evidence of a significant spatial dependence at a given location when its observed value is much greater than that observed in case of a random allocation of the score values to the neighbours of the location (see, e.g. Delicado and Broner, 2008). Accordingly, LISA maps display with colours only the tracts with significant spatial dependence, which are distinguished in: (i) locations with high score values surrounded by high score values (labelled as high–high, marked with red colour); (ii) areas with low score values with neighbours presenting low score values as well (labelled as low–low, marked with blue colour); (iii)–(iv) high score values surrounded by tracts with low values (labelled as high–low) or vice versa (labelled as low–high). The cases (i) and (ii) correspond to significantly high values of Local Moran's index; the cases (iii) and (iv) correspond to spatial outliers. The latter case has not occurred in the case of age distributions in Upper Austria (Fig. 14).

The graphical inspection of Fig. 13 suggests that the scores may include a spatial information. Indeed, a number of homogeneous areas – with similar patterns for males and females – can be recognized in the map. For instance, the district of *Gmunden* (see also Fig. 7) presents high scores for both women and men along the first component (i.e., high incidence of the old population). In fact, this region is not particularly industrialized and people moved to the areas of *Linz* (the capital of Upper Austria) and *Steyr* to live closer to their working place. This migration is well reported and analysed (see, e.g., Schöfecker and Larndorfer, 2012; Government of Upper Austria, Department of Statistics, 2011). The LISA maps show the same characteristics marking as high–high the district of *Gmunden*, the eastern part of *Kirchdorf* and – especially for men – the *Ennstal* (i.e., the valley formed by the river "*Enns*", which is part of the district *Steyr Land*). Here, the emigration of people to other parts of Upper Austria is high, with a decrease of population size of about 10% in many areas of the district of *Gmunden* (see, e.g., Government of Upper Austria, Department of Statistics, 2013a). The areas in the North and East of *Linz* behave differently, as many young families have moved to the green areas having good connection to *Linz*. Within these regions, the fertility rate is much higher than the average of Upper Austria and also the population growth is substantially higher than the average. This regional pattern is well captured by the LISA map in Fig. 7. For instance, the area marked in blue in the East of *Linz* (*Altenmarkt*, a part of the region *Urfahr Umgebung*) experienced a population growth of about 10% in the last 10 years (Government of Upper Austria, Department of Statistics, 2013b).

Regarding the second SFPC (bottom panels in Fig. 13), different regional structures appear in the North with respect to the rest of the map. Indeed, the district of *Rohrbach* is associated with pretty low scores for men (i.e., high incidence of 69–95 years old population), and very low scores for women. Similarly, very low scores appear for women in the *Schärd-*
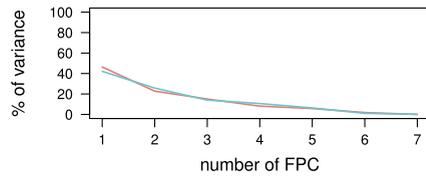
(a) Explained variance.
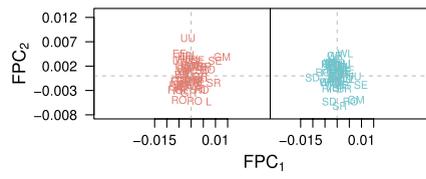
(b) Scores along FPC_1 and FPC_2.
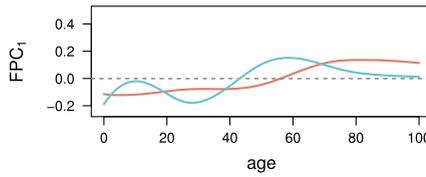
(c) FPC_1 (f: 46.39%; m: 42.12% of variability).
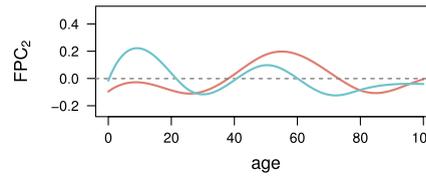
(d) FPC_2 (f: 22.76%; m: 25.77% of variability).
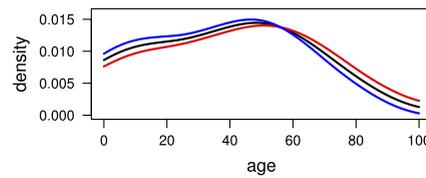
(e) f: Mean $\pm 2\sqrt{\rho_1}$ FPC_1.
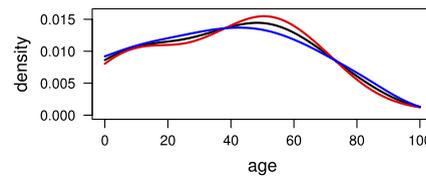
(f) f: Mean $\pm 2\sqrt{\rho_2}$ FPC_2.
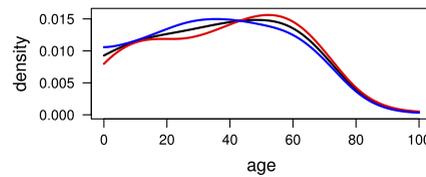
(g) m: Mean $\pm 2\sqrt{\rho_1}$ FPC_1.
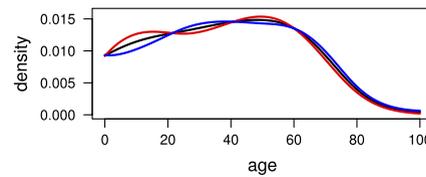
(h) m: Mean $\pm 2\sqrt{\rho_2}$ FPC_2.
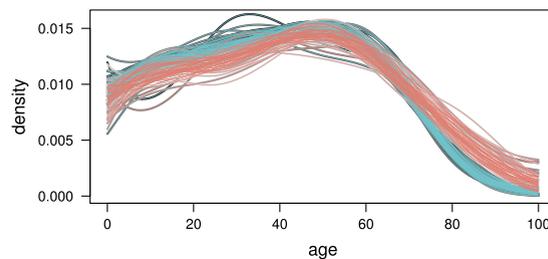
(i) Original densities.
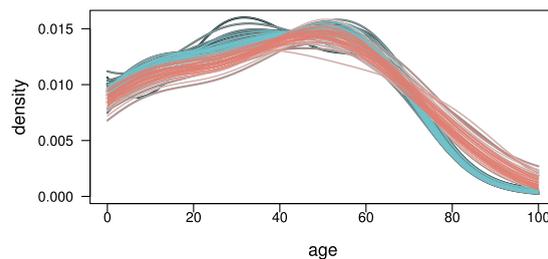
(j) Approximated densities (via FPC_1 and FPC_2).

Fig. 11. FPCA in the $L^2(I)$ geometry. In panels (e)–(h): the solid black curve indicates the mean function, the red curve indicates the mean minus the SFPC, the blue curve indicates the mean plus the SFPC. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(a) Original densities.   (b) Approximated densities (via $SFPC_1$ and $SFPC_2$).   (c) Approximated densities (via FPC_1 and FPC_2).
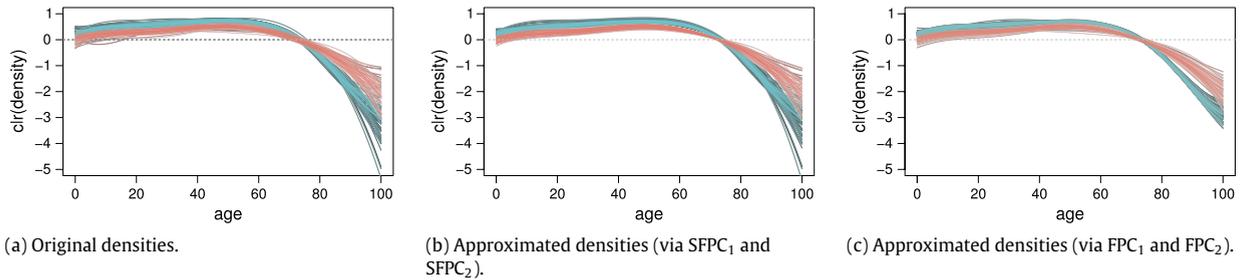
Fig. 12. Dimensionality reduction in $\mathcal{B}^2(I)$ and $L^2(I)$: approximation via the first two principal components and original data, in the clr-space.
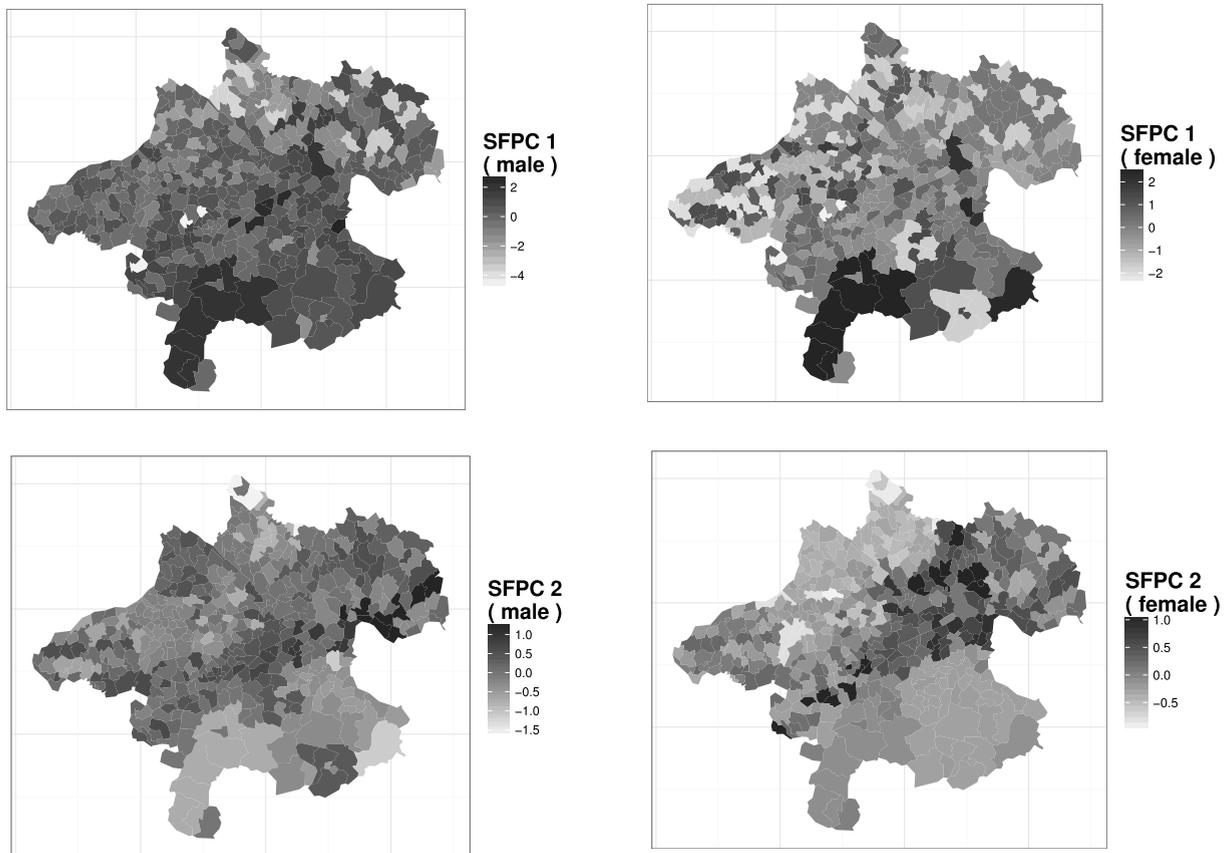

Fig. 13. Geographical regional representation of the first and second SFPCs scores.

ing district. These regions are mostly rural, with very few industries. In particular, these areas have low fertility rates and negative balance regarding the population size, see the area of *St. Florian am Inn* (in *Schärding*), which is a low–low area for men and women (Government of Upper Austria, Department of Statistics, 2013c). Instead, high scores appear in the regions around *Linz* and around the highway that crosses Upper Austria in direction E–NE to W–SW. This is particularly evident for women, marked with a red zone in the LISA map (Fig. 14). Several industries are placed nearby that highway. This makes the area particularly attractive for a large portion of the young population, which moves there to work.

7. Conclusions

The choice of an appropriate space to perform the analysis is crucial prior to any statistical processing using FDA methods. This is particularly evident in the presence of constrained data, such as functional compositions. We focused on the problem of dimensionality reduction for probability density functions. In this case, although the numerical problems resulting from the unit-integral constraint of densities could be overcome by applying an appropriate pre-processing (e.g., log-transformation), their inherent properties – as scale invariance and relative scale – are only captured when the theory of Bayes spaces is incorporated. The centred log-ratio transformation isometrically maps the Bayes space $\mathcal{B}^2(I)$ into the space
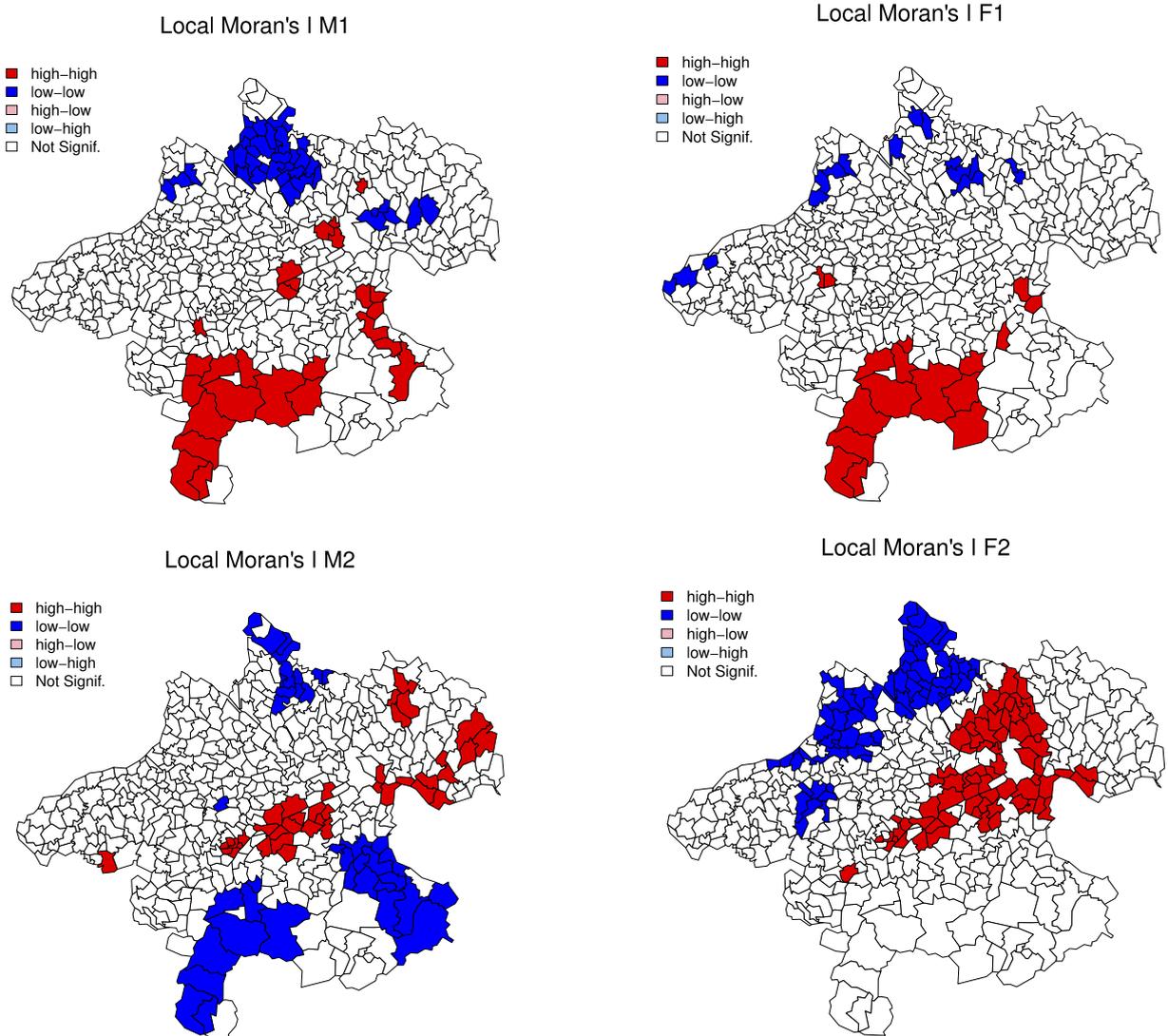
Fig. 14. Geographical regional representation of the first and second SFPCs scores using LISA maps. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$L^2(I)$, and provides a way to easily apply the well-known FDA methods in the presence of functional compositions. In this sense, it is possible to meaningfully use the standard tools for the interpretation of FPCA (e.g., the plot of mean $\pm$ eigenfunctions), if interpreted in the light of the Aitchison geometry which takes into account the relative information captured by density functions. We note that our proposal stands in continuity with the work of Delicado (2011), who pointed out that his most promising results were obtained through multidimensional scaling in the Bayes space $\mathcal{B}^2(I)$. However, our methodological developments provide a clear direction for the extension to density functions of several methods which are in use in FDA, besides opening a variety of challenges for the future. One of these is the possibility of considering alternative computational tools, such as the log-ratio transformations with respect to an orthonormal basis in the Bayes space, with the aim of avoiding the zero-integral constraint resulting from the clr-transformation. Even more promising could be the possibility of extending the support I to the general case of the whole real line (or any Borel subset) as proposed in Van den Boogaart et al. (2014). The choice of a non-uniform reference measure still needs to be thoroughly discussed in terms of applicative consequences, and certainly deserves to be further investigated.

Acknowledgements

Appendix. Computation of B-spline coefficients

In this Appendix, we briefly describe the procedure to obtain the B-splines coefficients that honour the zero integral constraint of clr-transformed densities. We refer to Machalová et al. (2015) for further details. Let $\Delta\tau := \tau_0 = a < \tau_1 < \cdots < \tau_g < b = \tau_{g+1}$ be a given sequence of knots in I. Then every polynomial spline s_k of degree $k > 0$ has a unique representation

$$s_k(t) = \sum_{i=-k}^{g} c_i \phi_i^{k+1}(t), \tag{14}$$

where $\mathbf{c} = (c_{-k}, \ldots, c_g)^\top$ is the vector of *B*-spline coefficients of s_k, and $\phi_i^{k+1}(t)$, $i = -k, \ldots, g$ are *B*-splines of degree k. We call

$$\boldsymbol{\Phi}_{k+1}(\mathbf{t}) = \begin{pmatrix} \phi_{-k}^{k+1}(t_1) & \cdots & \phi_g^{k+1}(t_1) \\ \vdots & \ddots & \vdots \\ \phi_{-k}^{k+1}(t_n) & \cdots & \phi_g^{k+1}(t_n) \end{pmatrix} \in \mathbb{R}^{n, g+k+1} \tag{15}$$

the collocation matrix and $\mathbf{S}_l = \mathbf{D}_l \mathbf{L}_l \ldots \mathbf{D}_1 \mathbf{L}_1 \in \mathbb{R}^{g+k+1-l, g+k+1}$ the upper triangular matrix, which is formed by the matrices $\mathbf{D}_j \in \mathbb{R}^{g+k+1-j, g+k+1-j}$ such that

$$\mathbf{D}_j = (k+1-j) \operatorname{diag}\left(d_{-k+j}, \ldots, d_g\right) \quad \text{with } d_i = \frac{1}{\tau_{i+k+1-j} - \tau_i} \ \forall i = -k+j, \ldots, g$$

and

$$\mathbf{L}_j := \begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix} \in \mathbb{R}^{g+k+1-j, g+k+2-j}.$$

Given a raw datum $x(t_i)$, observed at the time points $t_i \in I$, $i = 1, \ldots, n$, we seek for the coefficients of the smoothing spline (de Boor, 2001) that minimizes the functional

$$J_l(s_k) = \sum_{i=1}^{n} w_i \left[x(t_i) - s_k(t_i)\right]^2 + \lambda \int_I \left[s_k^{(l)}(t)\right]^2 dt, \tag{16}$$

and fulfils the condition

$$\int_I s_k(t) \, dx = 0, \tag{17}$$

where $w_i \geq 0$, $i = 1, \ldots, n$, $n \geq g+1$, are given weights and $\lambda \geq 0$ a given parameter. The parameter λ controls the impact of the differential penalization appearing in (16), and is thus associated with the smoothness of the resulting approximation. In the case study considered in Section 6, we set the weights as well as the value of the smoothing parameter λ to one, following the default setting of Machalová et al. (2015). For possible sensible determination of λ using, e.g., cross-validation, we refer to Kneip et al. (2012) and Liebl (2013). In general, for the purpose of setting the parameters, all the techniques which are in use in FDA can be employed in this case as well.

In our setting, an explicit expression can be derived for the optimal coefficients. Indeed, they are obtained as (Machalová et al., 2015)

$$\bar{\mathbf{c}}^* = \mathbf{D}\mathbf{K} \left[(\boldsymbol{\Phi}_{k+1}(\mathbf{t})\mathbf{D}\mathbf{K})^\top \mathbf{W} \boldsymbol{\Phi}_{k+1}(\mathbf{t})\mathbf{D}\mathbf{K} + \lambda (\mathbf{D}\mathbf{K})^\top \mathbf{N}_{kl} \mathbf{D}\mathbf{K} \right]^+ \mathbf{K}^\top \mathbf{D}^\top \boldsymbol{\Phi}_{k+1}^\top(\mathbf{t})\mathbf{W}\mathbf{y},$$

where $\mathbf{W} = \operatorname{diag}(\mathbf{w})$, $\mathbf{A}^+$ denotes Moore–Penrose pseudoinverse of a matrix $\mathbf{A}$,

$$\mathbf{D} = (k+1) \operatorname{diag}\left(\frac{1}{\tau_1 - \tau_{-k}}, \ldots, \frac{1}{\tau_{g+k+1} - \tau_g}\right) \in \mathbb{R}^{g+k+1, g+k+1},$$

$$\mathbf{K} = \begin{pmatrix} 1 & 0 & 0 & \cdots & -1 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \end{pmatrix} \in \mathbb{R}^{g+k+1, g+k+1}$$

and $\mathbf{N}_{kl} = \mathbf{S}_l^\top \mathbf{M}_{kl}\mathbf{S}_l$ is a positive semidefinite matrix, with

$$
\mathbf{M}_{kl} = \begin{pmatrix} \langle \phi_{-k+l}^{k+1-l}, \phi_{-k+l}^{k+1-l} \rangle_2 & \cdots & \langle \phi_g^{k+1-l}, \phi_{-k+l}^{k+1-l} \rangle_2 \\ \vdots & & \vdots \\ \langle \phi_{-k+l}^{k+1-l}, \phi_g^{k+1-l} \rangle_2 & \cdots & \langle \phi_g^{k+1-l}, \phi_g^{k+1-l} \rangle_2 \end{pmatrix} \in \mathbb{R}^{g+k+1-l,g+k+1-l}.
$$

References

Aitchison, J., 1986. The Statistical Analysis of Compositional Data. In: Monographs on Statistics and Applied Probability, Chapman and Hall Ltd., London, UK, p. 416. (Reprinted 2003 with additional material by The Blackburn Press).
Anselin, L., 1995. Local indicators of spatial association-LISA. Geogr. Anal. 2, 93–115.
de Boor, C., 2001. A Practical Guide to Splines. Springer, New York.
Delicado, P., 2007. Functional k-sample problem when data are density functions. Comput. Statist. 22, 391–410.
Delicado, P., 2011. Dimensionality reduction when data are density functions. Comput. Statist. Data Anal. 55, 401–420.
Delicado, P., Broner, S., 2008. Distance-based LISA maps for multivariate lattice data. Technical Report. Universitat Politecnica de Catalunya, Preprint http://hdl.handle.net/2117/2421.
Egozcue, J.J., 2009. Reply to "On the Harker Variation Diagrams; …" by J.A. Cortés. Math. Geosci. 41 (7), 829–834.
Egozcue, J.J., Díaz-Barrero, J.L., Pawlowsky-Glahn, V., 2006. Hilbert space of probability density functions based on Aitchison geometry. Acta Math. Sin. (Engl. Ser.) 22 (4), 1175–1182.
Egozcue, J.J., Pawlowsky-Glahn, V., 2006. Simplicial geometry for compositional data. In: Buccianti, A., Mateu-Figueras, G., Pawlowsky-Glahn, V. (Eds.), Compositional Data Analysis in the Geosciences: From Theory to Practice. In: Special Publications, vol. 264. Geological Society, London, pp. 145–160.
Egozcue, J.J., Pawlowsky-Glahn, V., Tolosana-Delgado, R., Ortego, M.I., van den Boogaart, K.G., 2013. Bayes spaces: use of improper distributions and exponential families. Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat. 107, 475–486.
Government of Upper Austria, Department of Statistics, 2011. https://www.wko.at/Content.Node/Interessenvertretung/Daten–Fakten–Zahlen/-Positionen-/b/ooe/Demografie-Serie3.pdf.
Government of Upper Austria, Department of Statistics, 2013a. http://www.land-oberoesterreich.gv.at/files/statistik/gesellschaftundsoziales/bevstand/bevstand_40706.pdf.
Government of Upper Austria, Department of Statistics, 2013b. http://www.land-oberoesterreich.gv.at/files/statistik/gesellschaftundsoziales/bevstand/bevstand_41602.pdf.
Government of Upper Austria, Department of Statistics, 2013c. http://www.land-oberoesterreich.gv.at/files/statistik/gesellschaftundsoziales/natbev/natbev_41418.pdf.
Horváth, L., Kokoszka, P., 2012. Inference for Functional Data with Applications. In: Springer Series in Statistics, Springer.
Johnson, R.A., Wichern, D.W., 2002. Applied Multivariate Statistical Analysis, fifth ed. Prentice Hall, London.
Jones, M.C., Rice, J.A., 1992. Displaying the important features of large collections of similar curves. Amer. Statist. 46 (2), 140–145.
Kneip, A., Sickles, R.C., Song, W., 2012. A new panel data treatment for heterogeneity in time trends. Econometric Theory 28, 590–628.
Kneip, A., Utikal, K., 2001. Inference for density families using functional principal component analysis. J. Amer. Statist. Assoc. 96, 519–542.
Liebl, D., 2013. Modeling and forecasting electricity spot prices: A functional data perspective. Ann. Appl. Stat. 7, 1562–1592.
Machalová, J., Hron, K., Monti, G.S., 2015. Preprocessing of centred logratio transformed density functions using smoothing splines. arXiv:1501.07047.
Menafoglio, A., Guadagnini, A., Secchi, P., 2014a. A kriging approach based on Aitchison geometry for the characterization of particle-size curves in heterogeneous aquifers. Stoch. Environ. Res. Risk Assess. 28 (7), 1835–1851.
Menafoglio, A., Secchi, P., Dalla Rosa, M., 2013. A universal kriging predictor for spatially dependent functional data of a Hilbert space. Electron. J. Stat. 7, 2209–2240.
Menafoglio, A., Secchi, P., Guadagnini, A., 2014b. A class-kriging predictor for functional compositions with application to particle-size curves in heterogeneous aquifers. MOX-report 58/2014.
Nerini, D., Ghattas, B., 2007. Classifying densities using functional regression trees: applications in oceanology. Comput. Statist. Data Anal. 51, 4984–4993.
Pawlowsky-Glahn, V., Buccianti, A. (Eds.), 2011. Compositional Data Analysis: Theory and Applications. John Wiley & Sons, Ltd., Chichester, UK, p. 378.
Pawlowsky-Glahn, V., Egozcue, J.J., 2001. Geometric approach to statistical analysis on the simplex. Stoch. Environ. Res. Risk Assess. 15 (5), 384–398.
Ramsay, J., Silverman, B.W., 2002. Applied Functional Data Analysis: Methods and Case Studies. Springer-Verlag, New York.
Ramsay, J., Silverman, B.W., 2005. Functional Data Analysis, second ed. Springer, New York.
Schöfecker, M., Larndorfer, E., 2012. Oberösterreich: Zahlen & Fakten Jahresausgabe 2013. Report of the Government of Upper Austria. pp. 1–72.
Shang, H.L., 2014. A survey of functional principal component analysis. Adv. Stat. Anal. 98, 121–142.
Van den Boogaart, K.G., Egozcue, J.J., Pawlowsky-Glahn, V., 2010. Bayes linear spaces. Statist. Oper. Res. Trans. 34 (2), 201–222.
Van den Boogaart, K.G., Egozcue, J.J., Pawlowsky-Glahn, V., 2014. Bayes Hilbert spaces. Aust. N. Z. J. Stat. 56 (2), 171–194.
Witting, H., 1985. Mathematische Statistik I. Parametrische Verfahren bei festem Stichprobenumfang. B.G. Teubner, Stuttgart, DE, p. 538.
Zhang, Z., Müller, H.G., 2011. Functional density synchronization. Comput. Statist. Data Anal. 55, 2234–2249.