Adv Data Anal Classif (2014) 8:231–255
DOI 10.1007/s11634-013-0158-y

REGULAR ARTICLE

Functional data clustering: a survey

Julien Jacques · Cristian Preda

Received: 28 March 2013 / Revised: 25 September 2013 / Accepted: 16 November 2013 /
Published online: 29 November 2013

Abstract Clustering techniques for functional data are reviewed. Four groups of clustering algorithms for functional data are proposed. The first group consists of methods working directly on the evaluation points of the curves. The second groups is defined by filtering methods which first approximate the curves into a finite basis of functions and second perform clustering using the basis expansion coefficients. The third groups is composed of methods which perform simultaneously dimensionality reduction of the curves and clustering, leading to functional representation of data depending on clusters. The last group consists of distance-based methods using clustering algorithms based on specific distances for functional data. A software review as well as an illustration of the application of these algorithms on real data are presented.

Keywords Functional data · Clustering · Basis expansion · Functional principal component analysis

Mathematics Subject Classification 62-07 · 62M99 · 62H30

1 Introduction

The aim of cluster analysis is to identify homogeneous groups (clusters) of observations representing realisations of some random variable X. Clustering is often used as a preliminary step for data exploration, the goal being to detect particular patterns

J. Jacques (✉) · C. Preda
Laboratoire Paul Painlevé, UMR CNRS 8524, Université Lille 1 and Inria Lille-Nord Europe,
U.F.R. de Mathématiques, 59655 Villeneuve d'Ascq Cédex, France
e-mail: julien.jacques@polytech-lille.fr

C. Preda
e-mail: cristian.preda@polytech-lille.fr

in data that have some convenient interpretation for the user. In the finite dimensional setting, X is a random vector with values in $\mathbb{R}^p$, $X = (X_1, \ldots, X_p)$, $p \geq 1$. Earliest methods relies on dissimilarity measures between pairs of observations. A popular dissimilarity measure is based on the distance between groups, introduced by Ward and Joe (1963) for hierarchical clustering. Similarly, the k-means algorithm (Hartigan and Wong 1978) is probably the most popular clustering algorithm among the geometric procedures. More recently, probabilistic approaches have been introduced to characterize the notion of cluster through their probability density (Banfield and Raftery 1993; Celeux and Govaert 1995; McLachlan and Peel 2000).

In the last years, researchers concentrated their efforts to solve problems (regression, clustering) when the dimension p is large, in its absolute value or with respect to the size of some learning sample of data. The curse of dimensionality was and is still a very active topic in statistics. A particular case is that of random variables taking values into an infinite dimensional space, typically a space of functions defined on some set $\mathcal{T}$ (time interval for instance). Such data is represented by curves (*functional data*) and the random variable underlying data is a stochastic process $X = \{X(t),\ t \in \mathcal{T}\}$. If this type of data was for long-time inaccessible for statistics (because of technological limitations), today it becomes more and more easy to observe, store and process large amounts of such data in medicine, economics, chemometrics and many others domains (see Ramsay and Silverman 2005 for an overview).

Functional data is represented by a set of curves belonging to an infinite dimensional space (Ferraty and Vieu 2006). It is well known that, theoretically, the infinite dimension is the main source of difficulty in modelling such data. For example, in the linear regression framework, the estimation of the regression function is inconsistent because the covariance operator is not inversible (Cardot et al. 1999). In the clustering context, the lack of a definition for the probability density of a functional random variable (Delaigle and Hall 2010) makes the probabilistic model-based methods not directly applicable. Also, without an explicit analytic expression of functions (they are often observed at discrete points), distances can not be exactly computed and thus distance-based methods fail. From a practical point of view, a common solution consists in representing the curves into a finite dimensional space of functions. Then, clustering algorithms for finite dimensional data can be performed, distance between functions can be approximated, etc. More recent works perform dimensionality reduction and clustering simultaneously

The aim of this paper is to propose a survey of clustering approaches for functional data. It is organized as follows. Section 2 introduces functional data and functional principal component analysis as the main tool for analysing and clustering functional data. Section 3 proposes a classification of the different clustering methods for functional data into four groups: raw-data methods working directly on the evaluation points of the curves; filtering methods which first approximate the curves into a finite basis of functions and second perform clustering using the basis expansion coefficients; adaptive methods performing simultaneously dimensionality reduction of the curves and clustering; distance-based methods using specific distances for functional data. Section 4 discusses the common problems of selecting the number of clusters and of choosing the appropriate representation for functional data. Section 5 presents some software for clustering functional data. In Sect. 6 the main clustering algorithms

are illustrated using real datasets. Some open problems in the framework of functional data clustering end the paper.

2 Functional data analysis

Functional data analysis (FDA) extends the classical multivariate methods when data are functions or curves. Some examples of such data are presented in Fig. 1: the top figure (a) plots the evolution of some stock-exchange index observed during 1 h; the bottom figure (b) presents the knee flexion angle observed over a gait cycle.

The first contributions to functional data analysis concern the factorial analysis and are mainly based on the Karhunen–Loeve expansion of a second order L_2-continuous stochastic process (Karhunen 1947; Loève 1945). A pioneer work paper on the subject is due to Deville (1974)—a one hundred pages paper in the *Annales de l'INSEE*—with some applications in economics. In Dauxois et al. (1982), Antoniadis and Beder (1989) the authors obtained asymptotic results for the elements derived from factorial analysis. The contributions of Besse (1979), Saporta (1981) extend to functional data the principal component analysis, the canonical analysis of two functional variables, the multiple correspondence analysis for functional categorical data and the linear regression on functional data. Notice also the contribution to functional categorical data due to Boumaza (1980).

Let us also remind the monographs on functional data by Ramsay and Silverman (2002, 2005) developing theory and applications on functional data, the book of Bosq

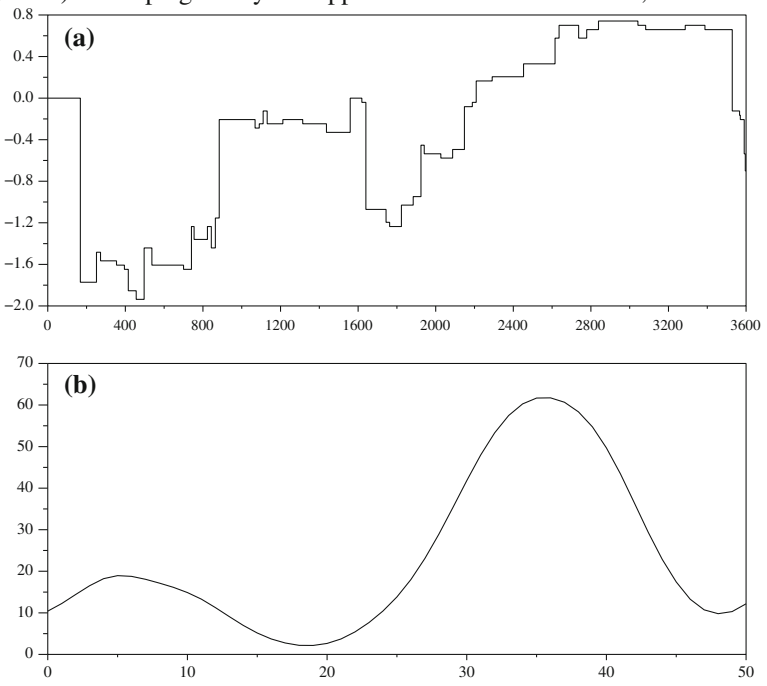

Fig. 1 Some examples of functional data. **a** Share index evolution during 1 h. **b** Knee flexion angle (degree) over a complete gait cycle

(2000) for modelling dependent functional random variables and the recent book of Ferraty and Vieu (2006) on non-parametric models for functional data containing a survey of the most recent contributions on this topic.

2.1 Functional data

According to Ferraty and Vieu (2006), a functional random variable X is a random variable with values in an infinite dimensional space. Then, *functional data* represents a set of observations $\{X_1, \ldots, X_n\}$ of X. The underlying model for X_i's is usually an i.i.d. sample of random variables drawn from the same distribution as X.

A well accepted model for this type of data is to consider it as paths of a stochastic process $X = \{X_t\}_{t \in \mathcal{T}}$ taking values in some Hilbert space, H, of functions defined on some set $\mathcal{T}$. Generally, $\mathcal{T}$ represents an interval of time, of wavelengths or any other subset of $\mathbb{R}$. We restrict our presentation to the case where H is a space of real-valued functions. For multivariate functional data (elements of H are $\mathbb{R}^p$-valued functions, $p \geq 2$) the reader can refer for instance to Jacques and Preda (2013b) for a recent work on multivariate functional data clustering.

The main source of difficulty when dealing with functional data relies on the fact that the observations are supposed to belong to an infinite dimensional space, whereas in practice one only has sampled curves observed into a finite set of observation points. Indeed, it is usual that we only have discrete observations X_{ij} of each sample path $X_i(t)$ at a finite set of knots $\{t_{ij} : j = 1, \ldots, m_i\}$. Because of this, the first step in FDA is often the reconstruction of the functional form of data from discrete observations. The most common solution to this problem is to consider that sample paths belong to a finite dimensional space spanned by some basis of functions (see, for example, Ramsay and Silverman 2005). An alternative way of solving this problem is based on non-parametric smoothing of functions (Ferraty and Vieu 2006).

Let us consider a basis $\Phi = \{\phi_1, \ldots, \phi_L\}$ generating some space of functions in H and assume that X admits the basis expansion

$$X_i(t) = \sum_{\ell=1}^{L} \alpha_{i\ell} \phi_\ell(t) \tag{1}$$

for some $L \in \mathbb{N}$, with $\alpha_{i\ell} \in \mathbb{R}$.

The sample paths basis coefficients are estimated from discrete observations by using an appropriate numerical method:

– If the sample curves are observed without error

$$X_{ij} = X_i\left(t_{ij}\right) \quad j = 1, \ldots, m_i,$$

an interpolation procedure can be used. For example, Escabias et al. (2005) propose quasi-natural cubic spline interpolation to reconstruct annual temperatures curves from monthly values.

– On the other hand, if the functional predictor is observed with error

$$X_{ij} = X_i\left(t_{ij}\right) + \varepsilon_{ij} \quad j = 1, \ldots, m_i,$$

least squares smoothing is used after choosing a suitable basis as, for example, trigonometric functions, B-splines or wavelets (see Ramsay and Silverman 2005 for a detailed study). In this case, the basis coefficients of each sample path $X_i(t)$ are approximated by

$$\widehat{\alpha}_i = \left(\Theta_i'\Theta_i\right)^{-1}\Theta_i'\tilde{X}_i,$$

with $\widehat{\alpha}_i = (\widehat{\alpha}_{i1}, \ldots, \widehat{\alpha}_{iL})'$, $\Theta_i = (\phi_\ell(t_{ij}))_{1\le j\le m_i, 1\le \ell\le L}$ and $\tilde{X}_i = (X_{i1}, \ldots, X_{im_i})'$.

Notice that the representation in (1) defines a functional representation of the curve $X_i(t)$, whatever are the positions of the observed evaluation points $\left\{t_{ij} : j = 1, \ldots, m_i\right\}$. In practice, that allows to deal with irregularly sampled curves with different evaluation points.

2.2 Functional principal component analysis

The optimal representation of the functional data $\{X_1, \ldots, X_n\}$ is the main purpose of Functional Principal Component Analysis (FPCA). Thanks to the practical interest of FPCA for interpretation and data presentation (graphics), it is one of the most used tools in literature for clustering functional data. For this reason, in what follows FPCA is presented in details.

In order to address this question in a formal way, we made the hypothesis that X is a L_2-continuous stochastic process:

$$\forall t \in \mathcal{T}, \quad \lim_{h\to 0} \mathbb{E}\left[|X(t+h) - X(t)|^2\right] = 0,$$

where $\mathbb{E}$ stands for the expectation operator.

The L_2-continuity is a quite general hypothesis, as most of the real data in applications satisfy it. Let $\mu = \{\mu(t) = \mathbb{E}[X(t)]\}_{t\in\mathcal{T}}$ denotes the mean function X.

If $L_2(\mathcal{T})$ is the space of the square-integrable functions defined on $\mathcal{T}$, then the covariance operator $\mathcal{V}$ of X:

$$\begin{aligned} \mathcal{V} &: L_2(\mathcal{T}) \to L_2(\mathcal{T}) \\ f &\xmapsto{\mathcal{V}} \mathcal{V}f = \int_0^T V(\cdot, t) f(t) dt, \end{aligned}$$

is an integral operator with kernel V defined by:

$$V(s,t) = \mathbb{E}\left[(X(s) - \mu(s))(X(t) - \mu(t))\right], \quad s, t \in \mathcal{T}.$$

Under the L_2-continuity hypothesis, the mean and the covariance function are continuous and the covariance operator $\mathcal{V}$ is a Hilbert-Schmidt one (compact, positive and of finite trace).

The spectral analysis of $\mathcal{V}$ provides a countable set of positive eigenvalues $\{\lambda_j\}_{j\ge 1}$ associated to an orthonormal basis of eigenfunctions $\{f_j\}_{j\ge 1}$:

$$\mathcal{V} f_j = \lambda_j f_j, \tag{2}$$

with $\lambda_1 \geq \lambda_2 \geq \ldots$ and $\int_0^T f_j(t) f_{j'}(t)dt = 1$ if $j = j'$ and 0 otherwise.

The *principal components* $\{C_j\}_{j\geq 1}$ of X are random variables defined as the projection of X on the eigenfunctions of $\mathcal{V}$:

$$C_j = \int_0^T (X(t) - \mu(t)) f_j(t)dt.$$

The principal components $\{C_j\}_{j\geq 1}$ are zero-mean uncorrelated random variables with variance $\lambda_j,\ j \geq 1$.

With these definitions, the Karhunen–Loeve expansion (Karhunen 1947; Loève 1945) holds:

$$X(t) = \mu(t) + \sum_{j\geq 1} C_j f_j(t), \qquad t \in \mathcal{T}. \tag{3}$$

Truncating (3) at the first q terms one obtains the best approximation in norm L_2 of $X(t)$ by a sum of quasi-deterministic processes (Saporta 1981),

$$X^{(q)}(t) = \mu(t) + \sum_{j=1}^{q} C_j f_j(t), \qquad t \in \mathcal{T}. \tag{4}$$

In practice, the curves are observed at discrete observation points (see Sect. 2.1). In order to compute the principal components, one needs some approximation presented in Appendix A.

2.3 Preprocessing functional data

Curves are usually observed at discrete observation points. For this reason a first step when working with functional data is to reconstruct the functional form of data. A second important step in functional data analysis is, usually, data registration (Ramsay and Silverman 2005, chap. 7). It consists of centring and scaling the curves in order to eliminate both phase and amplitude variations into the curve's dataset. In a clustering framework, the need of a preprocessing of data is questionable, and several schools of thought exist.

On the one hand, the majority of the existing works does not perform data registration, assuming that either this effect is only limited or it contains cluster information. For instance, in the well-known Canadian weather dataset (temperature and precipitation curves for Canadian weather stations (Ramsay and Silverman 2005), the geographical interpretation of the clusters of weather stations is mainly due to amplitude variability. In that sense, amplitude variation of the data is a source of differentiation between clusters, and must not been removed by a preprocessing step if we are interested in clusters defined by such information. Similarly, a clustering of the Growth

dataset (Ramsay and Silverman 2005; Tuddenham and Snyder 1954), which consists of growth curves for girls and boys, usually exhibit clusters related to the gender of children (Bouveyron and Jacques 2011; Chiou and Li 2007; Jacques and Preda 2013a). In Liu and Yang (2009), the authors perform on this dataset curves registration and clustering simultaneously, and they conclude to the absence of cluster in the data. That means that the main source of heterogeneity in data was due to phase and amplitude variations between curves. Recently, Slaets et al. (2012) define a clustering procedure which is devoted to identify clusters due to phase and amplitude variations.

On the other hand, several works consider that the registration of data must be carried out in a clustering framework. Rather than performing data registration before clustering, they define more efficient procedure which simultaneously align and cluster the curves. Refer for instance to Liu and Yang (2009), Sangalli et al. (2010a,b) for some interesting examples of such works. Usually, in presence of important phase and amplitude variations between curves, a clustering analysis mainly exhibits clusters due to these features. Depending on what the practitioner is looking for, such clusters can be helpful or not. If the answer is no, it is interesting to remove such variations in order to exhibit a new clusters structure (potentially hidden by the phase and amplitude variations) and the works previously cited can be helpful for this. Thus, the interpretation of this new clustering structure will allow to discover new sources of heterogeneity in data.

3 Functional data clustering approaches

Clustering functional data received particular attention from statisticians in the last decade. In this section we present a classification of the different approaches into four groups. This classification is illustrated by the Fig. 2 and described in what follows:

- raw data methods (Sect. 3.1): these methods consists of clustering directly the curves on the basis of their evaluation points.
- filtering methods (Sect. 3.2): these methods first approximate the curves into some basis of functions and second perform clustering using the basis expansion coefficients.
- adaptive methods (Sect. 3.3): these methods consider that the functional representation of data is depending on clusters, and perform simultaneously dimensionality reduction and clustering. Thus, depending on its cluster membership, an observation (a curve) could have different representations.
- distance-based methods (Sect. 3.4): these methods use clustering algorithms based on specific distances for functional data. Notice that, depending on the way these distances are computed, these methods can be related to either raw data or filtering methods.

3.1 Raw-data clustering approaches

The first naive approach of functional data clustering consists of clustering some discretization of data. This approach, quoted as *raw-data* in Fig. 2, is the most simple one, since the functions are usually already observed at some discrete observation points. So, in such situation there is no need to reconstruct the functional form of the

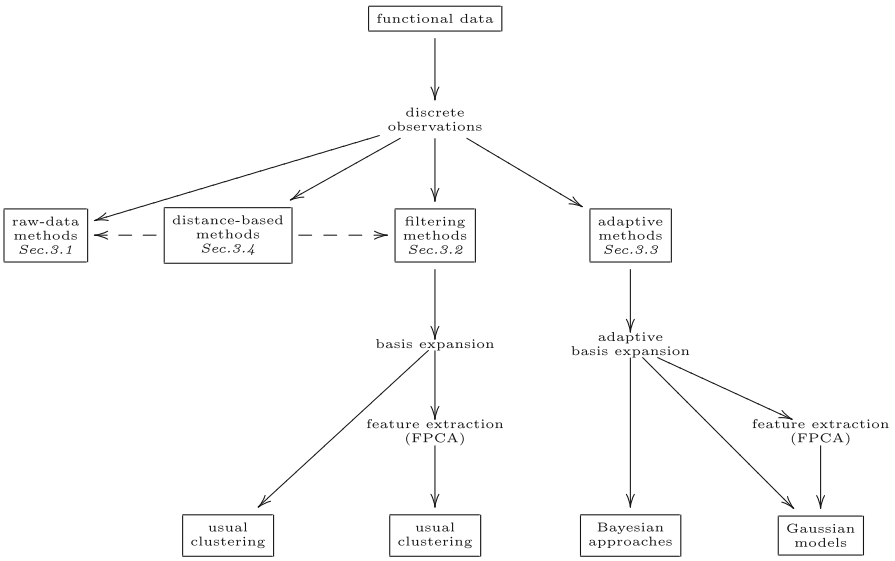

Fig. 2 Segmentation of the different clustering methods for functional data

data. The number of evaluation points being usually large, clustering techniques for high-dimensional vectorial data have to be used. These techniques are not discussed in this paper and we refer to Bouveyron and Brunet (2013) for a recent survey on this topic. Although leading sometimes to interesting results, the use of raw-data clustering does not take into account the functional feature of data (continuity, derivatives, etc). As an illustration, raw-data clustering is invariant to any permutation of the curves observation order. Extracting the true signal from noisy data and the use of some specific metric related to the functional feature of the data are important steps when dealing with functional data. In that sense, a loss of information can be expected. Moreover, the case of curves observed at different evaluation points can not be taken into account by such methods. For these reasons, we do not encourage the use of such methods.

In this category, we can also ranked the work of Boullé (2012), which proposes a clustering algorithm based on the probability distribution of the number of discrete observations of the curves in each interval of some partition of $\mathcal{T}$. If this clustering techniques is based on the discrete observation of the curves, it does not suffer from the latter limitation (curves can be sampled at different points), and has the advantage to do not depend on any assumption about the nature of the curves, since no basis expansion is used.

3.2 Filtering methods

Filtering methods consist of a first step, quoted as *filtering* step in James and Sugar (2003), in which the dimension of data is reduced. The curves are thus represented by a finite set of parameters (or coefficients). A second step of using clustering tools for finite dimensional data is then performed.

Table 1 Filtering methods for functional data clustering

References	Clustering method	Type of basis functions
Abraham et al. (2003)	k-means	B-spline
Rossi et al. (2004)	Self-organised map	B-spline
Peng and Müller (2008)	k-means	Eigenfunctions
Kayano et al. (2010)	Self-organised map	Gaussian

The filtering step consists usually in approximating the curves into a finite basis of functions. Spline basis is one of the most common choice because of their optimal properties (Wahba 1990). For instance, B-splines are considered in Abraham et al. (2003) and Rossi et al. (2004). Another popular dimension reduction technique is functional principal component analysis (FPCA, Sect. 2.2). As mentioned in Sect. 2.2, computing FPCA in practice needs usually to reconstruct the functional nature of the curves, by approximating them into a finite dimensional space of functions spanned by some basis of functions (see Appendix A).

Functional data being summarized either by their coefficients in a basis of functions or by a finite number of their principal component scores, usual clustering algorithms can be used to define clusters of functional data. In Abraham et al. (2003) and Peng and Müller (2008) the k-means algorithm is used on B-splines coefficients (Abraham et al. 2003) and on principal component scores (Peng and Müller 2008). In Peng and Müller (2008) the number of principal component scores is selected according to the percentage of explained variance, which is an usual criterion in principal component analysis. Let us also remark that in Peng and Müller (2008), the principal component scores are not directly used but transformed in a low-dimensional space thanks to a multi-dimensional scaling (Cox and Cox 2001). In Rossi et al. (2004) and Kayano et al. (2010) an unsupervised neural network, Self-Organised Map (Kohonen 1995), is applied respectively on B-spline and Gaussian coefficient's basis.

Table 1 summarizes these clustering methods based on a parametric representation of the data.

Other approaches, developed in more specific contexts, can be associated to this category. For instance, in the context of repeated measure, Serban and Jiang (2012) use a mixture of Gaussian distributions for the ANOVA's coefficients in the context of functional data.

3.3 Adaptive methods

In the filtering approaches, the curves are approximated in a finite basis of functions, and the curves are then identified to the basis expansion coefficients. In the present category of methods, instead of treating the basis expansion coefficients as parameters, they are considered as random variables having a cluster-specific probability distribution. Most of methods in this categoriy are based on a probabilistic modelling of either some basis expansion coefficients and or some FPCA scores. For this reason, a presentation of such probabilistic model-based techniques (summarized in Table 2) is given below, separating methods working with basis expansion coefficients from those working with FPCA scores.

Table 2 Adaptive methods for functional data clustering based on a modelling of basis expansion coefficients or FPCA scores

References	Model on	Type of model
James and Sugar (2003)	Basis expansion coef.	Gaussian (parsimonious)
Samé et al. (2011)	Basis expansion coef.	Gaussian with regime changes
Chiou and Li (2007)	FPCA scores	Gaussian spherical (k-means)
Bouveyron and Jacques (2011)	FPCA scores	Gaussian (parsimonious)
Jacques and Preda (2013a)	FPCA scores	Gaussian
Jacques and Preda (2013b)	FPCA scores	Gaussian
Heard et al. (2006)	Basis expansion coef.	Bayesian
Ray and Mallick (2006)	Basis expansion coef.	Bayesian
Giacofci et al. (2012)	Basis expansion coef.	Bayesian

We have to cite also two other adaptive methods, in which the representation of the curves is updated during the clustering process, which are not ranked in Table 2. The first one is the work of Hébrail et al. (2010), in which two dynamic programming algorithms which simultaneously perform clustering (using a k-means-like algorithm) and piecewise estimation of the cluster centres are proposed. In the second one (Yamamoto 2012) a new procedure to identify simultaneously optimal clusters of functions and optimal subspaces for clustering is proposed. For this purpose, an objective function is defined as the sum of the L_2-distances (d_0, Eq. 5) between the observations and their projections plus the L_2-distances between the projections and the cluster means (in the projection space). An alternating algorithm is used to optimize the objective function.

Adaptive clustering techniques using basis expansion coefficients modelling. In James and Sugar (2003), the authors consider that the basis expansion coefficients of the curves into a spline basis are distributed according to a mixture Gaussian distributions with means μ_k, specific to each cluster, and common variance Σ:

$$\alpha_i \sim \mathcal{N}(\mu_k, \Sigma).$$

Contrary to the filtering approaches, in which the basis expansion coefficients are considered fixed, they are considered as random variables, what allows inter alia to proceed efficiently with sparsely sampled curves. Parsimony assumptions on the cluster means μ_k allow to define parsimonious clustering models (quoted by *fclust*) and low-dimensional graphical representation of the curves.

The use of spline basis is convenient when the curves are regular, but are not appropriate for peak-like data as encountered in mass spectrometry for instance. For this reason, Giacofci et al. (2012) recently proposes a Gaussian model on a wavelet decomposition of the curves, which allows to deal with a wider range of functional shapes than splines.

An interesting approach has also been considered in Samé et al. (2011), by assuming that the curves arise from a mixture of regressions on a basis of polynomial functions, with possible changes in regime at each instant of observation. Thus, at each evaluation

point t_{ij}, the observation $X_i(t_{ij})$ is assumed to arise from one of the polynomial regression models specific to the cluster X_i belongs to.

Some Bayesian models have also been proposed. On the one hand, Heard et al. (2006) consider that the basis expansion coefficients are distributed as follows:

$$\alpha_i|\sigma_k \sim \mathcal{N}(\mu, \sigma_k \Sigma) \quad \text{and} \quad \sigma_k \sim \mathcal{IG}(u, v),$$

where $\mathcal{IG}$ is the Inverse-Gamma distribution. On the other hand, Ray and Mallick (2006) propose a hierarchical Bayesian model assuming further that Σ is modelled by two sets of random variables controlling the sparsity of the wavelets decomposition and a scale effect.

Adaptive clustering techniques using FPCA scores modelling. In Delaigle and Hall (2010), an approximation of the notion of probability density for functional random variables is proposed. This approximation is based on the truncation (4) of the Karhunen–Loeve expansion, and uses the density of principal components resulting from a FPCA of the curves. After an independence assumption on the principal components (which are uncorrelated), they consider a non-parametric kernel-based density estimation and use it to estimate the mean and the mode of some functional datasets. Using a similar approximation of the notion of density for functional random variables, Bouveyron and Jacques (2011) and Jacques and Preda (2013a) assume a Gaussian distribution of the principal components and define probabilistic model-based clustering techniques by the mean of the following mixture model (Jacques and Preda 2013a):

$$f_X^{(q)}(x;\theta) = \sum_{k=1}^{K} \pi_k \prod_{j=1}^{q_k} f_{C_{j|Z_k=1}}(c_{jk}(x);\lambda_{jk}),$$

where $\theta = (\pi_k, \lambda_{1k}, \ldots, \lambda_{q_k k})_{1\leq k\leq K}$ are the model parameters (cluster proportions and principal components variances) and q_k is the order of truncation of the Karhunen–Loeve expansion (3), specific to cluster k. The main interest of this model, called *funclust* by the authors, is that principal component scores $c_{jk}(x)$ of x are computed per cluster, by an EM-like algorithm, which iteratively computes the conditional probabilities of the curves to belong to each cluster, performs FPCA per cluster by weighting the curves according to these conditional probabilities, and computes the truncation orders q_k thanks to the scree-test of Cattell (1966). In Bouveyron and Jacques (2011), the q_k's are fixed to the maximum number of positive eigenvalues, L, which corresponds to the number of basis functions used in FPCA approximation (see Appendix A), and some parsimony assumptions on the variance λ_{jk} are considered to define a family of parsimonious sub-models, quoted as *FunHDDC* as an extension of the HDDC method for finite dimensional data (Bouveyron et al. 2007). The choice between these different sub-models is performed thanks to the BIC criterion (Schwarz 1978).

Previously to these works, Chiou and Li (2007) have considered a k-means algorithm based on a distance defined as the L_2 distance between truncations of the Karhunen–Loeve expansion at a given order q_k, named *k-centres*. This method is a particular case of Bouveyron and Jacques (2011) and Jacques and Preda (2013a),

Table 3 Distance-based clustering methods for functional data clustering

References	Clustering method	Proximity measure
Cuesta-Albertos and Fraiman (2000)	Trimmed k-means	d_0
Tarpey and Kinateder (2003)	k-means	d_0
Ferraty and Vieu (2006)	Hierarchical clustering	d_0, d_2
Tokushige et al. (2007)	k-means	d_0
Ieva et al. (2012)	k-means	d_0, d_1, $d_0 + d_1$

assuming that the variance of the principal components λ_{jk} are all equals (within and between clusters).

Notice finally, if the FPCA were not computed per cluster, such methods will correspond to Gaussian probabilistic model-based clustering applied on FPCA scores, and could be ranked in the category of filtering methods (Sect. 3.2).

Let emphasize once again that such methods are of particular interest, since assuming a probabilistic distribution on the basis expansion coefficients allow to take into account the uncertainty due to the functional data reconstruction.

3.4 Distance-based clustering methods

Several contributions to functional data clustering relies on the adaptation of popular geometric clustering algorithms, as k-means and hierarchical clustering, for functional data. For this, specific distances or dissimilarities between curves are considered. In this context, several works consider the following measures of proximity between two curves x_i and $x_{i'}$:

$$d_\ell(x_i, x_{i'}) = \left(\int_{\mathcal{T}} (x_i^{(\ell)}(t) - x_{i'}^{(\ell)}(t))^2 dt \right)^{1/2}. \tag{5}$$

where $x^{(\ell)}$ is the ℓth derivative of x. In Ferraty and Vieu (2006) the authors propose to use a hierarchical clustering algorithm combined with the distance d_0 (the L_2-metric) or with the semi-metric d_2. In Ieva et al. (2012) the k-means algorithm is used with d_0, d_1 and with $(d_0^2 + d_1^2)^{1/2}$. In Tarpey and Kinateder (2003) the authors investigate the use of d_0 with k-means for Gaussian processes. In particular, they prove that the cluster centres are linear combinations of FPCA eigenfunctions. The same distance d_0 with k-means is considered in Tokushige et al. (2007) defining time-dependent clustering. These methods are summarized in Table 3.

Let also mention the work of Cuesta-Albertos and Fraiman (2000), in which a robust version of the k-means algorithm using the distance d_0 is developed.

Connexions with other categories of methods. Depending on the method used to estimate the distance d_0, distance-based method can be assimilated to a raw-data method or to a filtering method. Indeed, if d_0 is approximated using directly the discrete observations of curves—using for instance the function *metric.lp()* of the *fda.usc* package for the **R** software—, the methods described in this section are equivalent to

raw-data clustering methods (Sect. 3.1). Similarly, if an approximation of the curves into a finite basis is used to approximate d_0—with function *semimetric.basis()* of *fda.usc*—, these methods are equivalent to filtering methods (Sect. 3.2).

4 Model selection problems

In addition to the well known problem of selecting the number of clusters, the choice of an appropriate basis for the curves approximation is a major question for a large number of functional data clustering methods. Indeed, this latter choice can be of important influence on the resulting clustering. Sections 4.1 and 4.2 summarized the model selection tools usually considered to answer these two questions in the functional data clustering context.

4.1 Choosing the approximation basis

Except the raw-data clustering methods, all the other clustering algorithms described in this survey needs the approximation of the curves into a finite dimensional basis of functions. The choice between different basis (trigonometric, spline,…) as well as the choice of the number of basis functions are of primary interest.

In Ramsay and Silverman (2005), the authors advise to choose the basis according to the nature of the functional data: for instance, Fourier basis can be suitable for periodic data, whereas spline basis is the most common choice for non-periodic functional data. Considering probabilistic model-based clustering algorithms allow to use less subjective criteria such as penalized likelihood criteria BIC (Schwarz 1978), AIC (Akaike 1974) or ICL (Biernacki et al. 2000). The reader can for instance to refer to Bouveyron and Jacques (2011), James and Sugar (2003), Samé et al. (2011), Serban and Jiang (2012) for such use.

This choice of an appropriate basis is also a well known question in the signal processing community, in signal compression applications for instance. In Coifman and Wickerhauser (1992), an entropy criterion on the basis expansion coefficients is used to define an algorithm leading to select the best basis functions into a given orthonormal basis. This criterion has been extended to curve classification in Saito and Coifman (1995).

4.2 Choosing the number of clusters

The choice of the number of clusters is a common problem to every clustering task. If classical likelihood-based model selection criteria as BIC, AIC or ICL are frequently used in the context of probabilistic model-based clustering for functional data (see for instance Bouveyron and Jacques 2011; Giacofci et al. 2012; Samé et al. 2011; Serban and Jiang 2012), more specific criteria have also been introduced.

First of all, Bayesian models for functional data clustering (Heard et al. 2006; Ray and Mallick 2006) define a framework in which the number of clusters can be directly estimated. For instance, Heard et al. (2006) considered a uniform prior over the range $\{1, \ldots, n\}$ for the number of clusters, which is then estimated when maximizing the posterior distribution.

More empirical criteria have also been used for functional data clustering. In Kayano et al. (2010), the clustering is repeated several times for each number of clusters and that leading to the highest stability of the partition is retained. Even more empirical and very sensitive, but also meaningful, Chiou and Li (2007) and Ieva et al. (2012) retain the number of clusters leading to a partition having the best physical interpretation.

In James and Sugar (2003), an original model selection criterion is considered. Proposed initially in Sugar and James (2003), this criterion is defined as the averaged Mahalanobis distance between the basis expansion coefficients α_i and their closest cluster centre. In Sugar and James (2003), it is shown for a large class of mixture distributions that this criterion choose the right number of clusters asymptotically with the dimension (here the number L of basis functions).

5 Software

While there are numerous software solutions for the finite dimensional case (refer to Bouveyron and Brunet 2013 for a recent survey), the clustering software dedicated to functional data are far fewer. We describe here the packages and toolboxes (Table 4) available under the MATLAB software (2010) and the **R** software (2012).

Filtering methods (Sect. 3.2) can be applied using:

- the *Curve Clustering Toolbox* (Gaffney 2004) for MATLAB, which implements a family of clustering algorithms based on Gaussian mixture combined with spline or polynomial basis approximation (can be ranked in the filtering methods),
- under **R**, the functions *kmeans* of the *stats* package for the algorithm of the same name or the packages *mclust*, *Rmixmod* and *HDclassif* for Gaussian mixture model-based clustering can be used. Each of these algorithm can be applied either on basis expansion coefficients or on FPCA scores which can both be computed thanks to the *fda* package.

All the other software solutions described below work under the **R** software.

Distance-based clustering algorithms (Sect. 3.4) can be applied using the *kmeans* or *hclust* (hierarchical clustering) functions of the *stats* package, combined with the d_ℓ distances (5) available from the *fda* or *fda.usc* packages.

Finally, specific packages (or functions) have been developed for some adaptive methods (Sect. 3.3):

Table 4 Packages for functional data clustering

References	Software	Package name
Gaffney (2004)	Matlab	Curve clustering toolbox
Filtering methods	**R**	fda + stats/mclust/ Rmixmod/HDclassif
Distance-based methods	**R**	fda/fda.usc + stats
Giacofci et al. (2012)	**R**	curvclust
Jacques and Preda (2013a)	**R**	Funclustering
Bouveyron and Jacques (2011)	**R**	*FunHDDC*[2]
James and Sugar (2003)	**R**	*fclust*[2]
Sangalli et al. (2010b)	**R**	fdakma

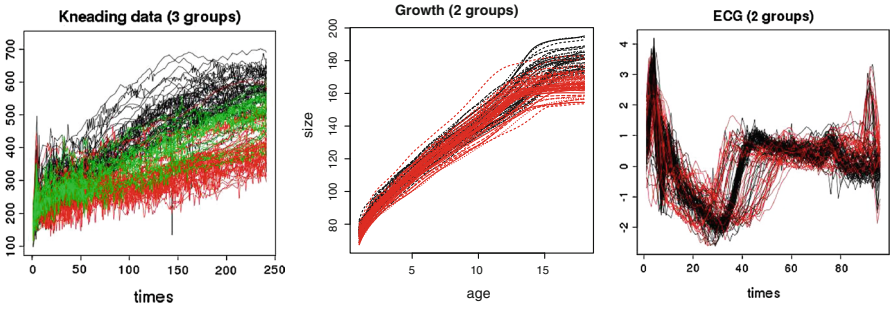

Fig. 3 *Kneading*, *Growth* and *ECG* datasets

- the package *curvclust* for **R** (Giacofci et al. 2012) implements the wavelets-based methods (Giacofci et al. 2012) described in Sect. 3.3,
- the package *Funclustering* for **R**, available on CRAN,[1] proposes clustering algorithm for functional data (potentially multivariate) based on a FPCA scores modelling (Jacques and Preda 2013a). Let notice that this package proposes also FPCA for multivariate functional data,
- the package *fdakma* for **R**, available on CRAN,[1] implements the algorithm of Sangalli et al. (2010b) which performs simultaneously curves alignment and clustering,
- function for FunHDDC (Bouveyron and Jacques 2011) is available from request upon to the authors of the method,
- function for fclust (James and Sugar 2003) is available directly from James's webpage.[2]

6 Numerical illustration

The evaluation of clustering algorithms is always a difficult task. For instance, techniques using simulations or classification datasets are both questionable (see for instance Guyon et al. (2009) for an interesting discussion on this topic). In this survey, we simply illustrate the clustering algorithms previously discussed on classification benchmark datasets. We emphasize that this numerical illustration must not be used to judge which clustering algorithm is the best. Moreover, this illustration will allow to show how to use the different **R** package for functional data clustering (**R** scripts are given in Appendix B).

6.1 The data

Three real datasets are considered: the *Kneading*, *Growth*, and *ECG* datasets. These three datasets are plotted in Fig. 3.

The Kneading dataset comes from Danone Vitapole Paris Research Center and concerns the quality of cookies and the relationship with the flour kneading process.

[1] http://cran.r-project.org/.

[2] author's function.

The kneading dataset is described in detail in Lévéder et al. (2004). There are 115 different flours for which the dough resistance is measured during the kneading process for 480 s. One obtains 115 kneading curves observed at 241 equispaced instants of time in the interval [0, 480]. The 115 flours produce cookies of different quality: 50 of them have produced cookies of *good* quality, 25 produced *medium* quality and 40 *low* quality. This data, have been already studied in a supervised classification context (Lévéder et al. 2004; Preda et al. 2007). They are known to be hard to discriminate, even for supervised classifiers, partly because of the medium quality class. Taking into account that the resistance of dough is a smooth curve but the observed one is measured with error, and following previous works on this data (Lévéder et al. 2004; Preda et al. 2007), least squares approximation on a basis of cubic B-spline functions (with 18 knots) is used to reconstruct the true functional form of each sample curve.

The Growth dataset comes from the Berkeley growth study (Tuddenham and Snyder 1954) and is available in the *fda* package of **R**. In this dataset, the heights of 54 girls and 39 boys were measured at 31 stages, from 1 to 18 years. The goal is to cluster the growth curves and to determine whether the resulting clusters reflect gender differences.

The ECG dataset is taken from the *UCR Time Series Classification and Clustering* website.[3] This dataset consists of 200 electrocardiograms from 2 groups of patients sampled at 96 time instants, and has already been studied in Olszewski (2001). For these two datasets, the same basis of functions as for the Kneading dataset has been arbitrarily chosen (20 cubic B-splines).

6.2 Experimental setup

All the clustering algorithms presented in Sect. 3 can not be tested since they are not all implemented in public software or package. In the present numerical experiments, only the clustering algorithms available as a package or function for the **R** software are considered:

- raw-data and filtering methods:
  - the clustering methods for finite dimensional data considered are k-means, hierarchical clustering, Gaussian Mixture Models (*mclust* package Banfield and Raftery 1993) and two methods dedicated to the clustering of high-dimensional data—*HDDC* (Bouveyron et al. 2007) and *MixtPPCA* (Tipping and Bishop 1999)—both available in the *HDclassif* package (Bergé et al. 2012),
  - these methods are applied: on the FPCA scores, with choice of the number of components thanks to the Cattell scree test (filtering); directly on discretizations of the curves at the observation points (raw-data); and on the coefficients in the cubic B-spline basis approximation (filtering),
- distance-based methods: k-means with distance d_0 and d_1 (Ieva et al. 2012),
- adaptive methods: funclust (Jacques and Preda 2013a), FunHDDC (Bouveyron and Jacques 2011), fclust (James and Sugar 2003), curvclust (Giacofci et al. 2012) and k-centres (Chiou and Li 2007) (results for the Growth dataset are available in their paper, but not software allow to proceed with the two other datasets).

[3] http://www.cs.ucr.edu/~eamonn/time_series_data/.

Table 5 Correct classification rates (CCR) in percentage for raw-data, adaptive and filtering methods on the kneading, growth and ECG datasets

	Kneading	Growth	ECG
Raw-data methods			
HDDC	66.09	56.99	74.5
MixtPPCA	65.22	62.36	74.5
mclust	63.48	65.59	81
k-means	62.61	65.59	74.5
hclust	63.48	51.61	73
Distance-based methods			
kmeans-d_0	62.61	64.52	74.5
kmeans-d_1	64.35	87.40	61.5
Filtering methods on spline coeff.			
HDDC	53.91	50.51	73.5
MixtPPCA	64.35	50.53	73.5
mclust	50.43	63.44	80.5
k-means	62.61	66.67	72.5
hclust	63.48	75.27	76.5
Filtering methods on FPCA scores			
HDDC	44.35	**97.85**	74.5
MixtPPCA	62.61	**97.85**	74.5
mclust	60	95.70	81.5
k-means	62.61	64.52	74.5
hclust	63.48	68.81	64
Adaptive methods			
funclust	**66.96**	69.89	**84**
FunHDDC	62.61	96.77	75
fclust	64	69.89	74.5
curvclust	65.21	67.74	74.5
k-centres	–	93.55	–

Bold values indicate the highest CCR for each dataset

The corresponding **R** codes are given in Appendix B.

The package *fdakma*, performing curves alignment additionally to clustering, is not considered since its clusters are not comparable with those of other methods (which do not perform curves alignment).

6.3 Clustering results and practical usefulness of the different clustering methods

The correct classification rates (CCR) according to the known partitions are given in Table 5. As previously discussed, these rates can not be used to judge which method is the best. They just indicates if the unsupervised clustering produced by the method is an adequacy with an external classification criteria (quality of the cookies for the Kneading dataset, gender for the Growth data). Nevertheless, these results can be used to illustrate some comments about the practical usefulness of the different clustering methods.

Curves representation The choice of the way the curves are represented is critical in a clustering context. For example, filtering methods can leads to very different results depending they are applied on the spline coefficients or on the FPCA scores. Indeed, for the Growth dataset, HDDC and MixtPPCA using the FPCA scores leads to a partition strongly related to the gender of children (CCR $= 97.85\,\%$), whereas the partition is very different if these methods are applied on the spline coefficients (CCR $\simeq 50.5\,\%$). A similar issue occurs for adaptive methods which need to choose an appropriate basis for the curve representation.

Probabilistic model-based clustering methods are very interesting in such situation, since they allow to use penalized likelihood criterion (BIC, AIC or ICL, see Sect. 4.1) to choose the most appropriate curves representation. Contrarily, distance-based methods can generally not use these criterion (except for k-means with the Euclidean distance which is equivalent to a particular Gaussian mixture model).

Characterisation of the clusters As mentioned in introduction, clustering is often used as a preliminary step for data exploration. It is of crucial interest to give an interpretation to the different clusters exhibited by the clustering analysis. In that sense, one usually uses statistics within clusters: mean, variance (matrix).

Methods using the FPCA scores provide additional visualizing tool: projection of the original curves into a subspace of low dimension (usually of dimension 2). The characterisation of the clusters by the mean functions and the principal components express different features of data. In Bouveyron and Jacques (2011), the authors perform clustering of the well-known Canadian weather dataset (temperature curves for Canadian weather stations Ramsay and Silverman 2005). After having noticed a geographical interpretation of the clusters (the four clusters correspond to continental, northern, Atlantic and Pacific stations), the interpretation of the principal components allows to exhibit some interesting feature of data: high variability within the cluster of Pacific stations, due to the presence of mountain stations ; time-shift effect within the cluster of continental station, due to the fact that some cities of this group have their seasons shifted, e.g. late entry and exit in winter; etc.

7 Conclusion and future challenge

This paper has presented a survey of the main existing algorithms for functional data clustering. A classification of these methods has been proposed, into four groups. The first group is raw-data methods working directly on the evaluation points of the curves, which has the disadvantage of not taking into account the functional nature of data. Indeed, the clustering is invariant to any permutation of the curves observations order. The second group is filtering methods, which first approximate the curves into a finite basis of functions and second perform clustering using the basis expansion coefficients. The third group is adaptive methods, which perform simultaneously dimensionality reduction of the curves and clustering. The main advantage of such approaches is that the dimensionality reduction, in other words how the curves are represented in a finite dimensional space, depends on clusters. Thus, two curves of two different clusters can be represented differently, what produce more flexible clustering tools. The last group is distance-based methods, which consists of using specific distances for

functional data with usual clustering algorithms. A software survey is proposed, and numerical experiments on three real datasets have provided the basis for a discussion on some practical usefulness of methods. In addition, the **R** codes used for the numerical experiments are given in the Appendix, what may help the reader in applying these clustering algorithms to his own data.

Future challenges We conclude this paper by given some future challenges related to functional data clustering.

We have seen in Sect. 4 that the choice of an appropriate basis for the curves approximation is a critical point, especially in a clustering context. Indeed, choosing a basis or another one can drastically change the clustering. The use of probabilistic model-based techniques allows to overcome this problem by the mean of likelihood-based model selection criteria. But for non probabilistic methods, this issue remains an open problem. A similar problem occur for the choice of an appropriate metric when using distance-based methods.

The clustering of spatial functional data is also a challenge because of the dependent nature of data. As an example, in Jacques and Preda (2013b), a clustering of the soil of Mars based on hyperspectral images is performed, in order to characterize the composition of the Mars surface by determining homogeneous zones with respect to the material. In this work, the spectrum corresponding to the pixels of the hyperspectral image are assumed to be independent, but it is evident that this is far from the truth. There is a need of clustering methods allowing to take into account dependent observations (α-mixing, auto-regressive models, random fields, etc). Some early answers can be provided by the recent works of Romano et al. (2011) and Secchi et al. (2011).

Another interesting issue is the categorical functional data (Boumaza 1980), in which X_t is a categorical random variable for each $t \in \mathcal{T}$. The marital status of individuals, the status of some patients with respect to some diseases are some examples of such data. This type of data has been already studied in the longitudinal data framework, where the curves are defined on a discrete space $\mathcal{T}$ (refer for instance to the *TraMineR* package for **R** and to the related publications). Despite of the evident interest of such data, we do not found any recent work on categorical functional data, in particular for clustering.

Appendix A: computational methods for FPCA

Let $\{x_1, \ldots, x_n\}$ be the observation of the sample $\{X_1, \ldots, X_n\}$. The estimators for $\mu(t)$ and $V(s,t)$, for $s, t \in \mathcal{T}$, are:

$$\hat{\mu}(t) = \frac{1}{n}\sum_{i=1}^{n} x_i(t) \quad \text{and} \quad \hat{V}(s,t) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i(s) - \hat{\mu}(s))(x_i(t) - \hat{\mu}(t)).$$

In Deville (1974) it has been shown that $\hat{\mu}$ and $\hat{V}$ converges to μ and V in L_2-norm with convergences rate of $O(n^{-1/2})$.

As previously discussed, the functional data are usually observed at discrete evaluation points and a common solution to reconstruct the functional form of data is to assume that functional data belong to a finite dimensional space spanned by some

basis of functions. Let $\alpha_i = (\alpha_{i1}, \ldots, \alpha_{iL})'$ be the basis expansion coefficient of the observed curve x_i in the basis $\Phi = \{\phi_1, \ldots, \phi_L\}$, such that:

$$x_i(t) = \Phi(t)'\alpha_i$$

with $\Phi(t) = (\phi_1(t), \ldots, \phi_L(t))'$.

Let $\tilde{A}$ be the $n \times L$-matrix, whose rows are the vectors α_i', and $M(t) = (x_1(t), \ldots, x_n(t))'$ the vector of the values $x_i(t)$ of functions x_i at $t \in \mathcal{T}$ $(1 \leq i \leq n)$. With these notations, we have

$$M(t) = \tilde{A}\Phi(t). \tag{6}$$

Under the basis expansion assumption (1), the estimator $\hat{V}$ of V, for all $s, t \in \mathcal{T}$, is given by:

$$\hat{V}(s,t) = \frac{1}{n-1}(M(s) - \hat{\mu}(s))'(M(t) - \hat{\mu}(t)) = \frac{1}{n-1}\Phi(s)'A'A\Phi(t), \tag{7}$$

where $M(s) - \hat{\mu}(s)$ means that the scalar $\hat{\mu}(s)$ is subtracted to each elements of $M(s)$, and $A = (I_n - \mathbb{1}_n(1/n, \ldots, 1/n))\tilde{A}$ where I_n and $\mathbb{1}$ are respectively the identity $n \times n$-matrix and the unit column vector of size n.

From (2) and (7), each eigen-function f_j belongs to the linear space spanned by the basis Φ:

$$f_j(t) = \Phi(t)'b_j \tag{8}$$

with $b_j = (b_{j1}, \ldots, b_{jL})'$.

Using the estimation $\hat{V}$ of V, the eigen problem (2) becomes

$$\int_0^T \hat{V}(s,t) f_j(t) dt = \lambda_j f_j(s),$$

which, by replacing $\hat{V}(s,t)$ and $f_j(s)$ by their expressions given in (7) and (8), is equivalent to

$$\frac{1}{n-1}\Phi(s)'A'A \underbrace{\int_0^T \Phi(t)\Phi(t)'dt}_{W} b_j = \lambda_j \Phi(s)'b_j, \tag{9}$$

where $W = \int_0^T \Phi(t)\Phi(t)'dt$ is the symmetric $L \times L$ matrix of the inner products between the basis functions.

Since (9) is true for all s, we have:

$$\frac{1}{n-1}A'AWb_j = \lambda_j b_j.$$

By defining $u_j = W^{1/2}b_j$, the multivariate functional principal component analysis is reduced to the usual PCA of the matrix $\frac{1}{\sqrt{n-1}}AW^{1/2}$:

$$\frac{1}{n-1}W^{1/2'}A'AW^{1/2}u_j = \lambda_j u_j.$$

The coefficient $b_j,\ j \geq 1$, of the eigen-function f_j are obtained by $b_j = (W^{1/2})^{-1}u_j$, and the principal component scores, are given by

$$C_j = AWb_j \quad j \geq 1.$$

Note that the principal components scores C_j are also the solutions of the eigenvalues problem:

$$\frac{1}{n-1}AWA'C_j = \lambda_j C_j.$$

Appendix B: R codes for curve clustering

In this Appendix are given the **R** codes used to perform functional data clustering on growth dataset.

B.1 Data loading

First, the values of the functional data at the evaluation points are loaded in the matrix `data`, and the true label in the vector `cls`:

```
> library(fda)
> data=cbind(matrix(growth$hgtm,31,39),matrix(growth
   $hgtf,31,54))
> cls=c(rep(1,39),rep(2,54))
```

The functional form is reconstructed using spline basis (for FPCA-based methods), and stored in an object of the class `fd` of the fda package:

```
> t=growth$age
> splines <- create.bspline.basis(rangeval=c(1, max(t)),
nbasis = 20,norder=4)
> fdata <- Data2fd(data, argvals=t, basisobj=splines)
```

The number of clusters is 2 for this dataset:

```
> K=2
```

B.2 Clustering with funclust

These method is available in the *Funclustering* package for **R**. The main funclust of this package can be applied directly on the `fd` object `fdata`:
```
> library('Funclustering')
> res=funclust(fd,K=K)
```
The resulting clusters are stored in `res$cls`.

B.3 Clustering with FunHDDC

The corresponding computer code are available from request to their authors Bouveyron and Jacques (2011).

FunHDDC can be applied directly on the `fd` object `fdata`:
```
> res=fun_hddc(fd,K=K,model='AkjBkQkDk')
```

FunHDDC proposing several sub-models, each of one have to be tested –`'AkjBkQkDk'`, `'AkjBQkDk'`, `'AkBkQkDk'`, `'AkBQkDk'`, `'ABkQkDk'`, `'ABQkDk'`–, and the one leading to the highest BIC criterion is retained (available from `res$bic`). The clusters are stored in `res$cls`.

B.4 Clustering with fclust

The corresponding computer code is available from James's webpage.

First, the data have to be stored in a list as follows:
```
> nr=nrow(data)
> N = ncol(data)
> fdat = list()
> fdat$x = as.vector(data)
> fdat$curve = rep(1:N,rep(nr,N))
> fdat$timeindex = rep(as.matrix(seq(1,nr,1)),N)
> grid = seq(1, nr, length = nr)
```

And then, the clustering can be estimated by:
```
> testfit=fitfclust(data=fdat,grid=grid,K=K)
```
the cluster being available from `fclust.pred(testfit)$class`

B.5 Clustering with curvclust

First, the values of functional data discretization are registered in a list `Y`, then transformed in an object of the class `CClustData`:
```
> library('curvclust')
> fdat= list()
> for (j in 1:ncol(data)) fdat[[j]] =data[,j]
> CCD = new("CClustData",Y=fdat,filter.number=1)
```

Dimension reduction is then performed:

```
> CCDred = getUnionCoef(CCD)
```

The number of clusters is specified in the class `CClustO`:

```
> CCO = new("CClustO")
> CCO["nbclust"] = K
> CCO["Gamma2.structure"] = "none"
```

and clustering is performed thanks to the function `getFCM`:

```
> CCR = getFCM(CCDred,CCO)
> summary(CCR)
```

The cluster are finally estimated by maximum a posteriori:

```
> cluster = apply(CCR["Tau"],1,which.max)
```

References

Abraham C, Cornillon PA, Matzner-Løber E, Molinari N (2003) Unsupervised curve clustering using B-splines. Scand J Stat Theory Appl 30(3):581–595. doi:10.1111/1467-9469.00350

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19:716–723 (system identification and time-series analysis)

Antoniadis A, Beder JH (1989) Joint estimation of the mean and the covariance of a Banach valued Gaussian vector. Statistics 20(1):77–93

Banfield J, Raftery A (1993) Model-based Gaussian and non-Gaussian clustering. Biometrics 49:803–821

Bergé L, Bouveyron C, Girard S (2012) HDclassif : an R package for model-based clustering and discriminant analysis of high-dimensional data. J Stat Softw 42(6):1–29

Besse P (1979) Etude descriptive d'un processus. Thèse de doctorat 3$^{\text{ème}}$ cycle Université Paul Sabatier, Toulouse

Biernacki C, Celeux G, Govaert G (2000) Assessing a mixture model for clustering with the inegrated completed likelihood. IEEE Trans Pattern Anal Mach Intell 22(4):719–725

Bosq D (2000) Linear processes in function spaces, Lecture Notes in Statistics, vol 149. Springer, New York (theory and applications)

Boullé M (2012) Functional data clustering via piecewise constant nonparametric density estimation. Pattern Recognit 45(12):4389–4401

Boumaza R (1980) Contribution a l'étude descriptive d'une fonction aléatoire qualitative. PhD thesis, Université Paul Sabatier, Toulouse, France

Bouveyron C, Brunet C (2013) Model-based clustering of high-dimensional data : a review. Technical report

Bouveyron C, Jacques J (2011) Model-based clustering of time series in group-specific functional subspaces. Adv Data Anal Classif 5(4):281–300

Bouveyron C, Girard S, Schmid C (2007) High dimensional data clustering. Comput Stat Data Anal 52: 502–519

Cardot H, Ferraty F, Sarda P (1999) Functional linear model. Stat Probab Lett 45:11–22

Cattell R (1966) The scree test for the number of factors. Multivar Behav Res 1(2):245–276

Celeux G, Govaert G (1995) Gaussian parsimonious clustering models. J Pattern Recognit Soc 28:781–793

Chiou JM, Li PL (2007) Functional clustering and identifying substructures of longitudinal data. J R Stat Soc Ser B Stat Methodol 69(4):679–699. doi:10.1111/j.1467-9868.2007.00605.x

Coifman R, Wickerhauser M (1992) Entropy-based algorithms for best basis selection. IEEE Trans Inf Theory 38(2):713–718

Cox T, Cox M (2001) Multidimensional scaling. Chapman and Hall, New York

Cuesta-Albertos J, Fraiman R (2000) Impartial trimmed k-means for functional data. Comput Stat Data Anal 51:4864–4877

Dauxois J, Pousse A, Romain Y (1982) Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. J Multivar Anal 12(1):136–154. doi:10.1016/0047-259X(82)90088-4

Delaigle A, Hall P (2010) Defining probability density for a distribution of random functions. Ann Stat 38:1171–1193

Deville J (1974) Méthodes statistiques et numériques de l'analyse harmonique. Annales de l'INSEE 15:3–101
Escabias M, Aguilera A, Valderrama M (2005) Modeling environmental data by functional principal component logistic regression. Environmetrics 16:95–107
Ferraty F, Vieu P (2006) Nonparametric functional data analysis, Springer Series in Statistics. Springer, New York
Gaffney S (2004) Probabilistic curve-aligned clustering and prediction with mixture models. PhD thesis, Department of Computer Science, University of California, Irvine, USA
Giacofci M, Lambert-Lacroix S, Marot G, Picard F (2012) Wavelet-based clustering for mixed-effects functional models in high dimension. Biometrics (in press)
Guyon I, Von Luxburg U, Williamson R (2009) Clustering: science or art. In: NIPS 2009 workshop on clustering theory
Hartigan J, Wong M (1978) Algorithm as 1326: a k-means clustering algorithm. Appl Stat 28:100–108
Heard N, Holmes C, Stephens D (2006) A quantitative study of gene regulation involved in the immune response of anopheline mosquitoes: an application of Bayesian hierarchical clustering of curves. J Am Stat Assoc 101(473):18–29. doi:10.1198/016214505000000187
Hébrail G, Hugueney B, Lechevallier Y, Rossi F (2010) Exploratory analysis of functional data via clustering and optimal segmentation. Neurocomput EEG Neurocomput 73(7–9):1125–1141
Ieva F, Paganoni A, Pigoli D, Vitelli V (2012) Multivariate functional clustering for the analysis of ecg curves morphology. J R Stat Soc Ser C Appl Stat (in press)
Jacques J, Preda C (2013a) Funclust: a curves clustering method using functional random variable density approximation. Neurocomputing. doi:10.1016/j.neucom.2012.11.042
Jacques J, Preda C (2013b) Model-based clustering for multivariate functional data. Comput Stat Data Anal. doi:10.1016/j.csda.2012.12.004
James G, Sugar C (2003) Clustering for sparsely sampled functional data. J Am Stat Assoc 98(462):397–408
Karhunen K (1947) Über lineare Methoden in der Wahrscheinlichkeitsrechnung. Ann Acad Sci Fennicae Ser A I Math-Phys 1947(37):79
Kayano M, Dozono K, Konishi S (2010) Functional cluster analysis via orthonormalized gaussian basis expansions and its application. J Classif 27:211–230
Kohonen T (1995) Self-organizing maps. Springer, New York
Lévéder C, Abraham P, Cornillon E, Matzner-Lober E, Molinari N (2004) Discrimination de courbes de prétrissage. In: Chimiométrie 2004, Paris, pp 37–43
Liu X, Yang M (2009) Simultaneous curve registration and clustering for functional data. Comput Stat Data Anal 53:1361–1376
Loève M (1945) Fonctions aléatoires de second ordre. C R Acad Sci Paris 220:469
MATLAB (2010) version 7.10.0 (R2010a) The MathWorks Inc., Natick, Massachusetts
McLachlan G, Peel D (2000) Finite mixture models. Wiley Series in Probability and Statistics. Applied Probability and Statistics, Wiley-Interscience, New York. doi:10.1002/0471721182
Olszewski R (2001) Generalized feature extraction for structural pattern recognition in time-series data. PhD thesis, Carnegie Mellon University, Pittsburgh, PA
Peng J, Müller HG (2008) Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions. Ann Appl Stat 2(3):1056–1077. doi:10.1214/08-AOAS172
Preda C, Saporta G, Lévéder C (2007) PLS classification of functional data. Comput Stat 22(2):223–235
R Core Team (2012) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/. ISBN: 3-900051-07-0
Ramsay JO, Silverman BW (2002) Applied functional data analysis. Springer Series in Statistics. Springer, New York (methods and case studies)
Ramsay JO, Silverman BW (2005) Functional data analysis, 2nd edn. Springer Series in Statistics. Springer, New York
Ray S, Mallick B (2006) Functional clustering by Bayesian wavelet methods. J R Stat Soc Ser B Stat Methodol 68(2):305–332. doi:10.1111/j.1467-9868.2006.00545.x
Romano E, Giraldo R, Mateu J (2011) Recent advances in functional data analysis and related topics, Springer, chap clustering spatially correlated functional data
Rossi F, Conan-Guez B, El Golli A (2004) Clustering functional data with the som algorithm. In: Proceedings of ESANN 2004. Bruges, Belgium, pp 305–312
Saito N, Coifman R (1995) Local discriminant bases and thier applications. J Math Imaging Vis 5(4):337–358

Samé A, Chamroukhi F, Govaert G, Aknin P (2011) Model-based clustering and segmentation of times series with changes in regime. Adv Data Anal Classif 5(4):301–322
Sangalli L, Secchi P, Vantini S, Vitelli V (2010a) Functional clustering and alignment methods with applications. Commun App Ind Math 1(1):205–224
Sangalli L, Secchi P, Vantini S, Vitelli V (2010b) k-mean alignment for curve clustering. Comput Stat Data Anal 54(5):1219–1233
Saporta G (1981) Méthodes exploratoires d'analyse de données temporelles. Cahiers du BURO 37–38
Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6(2):461–464
Secchi P, Vantini S, Vitelli V (2011) Recent advances in functional data analysis and related topics, Springer, chap Spatial Clustering of Functional Data
Serban N, Jiang H (2012) Multilevel functional clustering analysis. Biometrics 68(3):805–814
Slaets L, Claeskens G, Hubert M (2012) Phase and amplitude-based clustering for functional data. Comput Stat Data Anal 56(7):2360–2374
Sugar C, James G (2003) Finding the number of clusters in a dataset: an information-theoretic approach. J Am Stat Assoc 98(463):750–763
Tarpey T, Kinateder K (2003) Clustering functional data. J Classif 20(1):93–114
Tipping ME, Bishop C (1999) Mixtures of principal component analyzers. Neural Comput 11(2):443–482
Tokushige S, Yadohisa H, Inada K (2007) Crisp and fuzzy k-means clustering algorithms for multivariate functional data. Comput Stat 22:1–16
Tuddenham R, Snyder M (1954) Physical growth of california boys and girls from birth to eighteen years. Univ Calif Public Child Dev 1:188–364
Wahba G (1990) Spline models for observational data. SIAM, Philadelphia
Ward J, Joe H (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58:236–244
Yamamoto M (2012) Clustering of functional data in a low-dimensional subspace. Adv Data Anal Classif 6:219–247