Functional data clustering: a survey

Julien Jacques, Cristian Preda

Functional data clustering: a survey

Julien JACQUES, Cristian PREDA

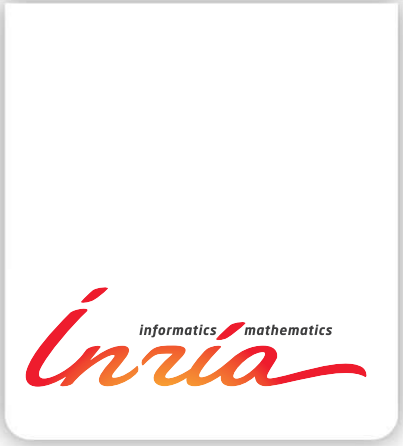
RESEARCH REPORT

N° 8198

January 2013

Project-Team MΘDAL

ISSN 0249-6399 ISRN INRIA/RR--8198--FR+ENG



Functional data clustering: a survey

Julien JACQUES*, Cristian PREDA†

Project-Team MΘDAL

Research Report n° 8198 — January 2013 — 24 pages

Abstract: The main contributions to functional data clustering are reviewed. Most approaches used for clustering functional data are based on the following three methodologies: dimension reduction before clustering, nonparametric methods using specific distances or dissimilarities between curves and model-based clustering methods. These latter assume a probabilistic distribution on either the principal components or coefficients of functional data expansion into a finite dimensional basis of functions. Numerical illustrations as well as a software review are presented.

Key-words: Functional data, Nonparametric clustering, Model-based clustering, Functional principal component analysis

* julien.jacques@polytech-lille.fr
† cristian.preda@polytech-lille.fr

RESEARCH CENTRE
LILLE – NORD EUROPE

Parc scientifique de la Haute-Borne
40 avenue Halley - Bât A - Park Plaza
59650 Villeneuve d'Ascq

Une revue des méthodes de classification automatique de données fonctionnelles

Résumé : Nous présentons dans cet article une revue des méthodes de classification automatique pour données fonctionelles. Ces techniques peuvent être classées en trois catégories: les méthodes procédant à une étape de réduction de dimension avant la classification, les méthodes non paramétriques qui utilisent des techniques de classification automatique classiques couplées à des distances ou dissimilarités spécifiques aux données fonctionnelles, et enfin, les techniques à base de modèles génératifs. Ces dernières supposent un modèle probabiliste soit sur les scores d'une analyse en composantes principales fonctionnelle, soit sur les coefficients des approximations des courbes dans une base de fonctions de dimension finie. Une illustration numérique ainsi qu'une revue des logiciels disponibles sont également présentées.

Mots-clés : Données fonctionnelles, Classification automatique, Méthodes non paramétriques, Modèles génératifs, Analyse en composantes principales fonctionnelles

1 Introduction

The aim of the cluster analysis is to build 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 identify particular patterns 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, such as hierarchical clustering [56] or k-means algorithm [24] are based on heuristic and geometric procedures. More recently, probabilistic approaches have been introduced to characterize the notion of cluster through their probability density [4, 13, 37].

In recent years, researchers concentrated their efforts to solve problems (regression, clustering) when p is large, in his absolute value or with respect to the size of some sample drawn from the distribution of X. The curse of dimensionality was and is still a very active topic. A particular case is that of random variables taking values into an infinite dimensional space, typically a space of functions defined on some continous set $\mathcal{T}$. Thus, 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 longtime inaccessible for statistics (because of technological limitations), today it becomes more and more easy to observe, to store and to process large amounts of such data in medicine, economics, chemometrics and many others domains (see [42] for an overview).

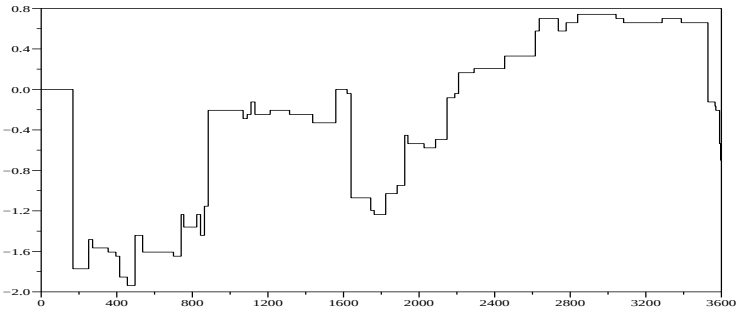
Clustering functional data is generally a difficult task because of the infinite dimensional space that data belong to. The lack of a definition for the probability density of a functional random variable, the defintion of distances or estimation from noisy data are some examples of such difficulties. Different approaches have been proposed along the years. The most popular approach consists of reducing the infinite dimensional problem to a finite one by approximating data with elements from some finite dimensional space. Then, clustering algorithms for finite dimensional data can be performed. On the other hand, nonparametric methods for clustering consist generally in defining specific distances or dissimilarities for functional data and then apply clustering algorithms as hierarchical clustering or k-means. Recently, model-based algorithms for functional data have been developed.

The aim of this paper is to propose a review of these 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 reviews the different clustering methods for functional data : two-stage methods which reduce the dimension before clustering, nonparametric methods and model-based methods. Section 4 discusses the common problems of selecting the number of clusters and of choosing appropriate representation for functional data approximation. Section 5 presents some software for clustering functional data. The numerical results of the application of some reviewed methods on real data are presented in Section 6. Some open problems related to functional data clustering end the paper.

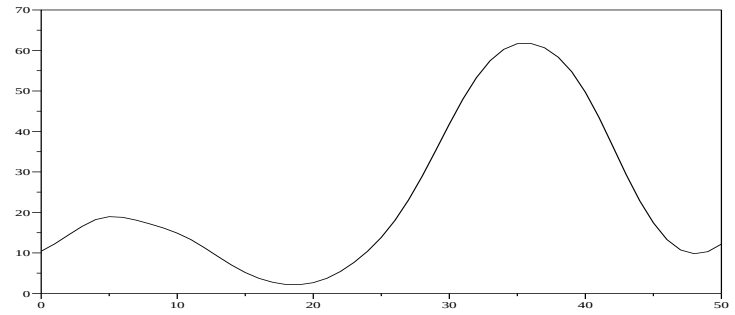
2 Functional Data Analysis

Functional data analysis (FDA) extends the classical multivariate methods when data are functions or curves. Some examples of such data are presented in Figure 1: the top Figure (a) plots the evolution of some stock-exchange index is observed during one hour; 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 [31, 36]. A pioneer work paper on the subject is due to Deville [18] – a one hunderd pages



(a) Share index evolution during one hour.



(b) Knee flexion angle (degree) over a complete gait cycle.

Figure 1: Some examples of functional data.

paper in the Annales de l'INSEE – with some applications in economy. In [16] and [3] the authors obtained asymptotic results for the elements derived from factorial analysis. The contributions of Besse [5] and of Saporta [47] extends 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. An important contribution to functional categorical data is due to [8].

More recently, important contributions to regression models for functional data are due to the research group working on functional statistics in Toulouse (STAPH[1]). Let us remind also the monographs on functional data by Ramsay and Silverman [41, 42] developing theory and applications on functional data, the book of Bosq [7] for modeling dependent functional random variables and the recent book of Ferraty and Vieu [20] on nonparametric models for functional data containing a review of the most recent contributions on this topic.

2.1 Functional Data

According to [20], 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.

[1] http://univ-tlse3.fr/STAPH

The underlying model for X_i's is generally 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 a Hilbert space H of functions defined on some set $\mathcal{T}$. Generally, $\mathcal{T}$ represents an interval of time, of wavelengths or any other continuous 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 [29] for a recent work on multivariate functional data clustering.

The main source of difficulty when dealing with functional data consists in 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 time-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, [42]). An alternative way of solving this problem is based on nonparametric smoothing of functions [20].

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

$$X_i(t) = \sum_{\ell=1}^{L} \alpha_{i\ell}\phi_\ell(t) \tag{2.1}$$

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

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

- If the sample curves are observed without error

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

  an interpolation procedure can be used. For example, [19] 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

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

  least squares smoothing is used after choosing a suitable basis as, for example, trigonometric functions, B-splines or wavelets (see [42] for a detailed study). In this case, the basis coefficients of each sample path $X_i(t)$ are approximated by

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

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

2.2 Functional Principal Component Analysis

From the set of functional data $\{X_1, \ldots, X_n\}$, one can be interested in optimal representation of curves into a function space of reduced dimension. The main tool to answer this request, Functional Principal Component Analysis (FPCA), is presented in this section. Althought the

practical interest of FPCA for interpretation and data presentation (graphics), it is one of the main tools considered when clustering functional data.
In order to address this question in a formal way, we need the hypothesis that considers X as a L_2-continuous stochastic process:

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

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

$$\begin{array}{rcl} \mathcal{V} : L_2(\mathcal{T}) & \to & L_2(\mathcal{T}) \\ f & \stackrel{\mathcal{V}}{\longmapsto} & \mathcal{V}f = \displaystyle\int_0^T V(\cdot, t) f(t) dt, \end{array}$$

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 $\{\lambda_j\}_{j \geq 1}$ associated to an orthonormal basis of eigenfunctions $\{f_j\}_{j \geq 1}$:

$$\mathcal{V} f_j = \lambda_j f_j, \tag{2.2}$$

with $\lambda_1 \geq \lambda_2 \geq \ldots$ and $\displaystyle\int_0^T f_j(t) f_{j'}(t) dt = 1$ 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 λ_j, $j \geq 1$.
With these definitions, the Karhunen-Loeve expansion [31, 36] holds:

$$X(t) = \mu(t) + \sum_{j \geq 1} C_j f_j(t), \qquad t \in \mathcal{T}. \tag{2.3}$$

Truncating (2.3) at the first q terms one obtains the best approximation in norm L_2 of $X(t)$ by a sum of quasi-deterministic processes [47] ,

$$X^{(q)}(t) = \mu(t) + \sum_{j=1}^{q} C_j f_j(t), \qquad t \in \mathcal{T}. \tag{2.4}$$

2.3 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 [18] 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 generally observed at discrete time 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 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 times $t \in \mathcal{T}$ $(1 \leq i \leq n)$. With these notations, we have

$$M(t) = \tilde{A}\Phi(t). \tag{2.5}$$

Under the basis expansion assumption (2.1), the estimator $\hat{V}$ of V, for all $s, t \in \mathcal{T}$, is given by:

$$\hat{V}(s,t) = \frac{1}{n-1}(M(s) - \hat{\mu}(s))'(M(t) - \hat{\mu}(t)) = \frac{1}{n-1}\Phi(s)'A'A\Phi(t), \tag{2.6}$$

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 $\mathbb{1}_n$ are respectively the identity $n \times n$-matrix and the unit column vector of size n.

From (2.2) and (2.6), each eigen-function f_j belongs to the linear space spanned by the basis Φ:

$$f_j(t) = \Phi(t)'b_j \tag{2.7}$$

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

Using the estimation $\hat{V}$ of V, the eigen problem (2.2) 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 (2.6) and (2.7), is equivalent to

$$\frac{1}{n-1}\Phi(s)'A'A\underbrace{\int_0^T \Phi(t)\Phi(t)'dt}_{W} b_j = \lambda_j \Phi(s)'b_j, \tag{2.8}$$

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 (2.8) 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 $\frac{1}{\sqrt{n-1}} AW^{1/2}$:

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

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:

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

2.4 Preprocessing functional data

Curves are generally observed at discrete instants of time. 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, generally, data registration [42, chap. 7]. It consists in centring and scaling the curves in order to eliminate both phase and amplitude variations into the curve's dataset. But, in our opinion, for clustering purpose registration is not necessarily. Indeed, the amplitude and phase variability of curves can be interesting elements to define clusters. For instance, in the well-known Canadian weather dataset (temperature and precipitation curves for Canadian weather stations [10, 28, 42]), the geographical interpretation of the clusters of weather stations is mainly due to amplitude variability. Nevertheless, several works perform curves registration before or simultaneously with clustering [35, 46] aiming to obtain new clusters which are not related to phase and amplitude variations. But, in this tentative, the conclusion is often the absence of cluster after registration. For instance, the Growth dataset [14, 28, 42, 54], which consists of growth curves for girls and boys, is considered by [35] for a clustering study simultaneously with data registration. The result being the absence of cluster, they failed in retrieving the gender of subjects, contrary to other methods [14, 28] which does not perform data registration.

3 Major functional data clustering approaches

Clustering functional data received particular attention from statisticians in the last decade. We present in this section a classification of the different approaches to functional data clustering into three groups. This classification, illustrated in Figure 2, is described below.
A first approach, quoted as *raw-data clustering* in Figure 2, consists in using directly discretization of the functions at some time points. This approach is the most simple one, since the functions are generally already observed at discrete instants of time. In this situation, there is no need to reconstruct the functional form of the data. Because of the large size of the discretization, clustering techniques for high-dimensional vectorial data must be used. These techniques are not discussed in this paper, and we refer to [11] for a complete review on the subject.
Thus, the first category of methods discussed in the sequel (Section 3.1) is *two-stage* methods, which first reduce the dimension of the data and second perform clustering.
The second category concerns nonparametric clustering methods and will be reviewed in Section 3.2. These methods consist generally in using specific distances or dissimilarities between curves

combined to classical non-probabilistic clustering algorithms designed for finite-dimensional data. The third category is model-based clustering techniques which assume a probability distribution underlying data. For functional data the notion of probability density generally does not exist [17]. Therefore, one can consider models involving probability density for some finite dimensional coefficients describing data. These coefficients can be either coefficients of curves into a basis approximation (splines, wavelets...) or principal components scores resulting from functional principal component analysis of the curves. These methods will be presented in Section 3.3.

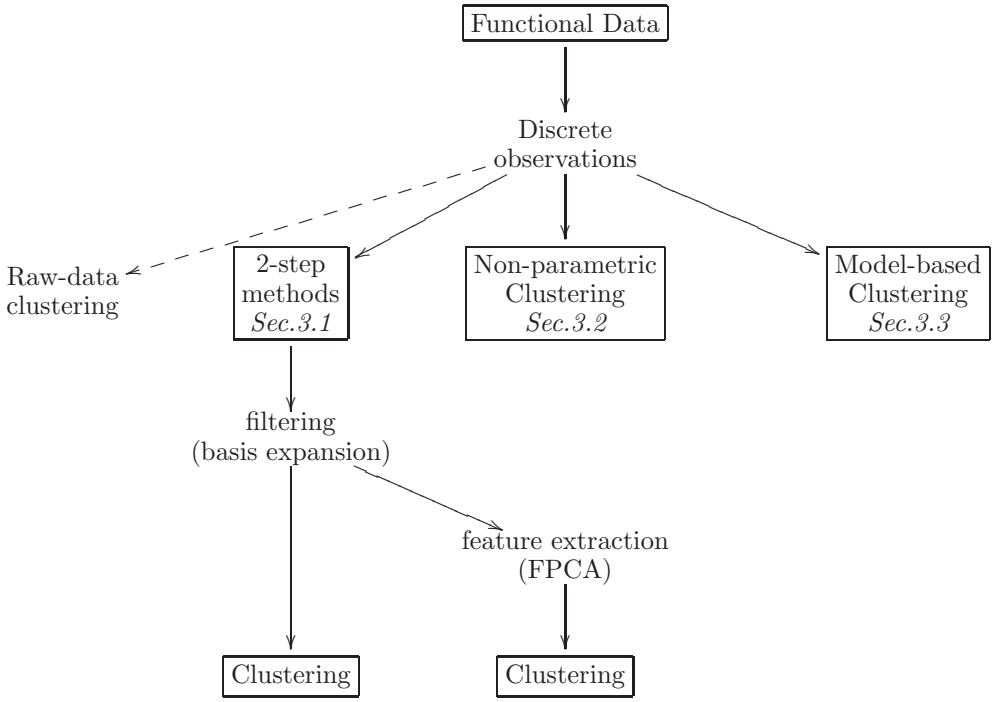


Figure 2: Classification of the different clustering methods for functional data.

3.1 Two-stage approaches

The two-stage approaches for functional data clustering consist of a first step, quoted as *filtering* step in [30], in which the dimension of data is reduced, and of a second step in which classical clustering tools for finite dimensional data are used.
The reducing dimension step consists generally in approximating the curves into a finite basis of functions. Spline basis [55] is one of the most common choice because of their optimal properties. For instance, B-splines are considered in [1, 44]. Another dimension reduction technique is the functional principal component analysis (see Section 2.2), for which, from a computational point of view, one needs generally to use also a basis approximation of the curves (see Section 2.3).
Functional data being summarized either by their coefficients in a basis of functions or by their first principal component scores, usual clustering algorithms can be used to estimate clusters of functional data. In [1] and [39] the k-means algorithm is used on B-splines coefficients [1] and on

a given number of principal component scores [39]. In [39] 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 remark also that in [39], the principal component scores are not directly used but transformed in a low-dimensional space thanks to a multi-dimensional scaling [15]. In [44] and [32] an unsupervised neural network, Self-Organised Map [33], is applied respectively on B-spline and Gaussian coefficient's basis.
Table 1 summarizes these two-stage approaches.
Let remark that there exist several other approaches developed in specific context. For instance, [49] decomposes a dataset of curves using a functional analysis of variance (ANOVA) model: taking into account repeated random functions the authors propose a clustering algorithm assuming a mixture of Gaussian distributions on the coefficients of the ANOVA model.

	type of basis functions		
clustering	B-spline	Gaussian	eigenfunctions
k-means	[1]		[39]
Self-Organised Map	[44]	[32]	

Table 1: Summary of two-stage clustering approaches for functional data.

3.2 Nonparametric approaches

Nonparametric approaches for functional data clustering are divided into two categories: methods who apply usual nonparametric clustering techniques (k-means or hierarchical clustering) with specific distances or dissimilarities, and methods which propose new heuristics or geometric criteria to cluster functional data.
In the first category of methods, several works consider the following measures of proximity between two curves x_i and $x_{i'}$:

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

where $x^{(\ell)}$ is the ℓ-th derivative of x. In [20] the authors propose to use hierarchical clustering combined with the distance d_0 – the L_2-metric – or with the semi-metric d_2. In [27] the k-means algorithm is used with d_0, d_1 and with $(d_0^2 + d_1^2)^{1/2}$. In [51] 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 [53] defining time-dependent clustering. These methods are summarized in Table 2.

	proximity measure			
clustering	d_0	d_1	$d_0 + d_1$	d_2
k-means	[27, 51, 53]	[27]	[27]	
hierarchical clustering	[20]			[20]

Table 2: Classical nonparametric clustering methods with proximity measures specific to functional data.

Remark: Following the method used to estimate the distance d_0, nonparametric methods can be assimilated to raw-data clustering or to a two-stage methods. 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 –, nonparametric methods are equivalent to raw-data clustering methods. Similarly, if an approximation of the curves into a finite basis is used to approximate d_0 – with function *semimetric.basis()* of *fda.usc* –, nonparametric methods are equivalent to two-stage methods with the same basis approximation.

The second category of nonparametric methods proposes new heuristics to cluster functional. In [26] two dynamic programming algorithms which simultaneously perform clustering and piecewise estimation of the cluster centres are proposed. Recently, [57] develops a new procedure to identify optimal clusters of functions and optimal subspaces for clustering, simultaneously. For this purpose, an objective function is defined as the sum of the distances between the observations and their projections plus the distances between the projections and the cluster means (in the projection space). An alternate algorithm is used to optimize the objective function.

3.3 Model-based approaches

Model on	Type of model	Reference
FPCA scores	Gaussian (parsimonious sub-models)	[10]
	Gaussian	[28, 29]
	Gaussian spherical (k-means)	[14]
basis expansion coefficients	Gaussian (parsimonious)	[30]
	Gaussian with regime changes	[45]
	Bayesian	[22, 25, 43]

Table 3: Model-based clustering approaches for functional data.

Model-based clustering techniques for functional data are not so straightforward as in the finite-dimensional setting, since the notion of density probability is generally not defined for functional random variable [17]. Thus, such techniques consists in assuming a density probability on a finite number of parameters describing the curves. But contrary to two-stage methods, in which the estimation of these coefficients is done previously to clustering, these two tasks are performed simultaneously with model-based techniques.

We divide model-based clustering techniques for functional data into two sets of methods, summarized in Table 3: those modelling the FPCA scores and those modelling directly the expansion coefficients in a finite basis of functions.

3.3.1 Model-based functional clustering techniques using principal components modelling

In [17], an approximation of the notion of probability density for functional random variables is proposed. This approximation is based on the truncation (2.4) 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 dataset. Using a similar approximation of the notion of density for functional random variables, [10] and [28] assume a Gaussian distribution of the principal components, and define model-based clustering techniques by the mean of the following mixture model [28]:

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

where $\theta = (\pi_k, \lambda_{1k}, \ldots, \lambda_{q_k k})_{1 \leq k \leq K}$ are the model parameters and q_k is the order of truncation of the Karhunen-Loeve expansion (2.3), specific to cluster k. The main interest of this model, called *funclust* by the authors, is in the fact that principal component scores $c_{jk}(x)$ of x are computed per cluster, thanks to 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 [12]. In [10], 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 Section 2.3), 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 [9]. The choice between these different sub-models is performed thanks to the BIC criterion [48].
Previously to this work, [14] 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. This model, named *k-centres*, is a particular case of [10, 28] assuming narrowly that the variance λ_{jk} are all equals within and between clusters.

3.3.2 Model-based functional clustering techniques using basis expansion coefficients modelling

To our knowledge, the first model-based clustering algorithm has been proposed in [30], under the name *fclust*. The authors consider that the expansion coefficients of the curves into a spline basis of functions 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 two-stage approaches, in which the basis expansion coefficients are considered fixed, they are considered as random variable, 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 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, [22] 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 [45], 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 time. Thus, at each time 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, [25] consider that the expansion coefficients are distributed as follows:

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

where $\mathcal{IG}$ is the Inverse-Gamma distribution. On the other hand, [43] 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.

3.4 Synthesis

We now present a short synthesis underlying the advantage and disadvantage of each categories of methods.

The use of raw-data clustering is probably the worst choice since it does not take into account the "time-dependent" structure of data, which is inherent to functional data.
The two-stage methods consider the functional nature of the data since the first stage consists of approximating the curves into a finite basis of function. The main weakness of these methods is that the filtering step is done previously to clustering, and then independently of the goal of clustering.
Nonparametric methods have the advantage of their simplicity: these methods are easy to understand and to implement. But their strength is also their weakness, since complex cluster structures can not be efficiently modelled by such approach. For instance, using k-means assumes in particular a common variance structure for each cluster, which is not always a realistic assumption.
In our opinion, the best methods are model-based clustering ones, because they take into account the functional nature of data, they perform simultaneously dimensionality reduction and clustering, and they allow to model complex covariance structure by modelling more or less free covariance operator with more or less parsimony assumptions. Both sub-categories of methods discussed in Section 3.3.2 and 3.3.1 are very efficient, and they both have their own advantages. On the one hand, model-based approaches built on the modelling of the basis expansion coefficients allow to model the uncertainty due to the approximation of the curve into a finite basis of function, what can be important especially for sparsely sample curve. On the other hand, the model-based approaches built on the principal component modelling define a general framework which can for instance be efficiently extended to the clustering of multivariate curves [29] or categorical curves.

4 Model selection

A common problem to clustering studies is the choice of the number of clusters. We present in Section 4.1 several criterion used for model selection in the functional data framework. A second model selection problem occurs for methods using an approximation of the curves into a finite basis of function, *i.e.* two-stage methods and model-based ones: the choice of an appropriate basis. Section 4.2 discusses this issue.

4.1 Choosing the number of clusters

If classical model selection tools, as BIC [48], AIC [2] or ICL [6] are frequently used in the context of model-based clustering to select the number of clusters (see for instance [10, 22, 45, 49]), more specific criteria have also been introduced.
First of all, Bayesian model for functional data clustering [25, 43] defines a framework in which the number of clusters can be directly estimated. For instance, [25] 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 the two-stage clustering method presented in [32], 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, [14, 27] retain the number of clusters leading to a partition having the best physical interpretation.
In [30], an original model selection criterion is considered. Proposed initially in [50], this criterion is defined as the averaged Mahalanobis distance between the basis expansion coefficients α_i and their closest cluster centre. In [50], 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).

4.2 Choosing the approximation basis

Almost all clustering algorithms for functional data needs the approximation of the curves into a finite dimensional basis of functions. Therefore, there is a need to choose an appropriate basis and thus, the number of basis functions. In [42], 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. The other solution is to use less subjective criteria such as penalized likelihood criteria BIC, AIC or ICL. The reader can for instance to refer to [30, 45, 49] for such use.

5 Software

Whereas there exist several software solutions for finite dimensional data clustering, the software devoted to functional data clustering is less developed.
Under the **R** software environment, two-stage methods can be performed using for instance the functions *kmeans* or *hclust* of the *stats* package, combined with the distances available from the *fda* or *fda.usc* packages.
Alternatively, several recent model-based clustering algorithms have been implemented by their authors and are available under different forms:

- **R** functions for funHDDC [10] and funclust [28] are available from request from their authors. An **R** package is currently under construction and will be available in 2013 on the CRAN[2] website,
- an **R** function for fclust [30] is available directly from James's webpage,
- the package *curvclust* for **R** [22] is probably the most finalized tool for curves clustering in **R**, and implements the wavelets-based methods [22] described in Section 3.3.2.

A MATLAB toolbox, *Curve Clustering Toolbox* [21], implements a family of two-stage clustering algorithms combining mixture of Gaussian models with spline or polynomial basis approximation.

6 Numerical illustration

The evaluation of clustering algorithms is always a difficult task [23]. In this review, we only illustrate the ability of the clustering algorithms previously discussed to retrieve the class labels of classification benchmark datasets.

6.1 The data

Three real datasets are considered: the *Kneading*, *Growth*, and *ECG* datasets. These three datasets are plotted in Figure 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. The kneading dataset is described in detail in [34]. There are 115 different flours for which the dough resistance is measured during the kneading process for 480 seconds. One obtains

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

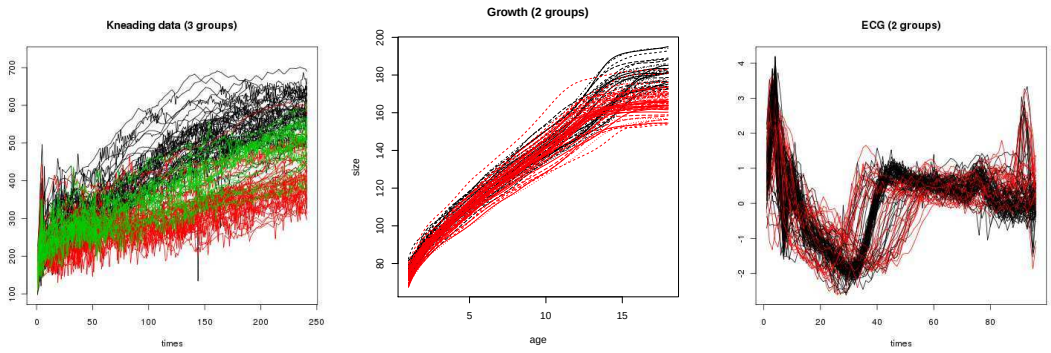


Figure 3: *Kneading*, *Growth* and *ECG* datasets.

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 [34, 40]. 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 [34, 40], 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 [54] 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 electrocardiogram from 2 groups of patients sampled at 96 time instants, and has already been studied in [38]. 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 algorithm presented in Section 3 can not tested, in particular because they are not all implemented in a software. The following clustering algorithms for functional data are considered:

- two-stage methods:
  - the classical clustering methods for finite dimensional data considered are k-means, hierarchical clustering and Gaussian Mixture Models (package *mclust* [4]) and two methods dedicated to the clustering of high-dimensional data: *HDDC* [9] and *MixtPPCA* [52],
  - these methods are applied on the FPCA scores with choice of the number of components with the Cattell scree test, directly on discretizations of the curves at the observation times points, and on the coefficients in the cubic B-spline basis approximation.

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

- non-parametric method: k-means with distance d_0 and d_1 [27],
- model-based clustering methods: Funclust [28], FunHDDC [10], fclust [30], k-centres [14] (results for the Growth dataset are available in their paper, but not software allow to proceed with the two other datasets), and curvclust [22].

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

6.3 Results

The correct classification rates (CCR) according to the known partitions are given in Table 4. Even if no numerical study can conclude to which method is the best, the present results suggest several comments:

- A first comment concerns the use of the different types of clustering methods: two-stage, nonparametric or model-based approaches. Two-stage methods can sometimes perform very well (to estimate the class label), but the main problem is that, in the present unsupervised context, we have no possibility to choose between working with the discrete data, with the spline coefficients or with the FPCA scores. For instance, HDDC and MixtPPCA are very well performing on the Growth dataset using the FPCA scores, but they are very poor using the discrete data or the spline coefficients. If nonparametric methods suffer from a similar limitation, due to the choice of the distance or dissimilarity to use, model-based clustering methods, which also require the choice of an appropriate basis, allow generally to use penalized likelihood criteria such as BIC to evaluate the different basis choices. In that sense, model-based approaches provide more flexible tools for functional data clustering.
- Concerning the model-based clustering methods, FunHDDC and Funclust are among the best methods on these datasets. On the contrary, fclust and curvclust lead to relatively poor clustering results. This is probably due to the nature of the data, which are regularly sampled and without peak whereas fclust and curvclust are especially designed for respectively irregularly sampled curves and peak-like data.

7 Conclusion and future challenge

This paper has presented a review of the main existing algorithms for functional data clustering. A classification of these methods has been proposed, into three main groups: 1. two-stage methods which perform dimension reduction before clustering, 2. nonparametric methods using specific distances or dissimilarities between curves and 3. model-based clustering methods which assume a probabilistic distribution on either the FPCA scores or the coefficients of curves into a basis approximation. A critical analysis has been proposed, which highlights the advantages of model-based clustering methods. Some numerical illustrations and a short software review have also been presented, and the corresponding **R** codes given in the appendix may help the reader in applying these clustering algorithms to his own data.

Literature on functional data clustering generally consider the case of functional data as realizations of a stochastic process $X = \{X_t\}_{t\in\mathcal{T}}$, with $X_t \in \mathbb{R}$, which is the subject of the present paper. Recently, some authors are interested in the case of multivariate functional data, *i.e.* $X_t \in \mathbb{R}^p$, in which a path of X is a set of p curves. An example of bivariate functional data is given in [42] with temperature and precipitation curves of Canadian weather stations. Few works have defined clustering algorithms for such multivariate functional data: [27, 29, 57]. Another case of interest is qualitative functional data [8], in which X_t lives in a categorical space.

	Kneading		Kneading		
	functional	2-stage methods	discretized (241 instants)	spline coeff. (20 splines)	FPCA scores (4 components)
Funclust	**66.96**	HDDC	66.09	53.91	44.35
FunHDDC	62.61	MixtPPCA	65.22	64.35	62.61
fclust	64	GMM	63.48	50.43	60
k-centres	-	k-means	62.61	62.61	62.61
curvclust	65.21	hclust	63.48	63.48	63.48
kmeans-d_0	62.61				
kmeans-d_1	64.35				
	Growth		Growth		
	functional	2-stage methods	discretized (350 instants)	spline coeff. (20 splines)	FPCA scores (2 components)
Funclust	69.89	HDDC	56.99	50.51	**97.85**
FunHDDC	96.77	MixtPPCA	62.36	50.53	**97.85**
fclust	69.89	GMM	65.59	63.44	95.70
k-centres	93.55	k-means	65.59	66.67	64.52
curvclust	67.74	hclust	51.61	75.27	68.81
kmeans-d_0	64.52				
kmeans-d_1	87.40				
	ECG		ECG		
	functional	2-stage methods	discretized (96 instants)	spline coeff. (20 splines)	FPCA scores (19 components)
Funclust	**84**	HDDC	74.5	73.5	74.5
FunHDDC	75	MixtPPCA	74.5	73.5	74.5
fclust	74.5	GMM	81	80.5	81.5
k-centres	-	k-means	74.5	72.5	74.5
curvclust	74.5	hclust	73	76.5	64
kmeans-d_0	74.5				
kmeans-d_1	61.5				

Table 4: Correct classification rates (CCR) in percentage for Funclust, FunHDDC (best model according BIC), fclust, kCFC, curvclust and usual non-functional methods on the Kneading, Growth and ECG datasets.

The marital status of individuals, the status of some patients with respect to some diseases are some examples of such data. To the best of our knowledge, there are no works considering this type of data in the functional data context and in particular, in the clustering topic.

A R codes for curve clustering

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

A.1 Data loading

First, the values of the functional data at the observations time 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
```

A.2 Clustering with Funclust and FunHDDC

The corresponding computer code are available from request to their authors.
Funclust and FunHDDC can be applied directly on the `fd` object `fdata`:
```
>res=funclust(fd,K=K)
```
and
```
>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$bic`).
For both methods, the clusters are stored in `res$cls`.

A.3 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`

A.4 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

[1] C. Abraham, P. A. Cornillon, E. Matzner-Løber, and N. Molinari. Unsupervised curve clustering using B-splines. *Scandinavian Journal of Statistics. Theory and Applications*, 30(3):581–595, 2003.

[2] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Automatic Control*, AC-19:716–723, 1974. System identification and time-series analysis.

[3] A. Antoniadis and J. H. Beder. Joint estimation of the mean and the covariance of a Banach valued Gaussian vector. *Statistics*, 20(1):77–93, 1989.

[4] J.D. Banfield and A.E. Raftery. Model-based Gaussian and non-Gaussian clustering. *Biometrics*, 49:803–821, 1993.

[5] P. Besse. Etude descriptive d'un processus. *Thèse de doctorat* $3^{ème}$ *cycle*, Université Paul Sabatier, Toulouse, 1979.

[6] C. Biernacki, G. Celeux, and G. Govaert. Assessing a mixture model for clustering with the inegrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(4):719–725, 2000.

[7] D. Bosq. *Linear processes in function spaces*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 2000. Theory and applications.

[8] R. Boumaza. *Contribution a l'étude descriptive d'une fonction aléatoire qualitative*. PhD thesis, Université Paul Sabatier, Toulouse, France, 1980.

[9] C. Bouveyron, S. Girard, and C. Schmid. High Dimensional Data Clustering. *Computational Statistics and Data Analysis*, 52:502–519, 2007.

[10] C. Bouveyron and J. Jacques. Model-based clustering of time series in group-specific functional subspaces. *Advances in Data Analysis and Classification*, 5(4):281–300, 2011.

[11] Bouveyron C. and Brunet C. Model-based clustering of high-dimensional data : A review. Technical report, Laboratoire SAMM, Université Paris 1 Panthéon-Sorbonne, 2012.

[12] R. Cattell. The scree test for the number of factors. *Multivariate Behav. Res.*, 1(2):245–276, 1966.

[13] G. Celeux and G. Govaert. Gaussian parsimonious clustering models. *The Journal of the Pattern Recognition Society*, 28:781–793, 1995.

[14] J-M. Chiou and P-L. Li. Functional clustering and identifying substructures of longitudinal data. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 69(4):679–699, 2007.

[15] T.F. Cox and M.A.A Cox. *Multidimensional Scaling*. Chapman and Hall, New York, 2001.

[16] J. Dauxois, A. Pousse, and Y. Romain. Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *Journal of Multivariate Analysis*, 12(1):136–154, 1982.

[17] A. Delaigle and P. Hall. Defining probability density for a distribution of random functions. *The Annals of Statistics*, 38:1171–1193, 2010.

[18] J.C. Deville. Méthodes statistiques et numériques de l'analyse harmonique. *Annales de l'INSEE*, 15:3–101, 1974.

[19] M. Escabias, A.M. Aguilera, and M.J. Valderrama. Modeling environmental data by functional principal component logistic regression. *Environmetrics*, 16:95–107, 2005.

[20] F. Ferraty and P. Vieu. *Nonparametric functional data analysis*. Springer Series in Statistics. Springer, New York, 2006.

[21] S. Gaffney. *Probabilistic Curve-Aligned Clustering and Prediction with Mixture Models*. PhD thesis, Department of Computer Science, University of California, Irvine, USA, 2004.

[22] M. Giacofci, S. Lambert-Lacroix, G. Marot, and F. Picard. Wavelet-based clustering for mixed-effects functional models in high dimension. *Biometrics*, in press, 2012.

[23] I. Guyon, U. Von Luxburg, and R.C. Williamson. Clustering: Science or art. In *NIPS 2009 Workshop on Clustering Theory*, 2009.

[24] J.A. Hartigan and M.A. Wong. Algorithm as 1326 : A k-means clustering algorithm. *Applied Statistics*, 28:100–108, 1978.

[25] N.A. Heard, C.C. Holmes, and D.A. Stephens. A quantitative study of gene regulation involved in the immune response of anopheline mosquitoes: an application of Bayesian hierarchical clustering of curves. *Journal of the American Statistical Association*, 101(473):18–29, 2006.

[26] G. Hébrail, B. Hugueney, Y. Lechevallier, and F. Rossi. Exploratory Analysis of Functional Data via Clustering and Optimal Segmentation. *Neurocomputing / EEG Neurocomputing*, 73(7-9):1125–1141, 03 2010.

[27] F. Ieva, A.M. Paganoni, D. Pigoli, and V. Vitelli. Multivariate functional clustering for the analysis of ecg curves morphology. *Journal of the Royal Statistical Society. Series C. Applied Statistics*, in press, 2012.

[28] J. Jacques and C. Preda. Funclust: a curves clustering method using functional random variable density approximation. *Neurocomputing*, in press, 2013.

[29] J. Jacques and C. Preda. Model-based clustering for multivariate functional data. *Computational Statistics and Data Analysis*, in press, 2013.

[30] G.M. James and C.A. Sugar. Clustering for sparsely sampled functional data. *Journal of the American Statistical Association*, 98(462):397–408, 2003.

[31] K. Karhunen. Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys.*, 1947(37):79, 1947.

[32] M. Kayano, K. Dozono, and S. Konishi. Functional Cluster Analysis via Orthonormalized Gaussian Basis Expansions and Its Application. *Journal of Classification*, 27:211–230, 2010.

[33] T. Kohonen. *Self-Organizing Maps*. Springer-Verlag, New York, 1995.

[34] C. Lévéder, P.A. Abraham, E. Cornillon, E. Matzner-Lober, and N. Molinari. Discrimination de courbes de prétrissage. In *Chimiométrie 2004*, pages 37–43, Paris, 2004.

[35] X. Liu and M.C.K. Yang. Simultaneous curve registration and clustering for functional data. *Computational Statistics and Data Analysis*, 53:1361–1376, 2009.

[36] M. Loève. Fonctions aléatoires de second ordre. *C. R. Acad. Sci. Paris*, 220:469, 1945.

[37] Geoffrey McLachlan and David Peel. *Finite mixture models*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience, New York, 2000.

[38] R.T. Olszewski. *Generalized Feature Extraction for Structural Pattern Recognition in Time-Series Data*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, 2001.

[39] J. Peng and H-G. Müller. Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions. *The Annals of Applied Statistics*, 2(3):1056–1077, 2008.

[40] C. Preda, G. Saporta, and C. Lévéder. PLS classification of functional data. *Comput. Statist.*, 22(2):223–235, 2007.

[41] J. O. Ramsay and B. W. Silverman. *Applied functional data analysis*. Springer Series in Statistics. Springer-Verlag, New York, 2002. Methods and case studies.

[42] J. O. Ramsay and B. W. Silverman. *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition, 2005.

[43] S. Ray and B. Mallick. Functional clustering by Bayesian wavelet methods. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 68(2):305–332, 2006.

[44] F. Rossi, B. Conan-Guez, and A. El Golli. Clustering functional data with the som algorithm. In *Proceedings of ESANN 2004*, pages 305–312, Bruges, Belgium, April 2004.

[45] A. Samé, F. Chamroukhi, G. Govaert, and P. Aknin. Model-based clustering and segmentation of times series with changes in regime. *Advances in Data Analysis and Classification*, 5(4):301–322, 2011.

[46] L.M. Sangalli, P. Secchi, S. Vantini, and V. Vitelli. k-mean alignment for curve clustering. *Computational Statistics & Data Analysis*, 54(5):1219–1233, 2010.

[47] G. Saporta. Méthodes exploratoires d'analyse de données temporelles. *Cahiers du Buro*, 37–38, 1981.

[48] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[49] N. Serban and H. Jiang. Multilevel functional clustering analysis. *Biometrics*, 68(3):805–814, 2012.

[50] C.A. Sugar and G.M. James. Finding the number of clusters in a dataset: an information-theoretic approach. *Journal of the American Statistical Association*, 98(463):750–763, 2003.

[51] T. Tarpey and K.J. Kinateder. Clustering functional data. *Journal of Classification*, 20(1):93–114, 2003.

[52] M. E. Tipping and C. Bishop. Mixtures of principal component analyzers. *Neural Computation*, 11(2):443–482, 1999.

[53] S. Tokushige, H. Yadohisa, and K. Inada. Crisp and fuzzy k-means clustering algorithms for multivariate functional data. *Computational Statistics*, 22:1–16, 2007.

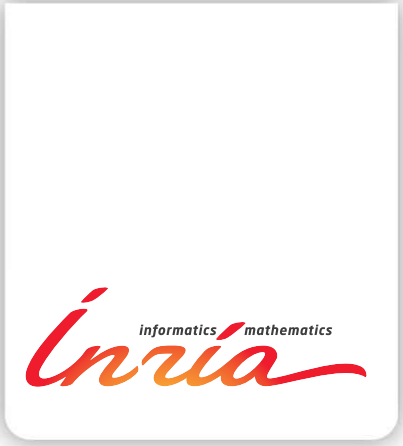
[54] R.D. Tuddenham and M.M. Snyder. Physical growth of california boys and girls from birth to eighteen years. *Universities of Califfornia Public Child Development*, 1:188–364, 1954.

[55] G. Wahba. *Spline models for observational data*. SIAM, Philadelphia, 1990.

[56] Joe H. Ward, Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58:236–244, 1963.

[57] M. Yamamoto. Clustering of functional data in a low-dimensional subspace. *Advances in Data Analysis and Classification*, 6:219–247, 2012.

Contents



RESEARCH CENTRE
LILLE – NORD EUROPE

Parc scientifique de la Haute-Borne
40 avenue Halley - Bât A - Park Plaza
59650 Villeneuve d'Ascq

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399