



HAL
open science

Model-based clustering for multivariate functional data

Julien Jacques, Cristian Preda

► **To cite this version:**

Julien Jacques, Cristian Preda. Model-based clustering for multivariate functional data. 2011. hal-00713334v1

HAL Id: hal-00713334

<https://hal.science/hal-00713334v1>

Preprint submitted on 29 Jun 2012 (v1), last revised 13 Oct 2012 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Model-based clustering for multivariate functional data

Julien Jacques*, Cristian Preda

Laboratoire Paul Painlevé, UMR CNRS 8524, University Lille I, Lille, France

MODAL team, INRIA Lille-Nord Europe, & Polytech'Lille

Abstract

Model-based clustering is considered for multivariate functional data. Based on multivariate functional principal components analysis (FPCA), a mixture model is defined and estimated by an EM-like algorithm. The main advantage of the proposed model is the ability to take into account the dependence among curves, thanks to the multivariate FPCA. Comparisons on simulated and real data show that the proposed method is a good alternative to conventional methods.

Keywords: Multivariate functional data, density approximation, model-based clustering, multivariate functional principal component analysis, EM algorithm

1. Introduction

Functional data analysis or "data analysis with curves" is an active topic in statistics with a wide range of applications. New technologies allow to record data with accuracy and at high frequency (in time or other dimension), generating large volume of data. In medicine one has growth curves of children and patient's state evolution, in climatology one records weather parameters over decades, chemometric curves are analysed in chemistry and physics (spectroscopy) and special attention is paid to the evolution of indicators coming from economy and finance. See Ramsay and Silverman (2005) for more details.

The statistical model underlying data represented by curves is a real-valued stochastic process with continuous time, $X = \{X_t\}_{t \in [0, T]}$. Most of the approaches dealing with functional data consider the univariate case, i.e. $X_t \in \mathbb{R}$, $\forall t \in [0, T]$, a path of X being represented by a single curve. Despite its evident interest, the multidimensional case,

$$\mathbf{X}(t) = (X_1(t), \dots, X_p(t)) \in \mathbb{R}^p, \quad p \geq 2$$

is, curiously, rarely considered in literature. In this case a path of \mathbf{X} is represented by a set of p curves. The dependency between these p measures provides the structure of \mathbf{X} . One

*Corresponding author. Tel.: +33 320 436 760, Fax: +33 320 434 302

Email addresses: julien.jacques@polytech-lille.fr (Julien Jacques),
cristian.preda@polytech-lille.fr (Cristian Preda)

finds in Ramsay and Silverman (2005) a brief example of bivariate functional data, $\mathbf{X}(t) = (X_1(t), X_2(t)) \in \mathbb{R}^2$, as a model for gait data (knee and hip measures) used in the context of functional principal component analysis (FPCA) as an extension of the univariate case. For a more theoretical framework, we must go back to the pioneer works of Besse (1979) on random variables with values into a general Hilbert space. Saporta (1981) provides a complete analysis of multivariate functional data from the point of view of factorial methods (principal components and canonical analysis).

In this paper we consider the problem of clustering multivariate functional data. Cluster analysis aims to identify homogeneous groups of data without using any prior knowledge on the group labels of data. The main difficulty in clustering functional data arises because of the infinite dimensional space data belong. Consequently, most of clustering algorithms for functional data consists of a first step of transforming the infinite dimensional problem into a finite dimensional one and of a second step of using a clustering method designed for finite dimensional data. Most of proposed approach consider the k-means algorithm applied to a B -spline fitting (Abraham et al., 2003), on defined principal points of curves (Tarpey and Kinateder, 2003) or on the truncated Karhunen-Loeve expansion (Chiou and Li, 2007). Sangalli et al. (2010) use also a k-means algorithm for clustering misaligned curves. As in the finite dimensional setting, where Gaussian model-based clustering generalizes the k-means algorithm, some other works introduce more sophisticated model-based techniques: James and Sugar (2003) define an approach particularly effective for sparsely sampled functional data, Ray and Mallick (2006) propose a nonparametric Bayes wavelet model for clustering of functional data based on a mixture of Dirichlet processes, Frühwirth-Schnatter and Kaufmann (2008) build a specific clustering algorithm based on parametric time series models, Bouveyron and Jacques (2011) extend the high-dimensional data clustering algorithm (HDDC, Bouveyron et al. (2007)) to the functional case and Jacques and Preda (2012) build a model-based clustering based on an approximation of the notion of density –which in general does not exists– for functional data.

The case of multivariate functional data is rarely considered in literature: Singhal and Seborg (2005) and Ieva et al. (2011) use a k-means algorithm based on specific distances between multivariate functional data, whereas Kayano et al. (2010) consider Self-Organizing Maps based on the coefficients of multivariate curves into orthonormalized Gaussian basis expansions. Tokushige et al. (2007) extend crisp and fuzzy k-means algorithms for multivariate functional data by considering an usual distance between functions, but applied their algorithms only on univariate functional data.

In the finite dimensional setting, model-based clustering algorithms consider that data is sampled from a mixture of density probabilities. This is not directly applicable to functional data since the notion of density probability generally does not exist for functional random variable (Delaigle and Hall, 2010). Consequently, model-based clustering algorithms assume a parametric distribution on some finite set of coefficients characterizing the curves. In Jacques and Preda (2012), the authors use the density surrogate defined in Delaigle and Hall (2010) to build a model-based clustering for univariate functional data. This density surrogate, based on

the truncation of the Karhunen-Loeve expansion, relies on the density probability of the first principal components of the curves (Ramsay and Silverman, 2005). In this paper we propose an extension of Jacques and Preda (2012) approach to multivariate functional data. We firstly introduce principal component analysis for multivariate functional data and assume a cluster-specific Gaussian distribution for the principal component scores. The elements derived from FPCA are estimated using approximation for the multivariate curves into a finite dimensional functional space. The number of first principal components used in the density surrogate as well as the computation of the principal component scores are cluster specific.

The main advantage of our model is its ability to take into account dependency between the p curves, thanks to the principal component analysis for multivariate functional data.

The paper is organized as follows. Section 2 introduces the model for multivariate functional data and present the principal components analysis for this type of data. Estimation and approximation details are provided. Section 3 defines an approximation of the probability density for multivariate functional random variable. The model-based clustering approach and parameters estimation by an EM-like algorithm are presented in Section 4. Comparisons with existing methods on simulated and real datasets are presented in Section 5.

2. Principal component analysis for multivariate functional data

Principal component analysis for multivariate functional data (MFPCA) has already been suggested in Ramsay and Silverman (2005) and Berrendero et al. (2011). In Ramsay and Silverman (2005) the authors propose to concatenate the observations of the functions at a fine grid of points (or the coefficients in a suitable basis expansion) into a single vector and then to perform a standard PCA on these concatenated vectors. When a basis expansion is used, this method forces to consider only orthonormal bases since the metric induces by the scalar product between the basis functions is not taken into account. In Berrendero et al. (2011), the authors propose to not summarize the curves with scalar principal components as in usual FPCA, but with functional ones. For this, they carry out classical multivariate PCA for each value of the domain on which the functions are observed and suggest an interpolation method to build functional principal components.

Our approach is close to Ramsay and Silverman (2005) but by taking into account the possible use of non orthonormal basis. In particular, our method allows to use different basis for each dimension of the multivariate curves.

Let Ω be a population space of statistical units and $s_n = \{\omega_1, \omega_2, \dots, \omega_n\}$ be a random sample of size n drawn from Ω . Let \mathbf{X} be a random variable defined on Ω associating to $\omega \in \Omega$ a set of p curves, $p \geq 2$, each one defined on the finite interval $[0, T]$, $0 < T < \infty$, i.e

$$\mathbf{X}(\omega) = \{(X_{\omega,1}(t), \dots, X_{\omega,p}(t)), t \in [0, T]\}.$$

The observation of \mathbf{X} on the sample s_n provides the set $\{\mathbf{X}(\omega_1), \dots, \mathbf{X}(\omega_n)\}$ of multivariate curves called *multivariate functional data*. From the s_n curves, one can be interested in optimal representation of curves in a reduced dimensional function space (principal component

analysis), or in clustering, by determining an optimal partition of s_n with respect to some distance or homogeneity criterion. In order to address these two questions in a formal way, we need the hypothesis that considers $\mathbf{X} = (X_1, \dots, X_p)$ is a L_2 continuous stochastic process,

$$\lim_{h \rightarrow 0} \mathbb{E} [\|\mathbf{X}(t+h) - \mathbf{X}(t)\|^2] = \lim_{h \rightarrow 0} \int_0^T \sum_{\ell=1}^p \mathbb{E} [(X_\ell(t+h) - X_\ell(t))^2] = 0.$$

Notice that L_2 -continuity of \mathbf{X} implies L_2 -continuity of each component of \mathbf{X} , *i.e.* X_ℓ is a L_2 -continuous stochastic process for all $\ell = 1, \dots, p$. The L_2 -continuity is a quite general hypothesis, most of real data applications satisfying this one.

Let denote by $\mu_\ell = \{\mu_\ell(t) = \mathbb{E}[X_\ell(t)], t \in [0, T]\}$ the mean function of X_ℓ and by

$$\boldsymbol{\mu} = (\mu_1, \dots, \mu_p) = \mathbb{E}[\mathbf{X}],$$

the mean function of \mathbf{X} .

The covariance operator of \mathbf{X} is defined as an integral operator C with kernel

$$C(t, s) = \mathbb{E}[(\mathbf{X}(t) - \boldsymbol{\mu}(t)) \otimes (\mathbf{X}(s) - \boldsymbol{\mu}(s))],$$

where \otimes is the tensor product on \mathbb{R}^p . Thus, $C(t, s)$ is a $p \times p$ matrix with elements

$$C(t, s)[i, j] = \text{Cov}(X_i(t), X_j(s)), \quad i, j = 1, \dots, p.$$

The covariance operator of \mathbf{X} , $C : L_2([0, T])^p \rightarrow L_2([0, T])^p$ is defined by

$$\mathbf{f} \xrightarrow{C} \mathbf{g}, \quad \mathbf{g}(t) = \int_0^T C(t, s) \mathbf{f}(s) ds, \quad t \in [0, T],$$

where $\mathbf{f} = (f_1, \dots, f_p)$ and $\mathbf{g} = (g_1, \dots, g_p)$ are elements of $L_2([0, T])^p$.

2.1. Principal components analysis of \mathbf{X}

Under the hypothesis of L_2 -continuity, C is an Hilbert-Schmidt operator, *i.e.* compact, self-adjoint and such that $\sum_{j \geq 1} \lambda_j^2 < +\infty$. The spectral analysis of C provides a countable set of positive eigenvalues $\{\lambda_j\}_{j \geq 1}$ associated to an orthonormal basis of eigen-functions $\{\mathbf{f}_j\}_{j \geq 1}$, $\mathbf{f}_j = (f_{j,1}, \dots, f_{j,p})$, called principal factors:

$$C\mathbf{f}_j = \lambda_j \mathbf{f}_j, \tag{1}$$

with $\lambda_1 \geq \lambda_2 \geq \dots$ and $\langle \mathbf{f}_i, \mathbf{f}_j \rangle_{L_2([0, T])^p} = \int_0^T \sum_{\ell=1}^p f_{i,\ell}(t) f_{j,\ell}(t) dt = \delta_{i,j}$ with $\delta_{i,j} = 1$ if $i = j$ and 0 otherwise.

The principal components C_j of \mathbf{X} are zero-mean random variables defined as the projections of \mathbf{X} on the eigenfunctions of C ,

$$C_j = \int_0^T \langle \mathbf{X}(t) - \boldsymbol{\mu}(t), \mathbf{f}_j(t) \rangle_{\mathbb{R}^p} dt = \int_0^T \sum_{\ell=1}^p (X_\ell(t) - \mu_\ell(t)) f_{j,\ell}(t) dt.$$

Let recall that, as in the univariate setting, the principal components $\{C_j\}_{j \geq 1}$ are zero-mean uncorrelated random variables with variance $\mathbb{V}(C_j) = \lambda_j$, $j \geq 1$.

The following Karhunen-Loeve expansion holds (Saporta, 1981),

$$\mathbf{X}(t) = \boldsymbol{\mu}(t) + \sum_{j \geq 1} C_j \mathbf{f}_j(t),$$

and the approximation of order q of \mathbf{X} , $q \in \mathbb{N}^*$,

$$\mathbf{X}^{(q)}(t) = \boldsymbol{\mu}(t) + \sum_{j=1}^q C_j \mathbf{f}_j(t),$$

is the best approximation of this form under the mean square criterion.

2.2. Estimation and computational methods

Let consider the random sample of size n , $s_n = \{\omega_1, \omega_2, \dots, \omega_n\}$, and denote by $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})' = \mathbf{X}(\omega_i)'$. The estimators for $\boldsymbol{\mu}$ and C are

$$\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \quad \text{and} \quad \hat{C}(t, s) = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i(t) - \hat{\boldsymbol{\mu}}(t)) \otimes (\mathbf{x}_i(s) - \hat{\boldsymbol{\mu}}(s)).$$

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

2.2.1. Approximation into a finite basis of functions

Often in practice, data are observed at discrete time points and with some noise. In order to get the functional feature of data, smoothing and interpolation methods are used considering the true curve belongs to a finite dimensional space spanned by some basis of functions. This approximation reduces also the eigen-analysis problem (1) to one in finite dimensional setting.

Let assume that each curve $x_{i\ell}$ ($1 \leq \ell \leq p$) can be expressed as a linear combination of basis functions $\Phi_\ell = (\phi_{\ell 1}, \dots, \phi_{\ell q_\ell})$:

$$x_{i\ell}(t) = \sum_{j=1}^{q_\ell} \xi_{i\ell j} \phi_{\ell j}(t).$$

This can be written with the matrix formulation

$$\mathbf{x}_i = \Phi \mathbf{a}_i'$$

with

$$\Phi = \begin{pmatrix} \phi_{11} & \dots & \phi_{1q_1} & 0 & & \dots & & & 0 \\ 0 & \dots & 0 & \phi_{21} & \dots & \phi_{2q_2} & 0 & \dots & 0 \\ & & & & \dots & & & & \\ 0 & & & \dots & & & 0 & \phi_{p1} & \dots & \phi_{pq_p} \end{pmatrix}$$

and

$$\mathbf{a}_i = (\xi_{i11}, \dots, \xi_{i1q_1}, \xi_{i21}, \dots, \xi_{i2q_2}, \dots, \xi_{ip1}, \dots, \xi_{ipq_p}).$$

The sample s_n provides n paths of \mathbf{X} , $\mathbf{x}_i = \mathbf{X}(\omega_i)$, $i = 1, \dots, n$. Put $\underline{\mathbf{x}} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$. we have

$$\underline{\mathbf{x}} = \tilde{\mathbf{A}}\Phi' \quad (2)$$

with $\tilde{\mathbf{A}}$ the $n \times \sum_{\ell=1}^p q_\ell$ -matrix, with rows the vectors \mathbf{a}_i ($i = 1, \dots, n$).

Under the previous basis expansion assumption, the covariance matrix estimator $\hat{C}(t, s)$ is

$$\hat{C}(t, s) = \frac{1}{n-1} (\underline{\mathbf{x}}(t) - \hat{\boldsymbol{\mu}}(t))' (\underline{\mathbf{x}}(s) - \hat{\boldsymbol{\mu}}(s)) = \frac{1}{n-1} \Phi(t) \mathbf{A}' \mathbf{A} \Phi'(s),$$

with $\mathbf{A} = (I_n - \mathbb{I}_n(1/n, \dots, 1/n)) \tilde{\mathbf{A}}$ where I_n and \mathbb{I}_n are respectively the identity $n \times n$ -matrix and the unit column vector of size n .

From (1), each principal factor \mathbf{f}_j belongs to the linear space spanned by the basis Φ :

$$\mathbf{f}_j = \Phi \mathbf{b}'_j$$

with $\mathbf{b} = (b_{j11}, \dots, b_{j1q_1}, b_{j21}, \dots, b_{j2q_2}, \dots, b_{jp1}, \dots, b_{jpq_p})$.

Then

$$C\mathbf{f}_j = \int_0^T C(t, s) \mathbf{f}_j(s) ds = \int_0^T \frac{1}{n-1} \Phi(t) \mathbf{A}' \mathbf{A} \Phi'(s) \mathbf{f}_j(s) ds \quad (3)$$

$$= \int_0^T \frac{1}{n-1} \Phi(t) \mathbf{A}' \mathbf{A} \Phi'(s) \Phi(s) \mathbf{b}'_j ds = \frac{1}{n-1} \Phi(t) \mathbf{A}' \mathbf{A} \underbrace{\int_0^T \Phi'(s) \Phi(s) ds}_{\mathbf{W}} \mathbf{b}'_j \quad (4)$$

where $\mathbf{W} = \int_0^T \Phi'(s) \Phi(s) ds$ is the symmetric block-diagonal $\sum_{\ell=1}^p q_\ell \times \sum_{\ell=1}^p q_\ell$ -matrix of the inner products between the basis functions.

The eigen-analysis problem (1) can be written as

$$\frac{1}{n-1} \Phi(t) \mathbf{A}' \mathbf{A} \mathbf{W} \mathbf{b}'_j = \lambda_j \Phi(t) \mathbf{b}'_j \quad (5)$$

which becomes, since (5) is true for all t ,

$$\frac{1}{n-1} \mathbf{A}' \mathbf{A} \mathbf{W} \mathbf{b}'_j = \lambda_j \mathbf{b}'_j.$$

By defining $\mathbf{u}_j = \mathbf{b}_j \mathbf{W}^{1/2}$, the multivariate functional principal component analysis is approximated by the usual PCA of the matrix $\frac{1}{\sqrt{n-1}} \mathbf{A}' \mathbf{A} \mathbf{W}^{1/2}$:

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

The coefficient \mathbf{b}_j , $j \geq 1$, of the principal factors are obtained by $\mathbf{b}_j = (\mathbf{W}^{1/2'})^{-1}\mathbf{u}'_j$, and the principal component scores,

$$C_j = \mathbf{A}\mathbf{W}\mathbf{b}'_j \quad j \geq 1.$$

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

$$\frac{1}{n-1}\mathbf{A}\mathbf{W}\mathbf{A}'C_j = \lambda_j C_j.$$

2.3. Normed principal component analysis

When the X_ℓ 's components of \mathbf{X} ($1 \leq \ell \leq p$) are of different natures (different measure units for example), the question of normalizing data occurs naturally. One way to introduce a metric in MFPCA is to consider it as a canonical analysis (Saporta, 1981).

The principal components are then defined as solutions of the following eigen problem:

$$\int_0^T \mathbf{P}_t(\tilde{C}) dt = \lambda \tilde{C}, \quad (6)$$

where \mathbf{P}_t is the orthogonal projection operator associated to $\mathbf{X}(t) = (X_1(t), \dots, X_p(t))$ defined by

$$\mathbf{P}_t(\tilde{C}) = \langle \mathbf{X}(t), [C(t,t)]^{-1} \mathbb{E}[\mathbf{X}(t)\tilde{C}] \rangle_{\mathbb{R}^p}. \quad (7)$$

Replacing (7) into (6) one obtains that

$$\tilde{C} = \int_0^T \langle \mathbf{X}(t) - \mu(t), \mathbf{f}(t) \rangle dt$$

where \mathbf{f} is solution of the eigenvector problem

$$\int_0^T [C(t,t)]^{-1} C(t,s) \mathbf{f}(s) ds = \lambda \mathbf{f}(t). \quad (8)$$

Clearly, $[C(t,t)]^{-1}$ must exist for each $t \in [0, T]$. Under this hypothesis, the principal factors of the normalized MFPCA are the eigenfunctions of the integral operator with kernel $[C(t,t)]^{-1} C(t,s)$ as in (8). Notice that with the orthogonality condition on the principal factors

$$\int_0^T \langle f_i(t), [C(t,t)] f_j(t) \rangle_{\mathbb{R}^p} = \delta_{ij}, \quad i, j \geq 1,$$

with $\delta_{ij} = 1$ if $i = j$, 0 otherwise. The Karhunen-Loeve expansion of \mathbf{X} becomes

$$\mathbf{X}(t) = \mu(t) + \sum_{j=1}^{\infty} \tilde{C}_j \cdot [C(t,t)] \cdot \mathbf{f}_j(t),$$

with $\tilde{C}_j = \int_0^T \langle \mathbf{X}(t) - \mu(t), \mathbf{f}_j(t) \rangle dt$, $\text{Var}(\tilde{C}_j) = \lambda_j$ and $\mathbb{E}[\tilde{C}_j] = 0$.

Remark. Observe that if $R(t, t)$ is the square root of the matrix $C(t, t)$, *i.e.*

$$C(t, t) = R(t, t)R(t, t)',$$

then the MFPCA of \mathbf{X} with metric $C(t, t)$ is equivalent to the MFPCA of \mathbf{Y} defined by

$$\mathbf{Y}(t) = R(t, t)^{-1}\mathbf{X}(t),$$

with identity metric as in Section 2.1.

In practice, if \mathbf{X} is observed at times t_1, \dots, t_r , $r > 1$, then \mathbf{Y} is defined from \mathbf{X} as

$$\mathbf{Y}(t_i) = R(t_i, t_i)^{-1}\mathbf{X}(t_i), \quad i = 1, \dots, r$$

and approximation and estimation methodology in Section 2.2.1 is applied to \mathbf{Y} .

3. Approximation of the density for multivariate functional data

As the notion of probability density is not well defined for functional data (Delaigle and Hall, 2010), we can use an approximation of the density based on the Karhunen-Loeve expansion, adapted here to the multidimensional nature of the data:

$$\mathbf{X}(t) - \mu(t) = \sum_{j=1}^{\infty} C_j \mathbf{f}_j(t). \quad (9)$$

Considering the principal components indexed upon the descending order of the eigenvalues ($\lambda_1 \geq \lambda_2 \geq \dots$), let denote by $\mathbf{X}^{(q)}$ the approximation of \mathbf{X} by truncating (9) at the q first terms, $q \geq 1$,

$$\mathbf{X}^{(q)}(t) - \mu(t) = \sum_{j=1}^q C_j \mathbf{f}_j(t). \quad (10)$$

Then, $\mathbf{X}^{(q)}$ is the best approximation of \mathbf{X} , under the mean square criterion, among all the approximations of the same type (linear combination of q deterministic functions of t with random coefficients, Saporta (1981)).

Based on the approximation of \mathbf{X} by $\mathbf{X}^{(q)}$, in Delaigle and Hall (2010) it is shown that the probability of \mathbf{X} to belong to a ball of radius h centred in $\mathbf{x} \in L_2([0, T])^p$ can be written as

$$\log P(\|\mathbf{X} - \mathbf{x}\| \leq h) = \sum_{j=1}^q \log f_{C_j}(c_j(\mathbf{x})) + \xi(h, q(h)) + o(q(h)), \quad (11)$$

where f_{C_j} is the probability density of C_j and $c_j(\mathbf{x})$ is the j th principal component score of x , $c_j(\mathbf{x}) = \langle \mathbf{f}_j, \mathbf{x} \rangle_{L_2([0, T])^p}$. The functions $q(h)$ and ξ are such that $q(h)$ grows to infinity when h decreases to zero and ξ is a constant depending on h and $q(h)$.

The equality (11) suggests the use of the multivariate probability density of the vector $C^{(q)} = (C_1, \dots, C_q)$ as a surrogate for the "density" of \mathbf{X} :

$$f_{\mathbf{X}}^{(q)}(\mathbf{x}) = \prod_{j=1}^q f_{C_j}(c_j(\mathbf{x})), \quad (12)$$

Jacques and Preda (2012) use successfully a similar density surrogate for the clustering of univariate functional data.

4. A model based-clustering for multivariate functional data

The aim of model-based clustering is to identify homogeneous groups of data sampled from a mixture densities model. In this section, we build a mixture model based on the surrogate (12) for the density of \mathbf{X} .

Let us consider that there exists a latent group variable Z , of K categories (K groups), $Z = (Z_1, \dots, Z_K) \in \{0, 1\}^K$ such that $Z_g = 1$ indicates that the multivariate curve \mathbf{X} belongs to the cluster g , $1 \leq g \leq K$, and $Z_g = 0$ otherwise. For each $i = 1, \dots, n$, let associate to \mathbf{X}_i the corresponding group indicator \mathbf{Z}_i .

In the following we suppose that $\mathbf{X}_{|Z_g=1}$ is such that each $X_{\ell|Z_g=1}$ is a zero-mean stochastic process ($1 \leq \ell \leq p$). The number K of groups is assumed to be known.

4.1. The mixture model

Let assume that each couple $(\mathbf{X}_i, \mathbf{Z}_i)$ is an independent realization of the random vector (\mathbf{X}, \mathbf{Z}) where \mathbf{X} has a density surrogate depending on its group belonging:

$$f_{\mathbf{X}|Z_g=1}^{(q_g)}(\mathbf{x}; \Sigma_g) = \prod_{j=1}^{q_g} f_{C_{j|Z_g=1}}(c_{j,g}(\mathbf{x}); \lambda_{j,g})$$

where q_g is the number of the first principal components retained in the approximation (12) for the group g , $c_{j,g}(\mathbf{x})$ is the j th principal component score of $\mathbf{X}_{|Z_g=1}$ for $\mathbf{X} = \mathbf{x}$, $f_{C_{j|Z_g=1}}$ its probability density and Σ_g the diagonal matrix of the principal components variances $(\lambda_{1,g}, \dots, \lambda_{q_g,g})$. Conditionally to the group, the probability density $f_{C_{j|Z_g=1}}$ of the j th principal component of \mathbf{X} is assumed to be the univariate Gaussian density with zero mean (the principal component are centred) and variance $\lambda_{j,g}$. This assumption is satisfied when $\mathbf{X}_{|Z_g=1}$ is a Gaussian process.

The vector $\mathbf{Z} = (Z_1, \dots, Z_K)$ is assumed to have multinomial distribution $\mathcal{M}_1(\pi_1, \dots, \pi_K)$ of order 1, with π_1, \dots, π_K the mixing proportions ($\sum_{g=1}^K \pi_g = 1$). Under this model we can deduce a surrogate for the unconditional density of \mathbf{X} :

$$f_{\mathbf{X}}^{(q)}(\mathbf{x}; \theta) = \sum_{g=1}^K \pi_g \prod_{j=1}^{q_g} f_{C_{j|Z_g=1}}(c_{j,g}(\mathbf{x}); \lambda_{j,g}) \quad (13)$$

where $\theta = \{(\pi_g, \lambda_{1,g}, \dots, \lambda_{q_g,g})_{1 \leq g \leq K}\}$ and $q = (q_1, \dots, q_K)$. From this density surrogate, we deduce a pseudo likelihood:

$$l^{(q)}(\theta; \mathbf{X}) = \prod_{i=1}^n \sum_{g=1}^K \pi_g \prod_{j=1}^{q_g} \frac{1}{\sqrt{2\pi\lambda_{j,g}}} \exp\left(-\frac{1}{2} \frac{C_{i,j,g}^2(\mathbf{X}_i)}{\lambda_{j,g}}\right) \quad (14)$$

where $C_{i,j,g}(\mathbf{X}_i)$ is the j th principal score of the curve X_i for the group g .

4.2. Parameter estimation

In the unsupervised context the estimation of the mixture model parameters is not so straightforward as in the supervised context since the groups indicators Z_i are unknown. On the one hand, we need to use an iterative algorithm which alternate the estimation of the group indicators, the estimation of the PCA scores for each group and then the estimation of the mixture model parameters. On the other hand, the parameter $q = (q_1, \dots, q_g)$ will be estimated by an empirical method, similar to those used to select the number of components in usual PCA.

A classical way to maximise a mixture model likelihood when data are missing (here the clusters indicators \mathbf{Z}_i) is to use the iterative EM algorithm (McLachlan and Peel, 2000). In this work we use an EM-like algorithm including, between the standard E and M steps, a first step in which the principal components scores of each group are updated and a second one in which the approximation order q are selected. Our EM-like algorithm consists in maximizing the pseudo completed log-likelihood

$$L_c^{(q)}(\theta; \mathbf{X}, \mathbf{Z}) = \sum_{i=1}^n \sum_{g=1}^K Z_{i,g} \left(\log \pi_g + \sum_{j=1}^{q_g} \log f_{C_{j|Z_g=1}}(C_{i,j,g}(\mathbf{X}_i); \lambda_{j,g}) \right).$$

Let $\theta^{(h)}$ be the current value of the estimated parameter at step h , $h \geq 1$.

E step. As the group indicators $Z_{i,g}$'s are unknown, the E step consists in computing the conditional expectation of the pseudo completed log-likelihood:

$$\mathcal{Q}(\theta; \theta^{(h)}) = E_{\theta^{(h)}}[L_c^{(q)}(\theta; \mathbf{X}, \mathbf{Z}) | \mathbf{X} = \mathbf{x}] = \sum_{i=1}^n \sum_{g=1}^K t_{i,g} \left(\log \pi_g + \sum_{j=1}^{q_g} \log f_{C_{j|Z_g=1}}(c_{i,j,g}(\mathbf{x}_i); \lambda_{j,g}) \right)$$

where $c_{i,j,g}(\mathbf{x}_i)$ is the value of the random variable $C_{i,j,g}(\mathbf{X}_i)$ for $\mathbf{X}_i = \mathbf{x}_i$, $t_{i,g}$ is the probability for the multidimensional curve \mathbf{X}_i to belong to the group g conditionally to $C_{i,j,g}(\mathbf{X}_i) = c_{i,j,g}(\mathbf{x}_i)$:

$$t_{i,g} = E_{\theta^{(h)}}[Z_{i,g} | \mathbf{X} = \mathbf{x}] \simeq \frac{\pi_g \prod_{j=1}^{q_g} f_{C_{j|Z_{i,g}=1}}(c_{i,j,g}(\mathbf{x}_i); \lambda_{j,g})}{\sum_{l=1}^K \pi_l \prod_{j=1}^{q_l} f_{C_{j|Z_{i,l}=1}}(c_{i,j,l}(\mathbf{x}_i); \lambda_{j,l})}. \quad (15)$$

The approximation in (15) is due to the use of a surrogate for the density of \mathbf{X} given by (12).

Principal score updating step. The computation of the principal component scores has been described in Section 2.2. Here, the principal component scores $C_{i,j,g}$ of the multidimensional curve \mathbf{X}_i in the group g , is updated depending on the current conditional probability $t_{i,g}$ computed in the previous E step. This computation is carried out by weighting the importance of each curve in the construction of the principal components with the conditional probabilities $T_g = \text{diag}(t_{1,g}, \dots, t_{n,g})$. Consequently, the first step consists in centring the curve \mathbf{X}_i within the group g by subtracting the mean curve computed using the $t_{i,g}$'s: the basis expansion coefficients matrix \mathbf{A} becomes $\mathbf{A}_g = (I_n - \mathbb{I}_n(t_{1,g}, \dots, t_{n,g}))\tilde{\mathbf{A}}$. The j th principal component scores $C_{j,g}$ is then the j th eigenvector of the matrix $\mathbf{A}_g W \mathbf{A}'_g T_g$ associated to the j th eigenvalue $\lambda_{j,g}$.

Group specific dimension q_g estimation step. The estimation of the group specific dimension q_g is an open problem. In this work we propose to use, once the group specific FPCA have been computed at the previous step, the scree-test of Cattell (1966) in order to select each group specific dimension q_g . The advantage of using this test is that one hyperparameter (the threshold of the Cattell scree-test) allows to estimate K approximation orders.

M step. The M step consists in computing the mixture model parameters $\theta^{(h+1)}$ which maximizes $\mathcal{Q}(\theta; \theta^{(h)})$ according to θ . It leads simply to the following estimators

$$\pi_g^{(h+1)} = \frac{1}{n} \sum_{i=1}^n t_{i,g}, \quad \text{and} \quad \lambda_{j,g}^{(h+1)} = \lambda_{j,g}, \quad 1 \leq j \leq q_g,$$

where $\lambda_{j,g}$ is the variance of the j th principal component of the cluster g computed in the principal score updating step.

Convergence and numerical considerations. Notice that the q_g estimation step could lead to modify the number of principal components used in the density surrogate at each iteration of the algorithm. A change of dimension will lead to an artificial change (increase or decrease) in the density surrogate, and then in the pseudo likelihood, without moving the values of the model parameters. Thus, during the iterations of the algorithm, the pseudo likelihood is not necessarily increasing, and we loose the usual properties of the EM algorithm. In particular, the pseudo likelihood is not strictly increasing, and the algorithm can not be stopped at its convergence as usually. We decide to stop the algorithm after a fixed number of iterations, and to choose the best reached solutions. A stopping criterion at the convergence of the classification (same classification for consecutive steps) has also been tested, but the results are more stable with the fixed number of iterations.

Moreover, in order to avoid the convergence to a local maximum, the estimation algorithm is launched several times with random initializations for a small number of iterations, and the best reached solution is used as the initialization point for a longer algorithm (Biernacki, 2004).

5. Numerical experiments

This section is devoted to compare our approach, we will call *Funclust* –as in univariate case (Jacques and Preda, 2012)– with other existing methods. The numerical experiments

are organized in three parts. First, Funclust is compared to clustering methods for univariate functional data using three real datasets. Then Funclust is compared to multivariate functional data clustering methods based on k-means using a simulated dataset in a second part and a climatology dataset in a third part.

5.1. Benchmark study in the case of univariate functional data

The data. Three real datasets are considered: the *Kneading*, *Growth*, and *ECG* datasets. These three datasets, already studied in Jacques and Preda (2012), are plotted on Figure 1.

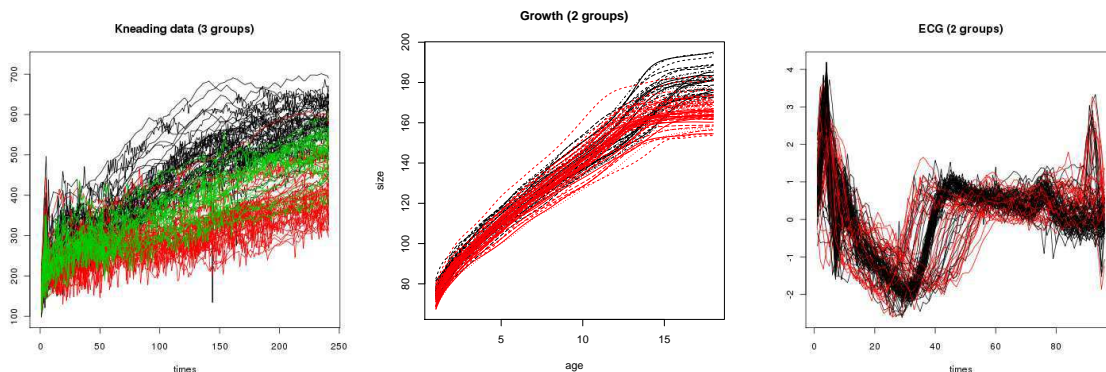


Figure 1: *Kneading*, *Growth* and *ECG* datasets.

The first dataset (*Kneading*) 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 Leveder et al. (2004). There are 115 different flours for which the dough resistance is measured during the kneading process for 480 seconds. 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. These data, have been already studied in a supervised classification context (Leveder et al., 2004; Preda et al., 2007). They are known to be hard to discriminate, even for supervised classifiers, partly because of the medium class. The second dataset (*Growth*) comes from the Berkeley growth study (Tuddenham and Snyder, 1954) and is available in the *fda* package of the software **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 last dataset (*ECG*) is taken from the *UCR Time Series Classification and Clustering* website¹. This dataset consists of 200 curves from 2 groups sampled at 96 time instants, and has already been studied in Olszewski (2001).

¹http://www.cs.ucr.edu/~eamonn/time_series_data/

Experimental setup. For each dataset, the labels indicating the group membership for each observation are available. These labels have been provided by human experts (Kneading and ECG datasets) or simply by the nature of the individuals (gender for the Growth dataset). In order to compare the clustering ability of the studied methods, we choose to use the correct classification rate (CCR) which measures the adequation of the resulting clustering with the known partition. This measure varies between 0 and 1, and larger the CCR is, better the clustering algorithm performs.

In this benchmark study, Funclust is compared to three challengers dedicated to the clustering of functional data: FunHDDC (Bouveyron and Jacques, 2011) and k -centres (kCFC, Chiou and Li (2007)) which are the closest methods and *fclust* James and Sugar (2003) which is known to be a good challenger. Notice that kCFC being not available as a computer code (at our knowledge), only the comparison on the Growth dataset has been possible thanks to the classification results presented in Chiou and Li (2007).

The maximum number of iterations is fixed to 200. Since the iterations corresponding to the retained solutions (according to the best pseudo-likelihood) were always relatively far from the last one, we judge this maximum number of iterations as sufficient. The threshold of the Cattell scree test allowing to select the approximation order q_k is fixed to 0.05. In order to avoid convergence to a local maximum of the pseudo likelihood, our EM-like algorithm has been initialized with the best solutions of 20 small EM-like algorithms with 20 iterations each Biernacki (2004). With this experimental setup, Funclust estimation is obtained in about 30 seconds for each dataset, on an usual laptop and with a code in **R** software.

Results. The estimated approximation orders q_g for Funclust are the following: Kneading ($q_1 = 2, q_2 = 1, q_3 = 3$), Growth ($q_1 = 2, q_2 = 3$), ECG ($q_1 = 9, q_2 = 4$). The correct classification rates (CCR) according to the known partitions are given in Table 1. Funclust performs better than the other methods on two datasets among three (Kneading and ECG). On the last dataset, the results are relatively poor (69.89% whereas some method are about 97%), but they can be greatly increased (95.70%) if the dimensions q_g are fixed to 2. This dataset illustrates that the choice of the approximation order is a very important question, and that further works have to be carried out in this direction.

| method | Kneading | Growth | ECG |
|----------|--------------|--------|-----------|
| Funclust | 66.96 | 69.89 | 84 |
| FunHDDC | 62.61 | 96.77 | 75 |
| fclust | 64 | 69.89 | 74.5 |
| kCFC | - | 93.55 | - |

Table 1: Correct classification rates (CCR) in percentage for Funclust, FunHDDC (best model according BIC), fclust and kCFC on the Kneading, Growth and ECG datasets.

5.2. Simulation study in the case of multivariate functional data

The data. In this simulation study, the number of clusters is assumed to be known: $K=2$. A sample of $n = 50$ curves are simulated according to the following model inspired by Ferraty and Vieu (2003) and Preda (2007):

$$\begin{aligned} \text{Class 1 : } \quad & X_1(t) = -5 + t/2 + U_2 h_3(t) + U_3 h_2(t) + \sqrt{0.1} \varepsilon(t), \quad t \in [1, 21], \\ & X_2(t) = -5 + t/2 + U_1 h_1(t) + U_2 h_2(t) + U_3 h_3(t) + \sqrt{0.5} \varepsilon(t), \quad t \in [1, 21], \\ \text{Class 2 : } \quad & X_1(t) = U_3 h_2(t) + \sqrt{10} \varepsilon(t), \quad t \in [1, 21], \\ & X_2(t) = U_1 h_1(t) + U_3 h_3(t) + \sqrt{0.5} \varepsilon(t), \quad t \in [1, 21], \end{aligned}$$

where $U_1 \sim \mathcal{N}(0.5, 1/12)$, $U_2 \sim \mathcal{N}(0, 1/12)$ and $U_3 \sim \mathcal{N}(0, 2/3)$ are independent Gaussian variables and $\varepsilon(t)$ is a white noise, independent of U_i 's and of unit variance. The function h_1 , h_2 and h_3 (plotted on Figure 2) are defined, for $t \in [1, 21]$, by $h_1(t) = (6 - |t - 11|)_+$ where $(\cdot)_+$ mean the positive part, $h_2(t) = (6 - |t - 7|)_+$ and $h_3(t) = (6 - |t - 15|)_+$.

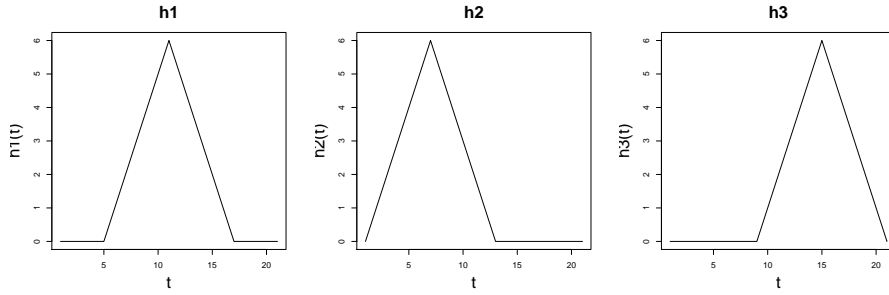


Figure 2: Function $h_1(t) = (6 - |t - 11|)_+$, $h_2(t) = (6 - |t - 7|)_+$ and $h_3(t) = (6 - |t - 15|)_+$ for $t \in [1, 21]$.

The mixing proportions π_i 's are chosen to be equal, and the curves are observed in 1001 equidistant points ($t = 1, 1.02, \dots, 21$). Figure 3 plots the simulated curves.

Experimental setup. Funclust is compared to the multivariate functional data clustering methods described in Ieva et al. (2011), based on k-means method with the following distances:

$$d_1(\mathbf{X}, \mathbf{Y}) = \sqrt{\sum_{j=1}^2 \int_0^T (X_j(t) - Y_j(t))^2 dt} \quad \text{and} \quad d_2(\mathbf{X}, \mathbf{Y}) = \sqrt{\sum_{j=1}^2 \int_0^T (X'_j(t) - Y'_j(t))^2 dt},$$

where $X'_j(t)$ is the derivative of $X_j(t)$. A discussion on these distances in the case of univariate functional data can be found in Ferraty and Vieu (2003). Since no public implementation is available for this method, we built our own implementation in the **R** software. In particular, the distance d_2 was computed using the package *fda*.

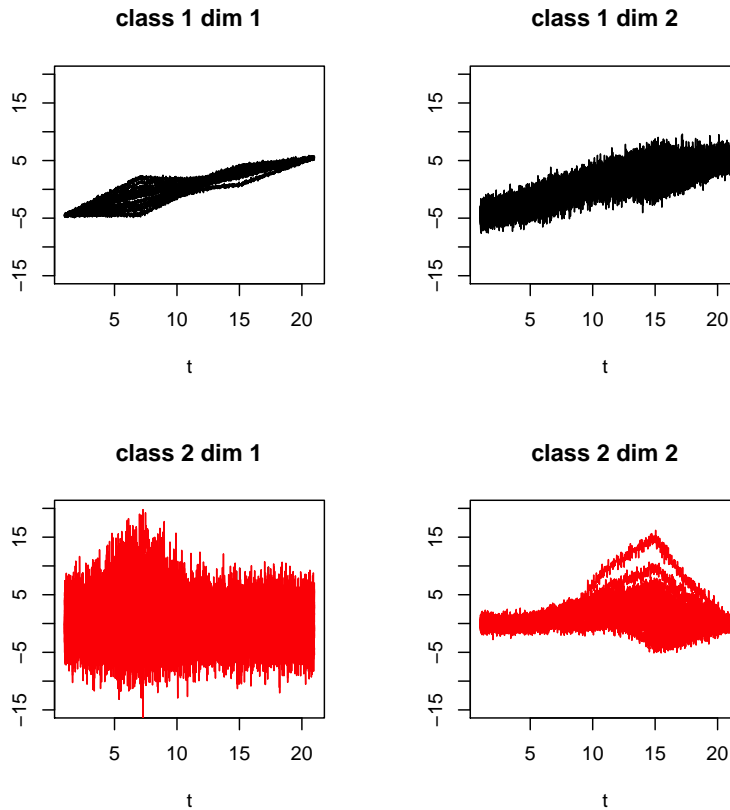


Figure 3: Bi-dimensional simulated curves for class 1 (top) and class 2 (bottom).

In addition to these methods, named $kmeans-d_1$ and $kmeans-d_2$ in the following, we consider usual k-means applied on the values of the functions at each observation points $t = 1, 1.02, \dots, 21$ (quoted as $kmeans-points$) and on the coefficients resulting from a linear spline smoothing with 30 equidistant knots ($kmeans-spline$). Linear spline smoothing has also been used by Funclust in the principal score updating step. The others Funclust setting (Cattell threshold, initialization) are the same as for the previous experiments in the univariate case. Since both components X_1 and X_2 have similar covariance structure, the curves have not been normalized.

Results. Table 2 presents the mean and standard deviation of the correct classification rates (CCR), for 100 simulations. The results confirm the good behaviour of Funclust that we have already noticed for univariate functional data.

5.3. Canadian temperature and precipitation

In this last application, the Canadian temperature and precipitation data (available in the **R** package *fda* and presented in detail in Ramsay and Silverman (2005)) are used to compare Funclust with other clustering methods for multivariate functional data ($kmeans-d_1$ and

| method | Funclust | kmeans- d_1 | kmeans- d_2 | kmeans-points | kmeans-spline |
|----------|--------------|---------------|---------------|---------------|---------------|
| Mean CCR | 86.80 | 86.32 | 85.76 | 80.60 | 86.14 |
| Std CCR | 14.51 | 16.30 | 10.28 | 14.94 | 15.80 |

Table 2: Mean and standard deviation (for 100 simulations) of correct classification rates (CCR) in percentage for Funclust, kmeans- d_1 , kmeans- d_2 and k-means applied on observations points and spline coefficients.

kmeans- d_2). The dataset consists in the daily temperature and precipitation at 35 different locations in Canada averaged over 1960 to 1994. The goal is to provide a classification into 4 groups, and to compare this classification with a natural and intuitive geographical classification.

Since the unit of both curves are different (Celsius degrees and millimetres), the data are normalized following methodology presented at the end of Section 2.3. Figure 4 plots original and reduced temperature and precipitation curves.

Figure 5 presents the classification into 4 groups of the Canadian weather stations using Funclust, with Fourier basis approximation (65 knots) and a Cattell scree test threshold fixed at 0.2. We can observe four distinct groups of stations. The green group is mostly made of northern continental stations, whereas Atlantic stations and southern continental stations are mostly gathered in the red group. The blue group mostly contains Pacific stations and the last group (black) contains only the northernmost station Resolute (N.W.T.). We recall that all these results have been obtained without using the geographical positions of the stations.

In comparison, Figure 6 shows the classification with $kmeans-d_1$ and $kmeans-d_2$ methods, using the same normalized curves and the same basis approximation. This classification seems less pertinent than Funclust classification since the blue and red group contains both Atlantic and Pacific stations. Nevertheless, the black group mainly contains continental stations whereas in Funclust classification, continental and Atlantic stations are gathered together. This last fact is probably due to the Resolute station which is so different from the others (the temperature and precipitation are from far the lowest) that Funclust classifies this station alone in a apart group.

6. Discussion

In this paper we propose a clustering procedure for multivariate functional data based on an approximation of the notion of density for multivariate random function. The main tool is the principal component analysis of multivariate functional data and the use of probability densities of principal components scores. Assuming that the multivariate functional data are sampled from a multivariate Gaussian process, the resulting mixture model is an extrapolation of the finite dimensional Gaussian mixture model to the infinite dimensional setting. In comparison to other clustering techniques for multivariate functional data, our procedure has the advantage to take into account the dependency between the components of the multivariate data. An EM-like algorithm is proposed for the parameter estimation, and applications on real

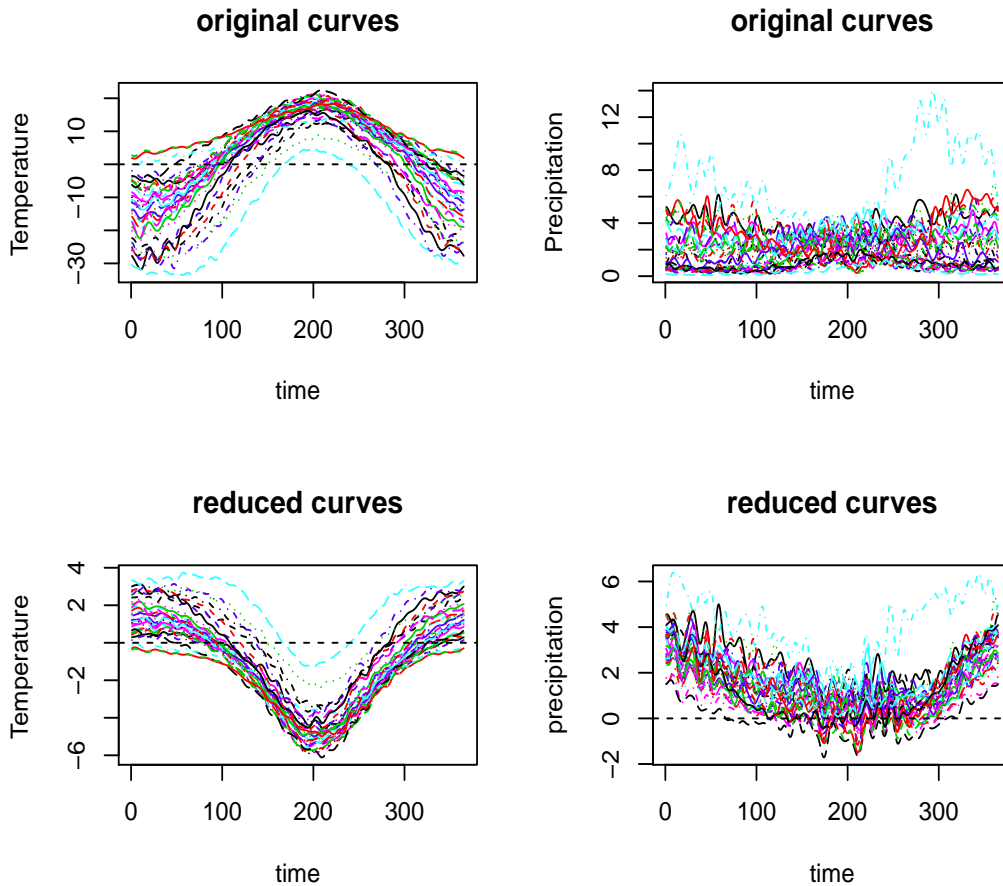


Figure 4: .Temperature and precipitation curves for 35 locations in Canada, averaged over 1960 to 1994. The top figures are the original curves and the bottom figures are the reduced ones.

and simulated data exhibit an interesting clustering accuracy of the proposed method.

Some questions still remain open and further research are to be undertaken to provide answers. First of all, as previously discussed, the selection of the approximation orders is a great challenge for which we actually use an empirical method. Moreover, since only an approximation of the likelihood is available, usual questions such as the selection of the number of clusters or the proofs of the convergence of the estimation algorithm are currently without response.

References

Abraham, C., Cornillon, P. A., Matzner-Løber, E., Molinari, N., 2003. Unsupervised curve clustering using B-splines. *Scandinavian Journal of Statistics. Theory and Applications*



Figure 5: Funclust classification using the reduced curves into 4 groups.

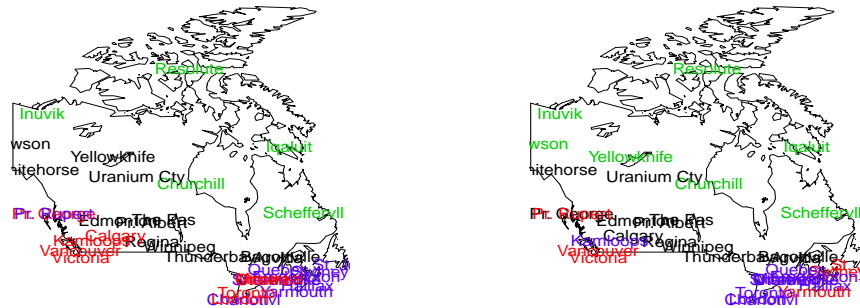


Figure 6: K-means classification using the reduced curves into 4 groups, with distance d_1 (left) and d_2 right.

30 (3), 581–595.

Berrendero, J., Justel, A., Svarc, M., 2011. Principal components for multivariate functional data. *Computational Statistics and Data Analysis* 55, 2619–263.

Besse, P., 1979. Etude descriptive d'un processus. Ph.D. thesis, Université Paul Sabatier,

Toulouse.

- Biernacki, C., 2004. Initializing EM using the properties of its trajectories in Gaussian mixtures. *Statistics and Computing* 14 (3), 267–279.
- Bouveyron, C., Girard, S., Schmid, C., 2007. High Dimensional Data Clustering. *Computational Statistics and Data Analysis* 52, 502–519.
- Bouveyron, C., Jacques, J., 2011. Model-based clustering of time series in group-specific functional subspaces. *Advances in Data Analysis and Classification* 5 (4), 281–300.
- Cattell, R., 1966. The scree test for the number of factors. *Multivariate Behaviour Research* 1 (2), 245–276.
- Chiou, J.-M., Li, P.-L., 2007. Functional clustering and identifying substructures of longitudinal data. *Journal of the Royal Statistical Society. Series B. Statistical Methodology* 69 (4), 679–699.
- Delaigle, A., Hall, P., 2010. Defining probability density for a distribution of random functions. *The Annals of Statistics* 38, 1171–1193.
- Deville, J., 1974. Méthodes statistiques et numériques de l’analyse harmonique. *Annales de l’INSEE* 15, 3–101.
- Ferraty, F., Vieu, P., 2003. Curves discrimination: a nonparametric approach. *Computational Statistics and Data Analysis* 44, 161–173.
- Frühwirth-Schnatter, S., Kaufmann, S., 2008. Model-based clustering of multiple time series. *Journal of Business and Economic Statistics* 26, 78–89.
- Ieva, F., Paganoni, A., Pigoli, D., Vitelli, V., 2011. ECG signal reconstruction, landmark registration and functional classification. In: *7th Conference on Statistical Computation and Complex System*. Padova.
- Jacques, J., Preda, C., 2012. Model-based clustering of functional data. In: *20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. Bruges, pp. 459–464.
- James, G., Sugar, C., 2003. Clustering for sparsely sampled functional data. *J. Amer. Statist. Assoc.* 98 (462), 397–408.
- Kayano, M., Dozono, K., Konishi, S., 2010. Functional Cluster Analysis via Orthonormalized Gaussian Basis Expansions and Its Application. *Journal of Classification* 27, 211–230.
- Leverdier, C., Abraham, C., Cornillon, P. A., Matzner-Løber, E., Molinari, N., 2004. Discrimination de courbes de pétrissage. In: *Chimiométrie 2004*. Paris, pp. 37–43.

- McLachlan, G., Peel, D., 2000. *Finite Mixture Models*. Wiley Interscience, New York.
- Olszewski, R., 2001. *Generalized feature extraction for structural pattern recognition in time-series data*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA.
- Preda, C., 2007. Regression models for functional data by reproducing kernel hilbert spaces methods. *Journal of Statistical Planning and Inference* 137, 829–840.
- Preda, C., Saporta, G., Lévêder, C., 2007. PLS classification of functional data. *Comput. Statist.* 22 (2), 223–235.
- Ramsay, J. O., Silverman, B. W., 2005. *Functional data analysis*, 2nd Edition. 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.
- Sangalli, L., Secchi, P., Vantini, S., Vitelli, V., 2010. K-means alignment for curve clustering. *Computational Statistics and Data Analysis* 54 (5), 1219–1233.
- Saporta, G., 1981. Méthodes exploratoires d'analyse de données temporelles. *Cahiers du Buro* 37–38.
- Singhal, A., Seborg, D., 2005. Clustering multivariate time-series data. *Journal of Chemometrics* 19, 427–438.
- Tarpey, T., Kinateder, K., 2003. Clustering functional data. *J. Classification* 20 (1), 93–114.
- Tokushige, S., Yadohisa, H., Inada, K., 2007. Crisp and fuzzy k-means clustering algorithms for multivariate functional data. *Computational Statistics* 22, 1–16.
- Tuddenham, R., Snyder, M., 1954. Physical growth of california boys and girls from birth to eighteen years. *Univ. Calif. Publ. Chld Devlpmnt* 1, 188–364.