

Model-based clustering of functional data

Julien Jacques and Cristian Preda

Laboratoire Paul Painlevé, UMR CNRS 8524, University Lille I, Lille, France
INRIA Lille-Nord Europe and Polytech'Lille

September 30, 2011

Abstract

Model-based clustering for functional data is considered. An alternative to model-based clustering using the functional principal components is proposed by approximating the density of functional random variables. The EM algorithm is used for parameter estimation and the maximum a posteriori rule provides the clusters. Simulation study and real data application illustrate the interest of the proposed methodology.

Key words: functional data, functional principal component analysis, model-based clustering, random function density, EM algorithm.

1 Introduction

Let X be a functional random variable with values in a functional space \mathcal{F} . For instance, we consider \mathcal{F} is the space of squared integrable functions, $L_2([0, T])$, $T > 0$, and X is a L_2 -continuous stochastic process, $X = \{X_t, t \in [0, T]\}$. Let X_1, \dots, X_n be an i.i.d sample of size n from the same probability distribution as X . Known as *functional data* (see [16]), the observation of X_i 's corresponds to n curves belonging to \mathcal{F} .

The aim of model-based clustering is to identify homogeneous groups of data from a mixture densities model. More precisely, the model-based clustering allows to predict the observation of an indicator vector $Z = (Z_1, \dots, Z_K)$ of the K clusters, such that, conditionally to the belonging to the g th group, $Z_g = 1$, X_i 's comes from a common distribution f indexed by some group-specific parameters, $f(\theta_g)$.

In finite dimensional setting (see for instance [2, 6]), the multivariate probability density function is the main tool for estimating such a model. For functional data, the notion of probability density is not well defined because of the infinite dimension of data. To overcome this difficulty, a pragmatic solution consists in using classical clustering tools, designed for the finite dimensional setting, onto the expansion coefficient of X on some finite basis of functions. The main drawback of such method is that the basis expansion is built independently of the clustering objective. Recent works [10, 4] overcome this problem by defining basis expansion specific to each cluster.

Our work is based on the idea developed in [7] where a "surrogate density" for X is proposed using the Karhunen-Loeve expansion (or principal component analysis (PCA)):

$$X(t) = \mu(t) + \sum_{j=1}^{\infty} C_j \psi_j(t), \quad (1)$$

where μ is the mean function of X , $C_j = \int_0^T (X_t - \mu(t)) \psi_j(t) dt$, $j \geq 1$, are zero-mean random variables (called principal components) and ψ_j 's form an orthonormal system of eigen-functions of the covariance operator of X :

$$\int_0^T \text{Cov}(X_t, X_s) \psi_j(s) ds = \lambda_j \psi_j(t), \forall t \in [0, T].$$

Notice that the principal components C_j 's are uncorrelated random variables of variance λ_j . Considering the principal components indexed upon the descending order of the eigenvalues ($\lambda_1 \geq \lambda_2 \geq \dots$), let denote by $X^{(q)}$ the approximation of X by truncating (1) at the q first terms, $q \geq 1$,

$$X^{(q)}(t) = \mu(t) + \sum_{j=1}^q C_j \psi_j(t). \quad (2)$$

Then, $X^{(q)}$ is the best approximation of X , under the mean square criterion, among all the approximations of the same type (linear combination of deterministic functions of t with random coefficients, [17]). Denoting by $\|\cdot\|$ the usual norm on $L_2([0, T])$, we have

$$\mathbb{E}(\|X - X^{(q)}\|^2) = \sum_{j \geq q+1} \lambda_j \quad \text{and} \quad \|X - X^{(q)}\| \xrightarrow[q \rightarrow \infty]{\text{m.s.}} 0. \quad (3)$$

Without loss of generality, we will suppose in the following that X is a zero-mean stochastic process, *i.e.* $\mu(t) = 0$, $\forall t \in [0, T]$.

Based on the approximation of X by $X^{(q)}$, [7] show that the probability of X to belong to a ball of radius h centered in $x \in L_2[0, T]$ can be written as

$$\log P(\|X - x\| \leq h) = \sum_{j=1}^q \log f_{C_j}(c_j(x)) + \xi(h, q(h)) + o(q(h)), \quad (4)$$

where f_{C_j} is the probability density of C_j and $c_j(x)$ is the j th principal component score of x , $c_j(x) = \langle x, \psi_j \rangle_{L_2}$. 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 (4) suggests the use of the multivariate probability density of the vector $C^{(q)} = (C_1, \dots, C_q)$ as an approximation for the "density" of X . Moreover, observe that we have, $\forall h > 0$, $x \in L_2[0, T]$,

$$P(\|X^{(q)} - x\| \leq h - \|X - X^{(q)}\|) \leq P(\|X - x\| \leq h) \leq P(\|X^{(q)} - x\| \leq h + \|X - X^{(q)}\|). \quad (5)$$

The relation (3) and (5) suggest also that the probability $P(\|X - x\| \leq h)$ could be approximated by $P(\|X^{(q)} - x\| \leq h)$.

Let denote by $f_X^{(q)}$ the joint probability density of $C^{(q)}$. If $x = \sum_{j \geq 1} c_j(x)\psi_j$ and $x^{(q)} = \sum_{j=1}^q c_j(x)\psi_j$ then

$$P(\|X^{(q)} - x\| \leq h) = \int_{\mathcal{D}_x^{(q)}} f_X^{(q)}(y) dy, \quad (6)$$

where $\mathcal{D}_x^{(q)} = \{y \in \mathbb{R}^q : \|y - x^{(q)}\|_{\mathbb{R}^q} \leq \sqrt{h^2 - \sum_{j \geq q+1} c_j^2(x)}\}$.

When X is a gaussian process, the principal components C_j are gaussian and independent. The density $f_X^{(q)}$ is then:

$$f_X^{(q)}(x) = \prod_{j=1}^q f_{C_j}(c_j(x)). \quad (7)$$

We use the functional defined by (7) to develop our model-based clustering methodology for functional data. Our approach is different of that consisting to perform classical model-based clustering on the first q principal components of X .

The paper is organized as follows. In Section 2 we define the model underlying the functional data and describe the parameter estimation procedure for the model-based clustering procedure. The choice of the approximation order q and the definition of the clustering rule are described. In Section 3 we present a simulation study as well as an application on real data (Danone) and compare our results with those provided by other clustering methods.

2 Model-based clustering for functional data

In the following we suppose that X is a zero-mean gaussian stochastic process. Let $\underline{X} = (X_1, \dots, X_n)$ be an i.i.d sample of size n of X and Z be a latent categorical random variable of dimension K , $1 \leq K < \infty$, associated to the K clusters X_i 's belong. For each $i = 1, \dots, n$, let associate to X_i the corresponding categorical variable Z_i indicating the group X_i belongs : $Z_i = (Z_{i,1}, \dots, Z_{i,K}) \in \{0, 1\}^K$ is such that $Z_{i,g} = 1$ if X_i belongs to the cluster g , $1 \leq g \leq K$, and 0 otherwise.

In a clustering setting, the X_i 's variables are observed but not the Z_i 's. The goal is to predict the Z_i 's knowing the X_i 's. For this, we define a parametric mixture model based on the approximation (7) of the density of a random function.

2.1 The mixture model

Let assume that each couple (X_i, Z_i) is an independent realization of the random vector (X, Z) where X has an approximated density depending on its group belonging:

$$f_{X|Z_g=1}^{(q_g)}(x; \Sigma_g) = \prod_{j=1}^{q_g} f_{C_j|Z_g=1}(c_{j,g}(x); \sigma_{j,g}^2)$$

where q_g is the number of the first principal components retained in the approximation (7) for the group g , $c_{j,g}(x)$ is the j th principal component score of $X|_{Z_g=1}$ for $X = x$, $f_{C_{j|Z_g=1}}$ its probability density and Σ_g the diagonal matrix $\text{diag}(\sigma_{1,g}^2, \dots, \sigma_{q_g,g}^2)$.

Conditionally on the group, the probability density $f_{C_{j|Z_g=1}}$ of the j th principal component of X is the univariate gaussian density with zero mean (the principal component are centered) and variance $\sigma_{j,g}^2$.

The vector $Z = (Z_1, \dots, Z_K)$ is assumed to have one order multinomial distribution

$$Z \sim \mathcal{M}_1(\pi_1, \dots, \pi_G)$$

with π_1, \dots, π_K the mixing probabilities ($\sum_{g=1}^K \pi_g = 1$). Under this model it follows that the unconditional (approximated) density of X is given by

$$f_X^{(q)}(x; \theta) = \sum_{g=1}^K \pi_g \prod_{j=1}^{q_g} f_{C_{j|Z_g=1}}(c_{j,g}(x); \sigma_{j,g}^2) \quad (8)$$

where $\theta = (\pi_g, \sigma_{1,g}^2, \dots, \sigma_{q_g,g}^2)_{1 \leq g \leq K}$ have to be estimated and $q = (q_1, \dots, q_K)$. As in the finite dimensional setting, we define an *approximated likelihood* of the sample of curves \underline{X} by:

$$l^{(q)}(\theta; \underline{X}) = \prod_{i=1}^n \sum_{g=1}^K \pi_g \prod_{j=1}^{q_g} \frac{1}{\sqrt{2\pi}\sigma_{j,g}} \exp -\frac{1}{2} \left(\frac{C_{i,j,g}}{\sigma_{j,g}} \right)^2 \quad (9)$$

where $C_{i,j,g}$ is the j th principal score of the curve X_i belonging to the group g .

2.2 Parameter estimation

In the unsupervised context the estimation of the mixture model parameters is not as 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 must be estimated by an empirical method, similar to those used to select the number of components in usual PCA.

2.2.1 Mixture model and component scores estimation

A classical way to maximise a mixture model likelihood when data are missing (here the clusters indicators Z_i) is to use the iterative EM algorithm [8, 12, 13]. In this work we use an EM like algorithm for the maximization of the approximated likelihood (9). This algorithm includes, between the standard E and M steps, a step in which the principal components scores of each group are updated.

The EM algorithm consists in maximizing the approximated completed log-likelihood

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

which is known to be easier to maximise than its incomplete version (9), and leads to the same estimate. Let $\theta^{(h)}$ be the current value of the estimated parameter at step h , $h \geq 1$.

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

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

where $t_{i,g}$ is the probability for the curve X_i to belong to the group g conditionally to $C_{i,j,g} = c_{i,j,g}$:

$$t_{i,g} = E_{\theta^{(h)}}[Z_{i,g} | \underline{X} = \underline{x}] \simeq \frac{\pi_g \prod_{j=1}^{q_g} f_{C_{j|Z_{i,g}=1}}(c_{i,j,g})}{\sum_{l=1}^K \pi_l \prod_{j=1}^{q_g} f_{C_{j|Z_{i,g}=1}}(c_{i,j,g})}. \quad (10)$$

The approximation in (10) is due to the use of an approximation of the density of X .

Principal score updating step. The computation of FPCA eigenfunctions and scores within a given cluster follows [16]. In general, this computation needs some approximation. The most usual one is to assume that the curve admits an expansion into a basis of functions $\phi = (\phi_1, \dots, \phi_L)$. Let Γ be the $n \times L$ expansion coefficients matrix and $W = \int \phi \phi'$ be the matrix of the inner products between the basis functions. Here, the computation of the principal component scores $C_{i,j,g}$ of the curve X_i in the group g is updated depending of the current conditional probability $t_{i,g}$ computed in the previous E step. This computation is carried out by ponderating 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 centering the curve X^i within the group g by subtraction of the mean curve computed using the $t_{i,g}$'s. The principal component scores $C_{i,j,g}$ are then given by

$$C_{i,j,g} = (\lambda_{j,g})^{-1/2} \gamma_{i,g} W \beta_{j,g}$$

where $\beta_{j,g} = W^{-1/2} \mathbf{u}_{j,g}$, $\mathbf{u}_{j,g}$ and $\lambda_{j,g}$ being the j th eigenvector and respectively eigenvalue of the matrix $n^{-1} W^{1/2} \Gamma' T_g \Gamma W^{1/2}$.

Group specific dimension q_g estimation step. The estimation of the group specific dimension q_g is an open problem. It can not be solved by the use of such likelihood-based method. Indeed, the approximation of the density (7) is the product of the density of the q first principal component scores. Therefore, when the density distributions of the principal components are not too peaked (variance lower than $(2\pi)^{-1}$ for gaussian densities), their values are lower than 1, and then the likelihood necessarily decreases when q grows.

In this work we propose to use, once the group specific FPCA have been computed, classical empirical criteria as the proportion of the explained variance or the scree-test of Cattell [5] in order to select each group specific dimension q_g .

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

$$\pi_g^{(h+1)} = \frac{1}{n} \sum_{i=1}^n t_{i,g}, \quad \text{and} \quad \sigma_{j,g}^{2(h+1)} = \lambda_{j,g}, \quad 1 \leq j \leq q_g$$

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

The EM algorithm stops when the difference of the approximated likelihood value of two consecutive steps is lower than a given threshold ϵ (typically $\epsilon = 10^{-6}$).

2.2.2 Model selection

We provided an EM procedure for fitting the model-based clustering for functional data. However, there is a discrete parameter to estimate: the number K of clusters. We propose to use an approximation of the BIC criterion [18] built from the approximated log-likelihood (9):

$$BIC^{(q)} = 2 \log l^{(q)}(\hat{\theta}; \underline{X}) - \nu \log n,$$

where $\nu = 2 * K - 1$ is the number of parameters of the model (mixing proportions and principal scores variances) and $l^{(q)}(\hat{\theta}; \underline{X})$ is the maximum achieved by the likelihood. The number K of clusters maximizing this criterion could be an appropriate choice.

2.3 Classification step

Once the mixture model parameters have been estimated, we proceed to the classification of the observed curves in order to complete our clustering approach. The group belonging can be estimated by the rule of maximum a posteriori (MAP), which consists in classifying a curve x^i into the group g maximizing the conditional probability $P(Z_{ig} = 1 | X_i = x_i)$. At the convergence of the EM algorithm, this probability is given by (10).

Link with related methods. If the principal component scores of each curve are not computed conditionally to their group belonging (here the FPCA are carried out by group), then our approach corresponds exactly to a Gaussian mixture model on the principal component scores. The closest method to our approach is that proposed in [4] (called *fun-HDDC*), which assumes, conditionally to the group, a Gaussian mixture model on the coefficients of the eigen-function expansion. Our approach is different since we assume a Gaussian distribution for the principal component scores, which is true if the curves are sample paths of a Gaussian process. This is a reasonable hypothesis.

3 Numerical experiments

In order to compare our model (quoted in the following by *funclust*) to other approaches, a simulation study and an application on real data are presented in this section. The

simulation study allows to compare *funclust* to the usual clustering procedures, *kmeans* and gaussian mixture model (GMM, [2, 6], through the **R** package *mclust*) applied directly on the FPCA scores. The application on real data consists in clustering Danone kneading curves. We illustrate the accuracy of *funclust* with respect to usual clustering methods such as HDDC [3], MixtPPCA [19], *kmeans*, GMM [2, 6] and hierarchical clustering (*hclust*, **R** package). All these methods are successively applied on the discretized data, on the expansion coefficients in a natural cubic splines basis and on the functional PCA scores. For both, simulation study and application, the number of clusters is assumed to be known.

3.1 Simulation study

In this simulation, the number of clusters is assumed to be known: $K=2$. A sample of $n = 100$ curves are simulated according to the following model inspired by [9, 14]:

$$\begin{aligned} \text{Class 1} & : & X(t) &= U_1 h_1(t) + U_2 h_2(t) + \epsilon(t), & t \in [1, 21], \\ \text{Class 2} & : & X(t) &= U_1 h_1(t) + \epsilon(t), & t \in [1, 21], \end{aligned}$$

where U_1 and U_2 are independent gaussian variables such that $\mathbb{E}[U_1] = \mathbb{E}[U_2] = 0$, $\text{Var}(U_1) = 1/2$, $\text{Var}(U_2) = 1/12$ and $\epsilon(t)$ is a white noise, independent of U_i 's and such that $\text{Var}(\epsilon_t) = 1/12$. The function h_1 and h_2 (plotted on Figure 1) are defined, for $t \in [1, 21]$, by $h_1(t) = 6 - |t - 7|$ and $h_2(t) = 6 - |t - 15|$.

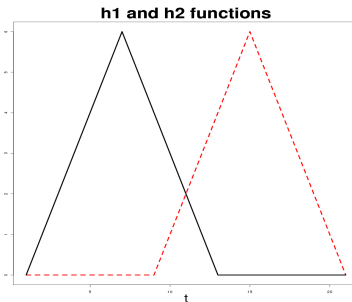


Figure 1: Plots of the functions $h_1(t)$ (solid line) and $h_2(t)$ (dashed line).

The mixing proportions π_i 's are chosen to be equal, and the curves are observed in 101 equidistant points ($t = 1, 1.2, \dots, 21$). Figure 2 plots the simulated curves.

The principal components of X are approximated from $\{X_t\}_{t=1, \dots, 21}$ and are computed using linear spline smoothing (with 30 equidistant knots). For *funclust*, the group specific dimensions q_g are estimated such that 90% of the total variance was explained by the first q_g principal components. For the classical clustering procedures, *kmeans* and *gaussian mixture model* (GMM, [2, 6]), the number of FPCA scores used is selected in the same way. Corresponding dimensions and correct classification rates, averaged on 100 simulations, are given in Table 1.

As we can expect, for this dataset with specific principal spaces of different dimensions, *funclust* outperforms classical clustering methods for the multivariate setting.

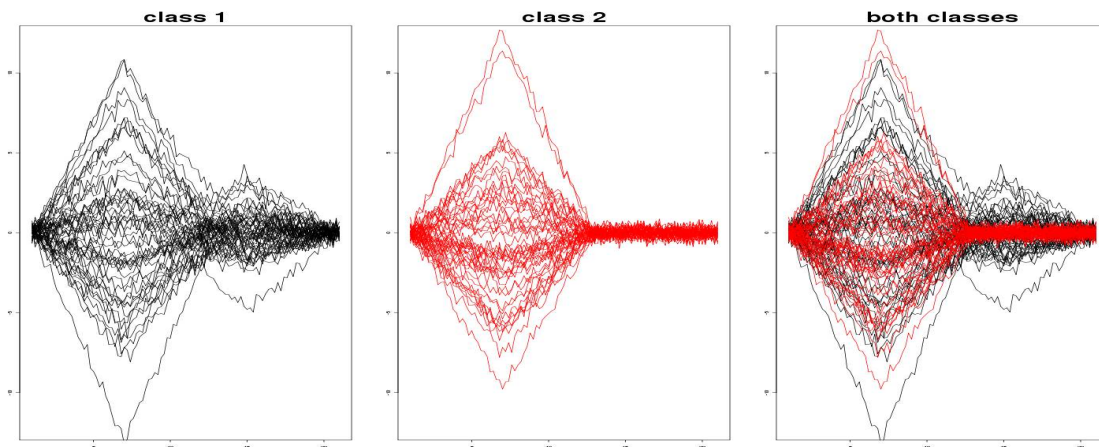


Figure 2: Class 1 (left), Class 2 (center) and both classes (right).

method	correct classif. rate	q_1	q_2
funclust	79.68	1.88	1.90
GMM	56.58	1.10	
kmeans	54.46	1.10	

Table 1: Correct classification rates, group specific dimension q_g for funclust and number of FPCA scores for GMM and kmeans (averaged on 100 simulations), for the simulation study.

3.2 Application

The dataset we use comes from Danone Vitapole Paris Research Center and it concerns the quality of cookies and the relationship with the flour kneading process. The kneading data set is described in detail in [11].

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 *adjustable* quality and 40 bad quality. The Figure 3 presents the set of the 115 kneading curves.

In a supervised classification context, this data is used in [11, 15, 1] for fitting linear and non-parametric prediction models for the cookie's quality. From these studies, it appears that it is difficult to discriminate between the three classes, even for supervised classifiers, partly because of the adjustable class.

Let us consider that the 115 kneading curves are sample-paths of a second order stochastic process X . In order to get the functional feature of data, each curve is approximated using cubic B-spline basis expansion with the following 16 knots [11] : 10, 42, 84, 88, 108, 134, 148, 200, 216, 284, 286, 328, 334, 380, 388, 478. Thus, each curve X_i is represented by a set of 18 coefficients. Therefore, the FPCA of X is approximated using the smoothed curves (for more details, see [16]). The group specific dimensions

q_g are estimated such that at least 95% of the total variance was explained. Resulting dimensions are $q_1 = 2, q_2 = 1, q_3 = 1$.

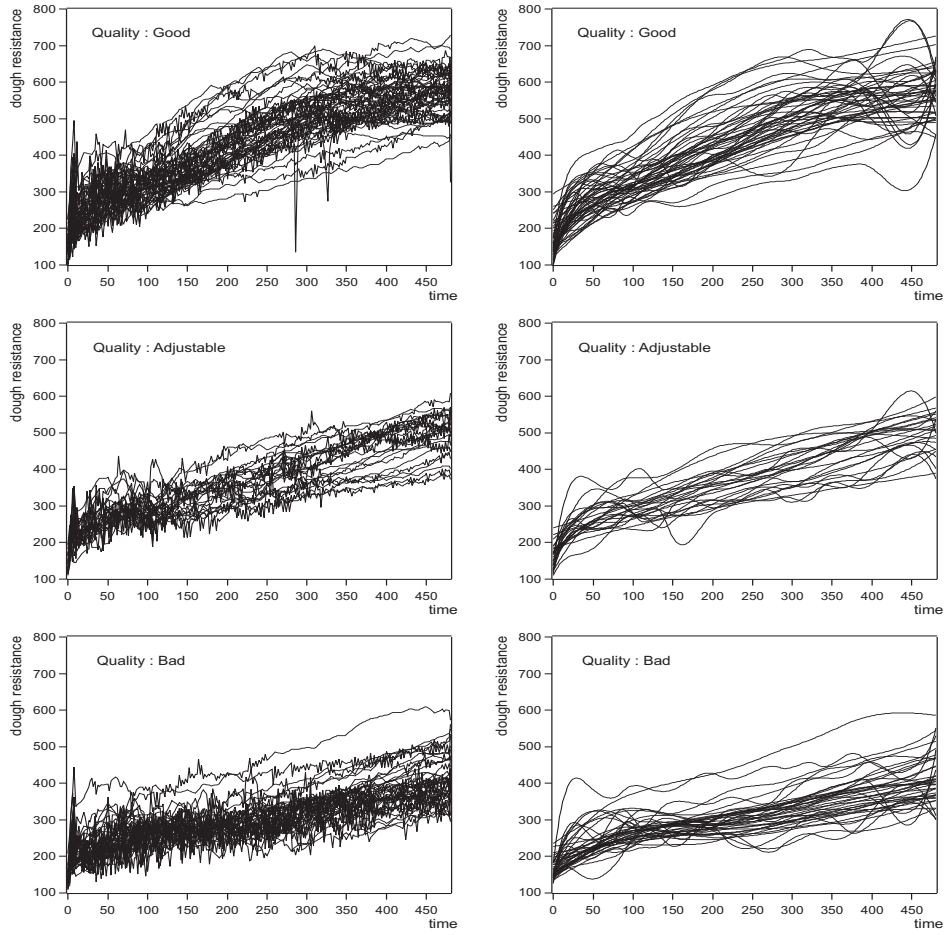


Figure 3: Kneading data : 115 flours observed during 480 seconds. *Left* : observed data. *Right* : smoothed data using cubic B-splines

Table 2 presents the results obtained with different clustering methods. Our method *funclust* performs better than fun-HDDC [4] which similarly to *funclust* considers group specific subspaces but assume a Gaussian mixture model on the coefficients of the eigenfunction expansion, and not on the principal score as *funclust*. The methods from the multivariate finite setting are also outperformed by *funclust*.

4 Conclusion

In this paper we propose a clustering procedure for functional data based on an approximation of the notion of *density of a random function*. The main tool is the use of the probability densities of the principal components scores. Assuming that the functional data are sample of a Gaussian process, the resulting mixture model is an extrapolation

2-steps methods	discretized (241 instants)	spline coeff. (20 splines)	FPCA scores (4 components)	functional methods	
HDCC	66.09	53.91	44.35	fun-HDCC ¹	62.61
MixtPPCA	65.22	64.35	62.61	funclust	67.82
mclust	63.48	50.43	60		
kmeans	62.61	62.61	62.61		
hclust	63.48	63.48	63.48		

Table 2: Percentage of correct classification for the Kneading dataset

of the finite dimensional Gaussian mixture model to the infinite dimensional setting. We defined an EM like algorithm for the parameter estimation and performed a simulation study, as well as an application on real data, in order to show the performance of this approach with respect to other clustering procedures.

The approximation of the density of a random function, based on the principal components densities, opens numerous perspectives for futur works. Indeed, a clustering procedure for multivariate functional data (several curves observed for a same individual) can be defined similarly. The difficult task in such a multivariate functional setting is to define dependence between univariate functions. This challenge can be met by the FPCA of multivariate curve [16].

References

- [1] A.M. Aguilera, M. Escabiasa, C. Preda, and G. Saporta. Using basis expansions for estimating functional pls regression. applications with chemometric data. *Chemometrics and Intelligent Laboratory Systems*, 104(2):289–305, 2011.
- [2] J.D. Banfield and A.E. Raftery. Model-based gaussian and non-gaussian clustering. *Biometrics*, 49:803–821, 1993.
- [3] C. Bouveyron, S. Girard, and C. Schmid. High Dimensional Data Clustering. *Computational Statistics and Data Analysis*, 52:502–519, 2007.
- [4] C. Bouveyron and J. Jacques. Model-based clustering of time series in group-specific functional subspaces. *Advances in Data Analysis and Classification*, in press, 2011.
- [5] R. Cattell. The scree test for the number of factors. *Multivariate Behav. Res.*, 1(2):245–276, 1966.
- [6] G. Celeux and G. Govaert. Gaussian parsimonious clustering models. *The Journal of the Pattern Recognition Society*, 28:781–793, 1995.
- [7] A. Delaigle and P. Hall. Defining pobability density for a distribution of random functions. *The Annals of Statistics*, 38:1171–1193, 2010.

- [8] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 39(1):1–38, 1977.
- [9] F. Ferraty and P. Vieu. Curves discrimination: a nonparametric approach. *Computational Statistics and Data Analysis*, 44:161–173, 2003.
- [10] G.M. James and C.A. Sugar. Clustering for sparsely sampled functional data. *J. Amer. Statist. Assoc.*, 98(462):397–408, 2003.
- [11] C. Lévêder, P.A. Abraham, E. Cornillon, E. Matzner-Lober, and N. Molinari. Discrimination de courbes de pétrissage. In *Chimiométrie 2004*, pages 37–43, Paris, 2004.
- [12] G. McLachlan and T. Krishnan. *The EM algorithm and extensions*. Wiley Interscience, New York, 1997.
- [13] G. McLachlan and D. Peel. *Finite Mixture Models*. Wiley Interscience, New York, 2000.
- [14] C. Preda. Regression models for functional data by reproducing kernel hilbert spaces methods. *Journal of Statistical Planning and Inference*, 137:829–840, 2007.
- [15] C. Preda, G. Saporta, and C. Lévêder. PLS classification of functional data. *Comput. Statist.*, 22(2):223–235, 2007.
- [16] J. O. Ramsay and B. W. Silverman. *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition, 2005.
- [17] G. Saporta. Méthodes exploratoires d’analyse de données temporelles. *Cahiers du Buro*, 37–38, 1981.
- [18] G. Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6(2):461–464, 1978.
- [19] M. E. Tipping and C. Bishop. Mixtures of principal component analyzers. *Neural Computation*, 11(2):443–482, 1999.