Mixture of functional principal component regression

Valérie Garès[1] & Madison Giacofci [2] & Valérie Monbet[3]

[1] *Univ. Rennes INSA, CNRS, IRMAR - UMR 6625, Rennes, France, valerie.gares@insa-rennes.fr*
[2] *Univ. Rennes 2, CNRS, IRMAR - UMR 6625, Rennes, France, joyce.giacofci@univ-rennes2.fr*
[3] *Univ. Rennes, CNRS, IRMAR - UMR 6625, Rennes, France, valerie.monbet@univ-rennes1.fr*

Résumé. Basé sur une application industrielle réelle, nous proposons un modèle de mélange de régression sur composantes principales. Les réponses observées sont fonctionnelles et sont caractérisées par des covariables scalaires. La nature infinie dimensionnelle des observations est prise en compte à travers une analyse en composantes principales fonctionnelle (ACPf). Le lien entre les réponses fonctionnelles et les covariables n'étant pas homogène pour tous les individus, nous considérons un modèle de mélange de régressions. L'originalité de ce travail réside dans le fait que le partitionnement et la prédiction des scores individuels des composantes principales sont réalisés de manière simultanée. Nous adoptons pour cela une approche probabiliste de l'analyse en composantes principales. Les paramètres du modèle sont estimés à l'aide de l'algorithme EM dans lequel à la fois les classes et les scores individuels de l'ACPf sont latents. Des simulations préliminaires et une application sur des données réelles sont présentées pour évaluer la performance de la méthode.

Mots-clés. ACP fonctionnelle, ACP probabiliste, Mélange de régression, Algorithme EM

Abstract. Motivated by a real data application for industry, we develop a mixture model of functional principal regression. The responses observed are functional data that are characterized by multivariate covariates. The infinite-dimension nature of the data is accounted through a funtional principal component analysis (fPCA) approach. The relationship between functional responses and covariates being not homogeneous for all individuals, we consider a mixture of regression model. The originality of this work lies in the simultaneous clustering and prediction of the individual principal components scores. This is achieved by adopting a probabilistic approach of the principal components analysis. Estimation of the parameters of the models is driven by the EM algorithm where both cluster labels and individuals fPCA scores are latent. Preliminary simulations and application to the real data example are presented to assess performance of the proposed model.

Keywords. Functional PCA, Probabilistic PCA, Mixture of regression, EM Algorithm

1 Motivation

The present work is motivated by the study of force curves recorded on pieces of rubber seal caught between two rigid parts and subject to increasing compression loads. Namely, the compression on the rubber seal is recorded in Newton for decreasing gaps between the rigid parts producing a force curve whose shape rely on the geometrical features of the seal as height, thickness or shape. The force curves under study comes from numerical simulations based on equations from material physic. Such simulations are computationally expensive and hence limit the number of geometrical configurations to consider.

Our aim is to propose a statistical model to predict a force curve based on the geometric characteristics of a rubber seal.

2 Mixture of functional principal component regression

Let the observed functional responses $Y_1(t), \ldots, Y_n(t)$ be independent realizations of a L_2-continuous stochastic process $Y = \{Y(t)\}_{t\in[0,T]}$ with finite covariance. The functional principal component analysis (FPCA) explores the main modes of variability in a set of curves (see *e.g* Wang et al. (2016) and references therein).

Using a truncated Karhunen-Loève expansion of the process covariance, individual random curve can be expressed as

$$Y_i(t) = m(t) + \sum_{j=1}^{q} \alpha_{ij}\Psi_j(t) + \varepsilon_i(t), \qquad i = 1, \ldots, n, \tag{1}$$

where $m(t)$ is the mean function, $\Psi_1(t), \ldots, \Psi_q(t)$ are the orthogonal functional principal components (FPCs) associated with the q largest eigenvalues $\lambda_1 \geq \ldots \geq \lambda_q$ whereas α_{ij} is the score of the ith individual on the jth FPC and $\varepsilon_i(t)$ are residual curves.

A common fixed B-spline basis $\varphi_1, \ldots, \varphi_p$ is used to expand functions in model (1), leading model (1) to be expressed in the coefficient domain as

$$\boldsymbol{C}_i = \boldsymbol{\mu} + \boldsymbol{Q}\boldsymbol{\alpha_i} + \boldsymbol{\epsilon}_i, \qquad i = 1, \ldots, n, \tag{2}$$

where $\boldsymbol{C}_i$, $\boldsymbol{\mu}$ and $\boldsymbol{\epsilon}_i$ are $p \times 1$ vectors of spline coefficients associated with, respectively, $Y_i(t)$, $m(t)$ and $\varepsilon_i(t)$. The $q \times p$ matrix $\boldsymbol{Q}$ contains spline coefficients of the FPCs $\{\Psi_j(t)\}_{j=1,\ldots,q}$ stacked in columns and $p \times 1$ residual vectors are assumed to follow a zero-mean Gaussian distribution. Model (2) in the coefficient domain hence resumes to a standard PCA model.

When scalar predictors are available, they can be linked to individual scores through a regression model. Such approach is referenced as function-on-scalar (Chen & al., 2016). Assume we observe scalars covariates $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ir})^T$ for $i = 1, \ldots, n$ that are linked to individual FPC's scores through the following model

$$\boldsymbol{\alpha}_{ij} = \boldsymbol{\beta}\boldsymbol{x}_i + \boldsymbol{\eta}_i,$$

where $\boldsymbol{\beta}$ is the $r \times 1$ vector of regression coefficients and $\boldsymbol{\eta}_i$ the $r \times 1$ vectors of centered uncorrelated measurements errors, independent of $\boldsymbol{\epsilon}_i$.

In real-life applications and in particular for the application that motivates this work, the relationship between functional responses and the scalar predictors can not be assumed to be identical for all subjects. One way to account for such variations is to assume individuals to belong to K clusters in which the mechanisms driving the response may differ. We denote by $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{iK})$ the cluster membership random variable where, for $k = 1, \ldots, K$, Z_{ik} is a binary random variable equal to 1 if individual i is in the group k and 0 otherwise. $\boldsymbol{Z}_i$ is thus assumed to follow a multinomial distribution with parameters 1 and $(\pi_1, \ldots, \pi_K)$. Assuming besides $\boldsymbol{\eta}_i$ and $\boldsymbol{\epsilon}_i$ to be independent Gaussian distributed errors, our final mixture of functional principal component regression is then written, hierarchically, as

$$\begin{cases} \boldsymbol{C}_i | \{Z_{ik} = 1\} \sim \mathcal{N}(\boldsymbol{\mu}_k + \boldsymbol{Q}_k \boldsymbol{\alpha}_{ik}, \sigma^2 \boldsymbol{I}_p) \\ \boldsymbol{\alpha}_{ik} \sim \mathcal{N}(\boldsymbol{\beta}_k \boldsymbol{x}_i, \theta^2 \boldsymbol{I}_q) \end{cases} \tag{3}$$

where Q_k is a $p \times q$ matrix of cluster-specific FPCs, α_{ik} is a $q \times 1$ vector of cluster-specific random scores and β_k is a $r \times q$ matrix of regression coefficients. A related regression model have been considered by Yao et al. (2011) in a scalar-on-function regression framework. However, the particularity of our model lies in the cluster-dependent FPCA leading the curves to be represented in their own functional subspaces, as proposed by Bouveyron & Jacques (2011) in a clustering framework.

3 Estimation

Our model exhibits two types of latent variables as both cluster-membership variables and individual random FPC's scores are not observed. When dealing with mixture models, the expectation-maximisation (EM) algorithm is the far most used algorithm. It consist in maximizing a likelihood. Besides, an alternative point of view on standard PCA, that conceives it as a likelihood-based model was proposed by Tipping & Bishop (1999). It is known as probabilistic PCA and parameters are also estimated thanks to the EM algorithm. Estimation of the parameters in model (3) is thus driven by an EM algorithm where cluster labels and individuals fPCA scores are latent.

Namely, given the coefficients vectors $\{\boldsymbol{C}_i\}_{i=1,\ldots,n}$ of the observed curves, the complete log-likelihood of the data under model (3) decomposes in

$$\begin{aligned} \log \mathcal{L}(\boldsymbol{C}_i, \boldsymbol{\alpha}_i, \boldsymbol{Z}_i; \boldsymbol{\Theta}) = \log \mathcal{L}(\boldsymbol{Z}_i, \pi) + \log \mathcal{L}(\alpha_{ik} | Z_i, \beta_k, \theta^2) \\ + \log \mathcal{L}(C_i | \alpha_{ik}, Z_i, Q_k, \mu_k, \sigma^2), \end{aligned} \tag{4}$$

where $\boldsymbol{\Theta}$ denotes the set of all model parameters. Direct maximisation of the complete log-likelihood is untractable and is done through the EM algorithm. We next describe the two steps of the algorithm.

E-step The E-step consists in replacing the unobserved variables by their conditional expectations. Hence, posterior cluster membership probabilities are updated with

$$\tau_{ik} = E[Z_{ik}|\boldsymbol{C}_i] = P(Z_{ik} = 1|\boldsymbol{C}_i) = \frac{\pi_k f(\boldsymbol{C}_i\,;\,\boldsymbol{\mu}_k + \boldsymbol{Q}_k\boldsymbol{\beta}_k\boldsymbol{x}_i\,,\,\theta^2\boldsymbol{Q}_k\boldsymbol{Q}_k^T + \sigma^2\boldsymbol{I}_p)}{\sum_\ell \pi_\ell f(\boldsymbol{C}_i\,;\,\boldsymbol{\mu}_\ell + \boldsymbol{Q}_\ell\boldsymbol{\beta}_\ell\boldsymbol{x}_i\,,\,\theta^2\boldsymbol{Q}_\ell\boldsymbol{Q}_\ell^T + \sigma^2\boldsymbol{I}_p)},$$

where $f(\,\cdot\,;\boldsymbol{\mu},\boldsymbol{\Sigma})$ is the density of a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Computation of the expected likelihood requires also conditional expectations depending of individual scores as $E[\boldsymbol{\alpha}_{ik}|\boldsymbol{C}_i]$, $E[\boldsymbol{\alpha}_{ik}^T\boldsymbol{\alpha}_{ik}|\boldsymbol{C}_i]$, $E[\boldsymbol{\alpha}_{ik}\boldsymbol{\alpha}_{ik}^T|\boldsymbol{C}_i]$ and $E[\boldsymbol{\alpha}_{ik}^T Q_k^T Q_k\boldsymbol{\alpha}_{ik}|\boldsymbol{C}_i])$. Thanks to conditioning of multivariate Gaussian distribution and basic matrix properties, we have

$$\begin{aligned}
\widehat{\boldsymbol{\alpha}}_{ik} = E[\boldsymbol{\alpha}_{ik}|\boldsymbol{C}_i] &= \boldsymbol{\beta}_k\boldsymbol{x}_i + (\theta^2\boldsymbol{Q}_k^T)(\theta^2\boldsymbol{Q}_k\boldsymbol{Q}_k^T + \sigma^2\boldsymbol{I}_p)^{-1}(\boldsymbol{C}_i - \boldsymbol{\mu}_k - \boldsymbol{Q_k}\boldsymbol{\beta}_k\boldsymbol{x}_i)\\
E(\boldsymbol{\alpha}_{ik}^T\boldsymbol{\alpha}_{ik}|\boldsymbol{C}_i) &= \mathrm{Tr}\left(\theta^2\left(\boldsymbol{I}_q + \theta^2/\sigma^2\boldsymbol{Q}_k^T\boldsymbol{Q}_k\right)^{-1}\right) + E(\boldsymbol{\alpha}_{ik}|\boldsymbol{C}_i)^T E(\boldsymbol{\alpha}_{ik}|\boldsymbol{C}_i)\\
E(\boldsymbol{\alpha}_{ik}^T\boldsymbol{Q}_k^T\boldsymbol{Q}_k\boldsymbol{\alpha}_{ik}|\boldsymbol{C}_i) &= \mathrm{Tr}\left(\theta^2\boldsymbol{Q}_k\left(\boldsymbol{I}_q + \theta^2/\sigma^2\boldsymbol{Q}_k^T\boldsymbol{Q}_k\right)^{-1}\boldsymbol{Q}_k^T\right) + E(\boldsymbol{\alpha}_{ik}|\boldsymbol{C}_i)^T\boldsymbol{Q}_k^T\boldsymbol{Q}_k E(\boldsymbol{\alpha}_{ik}|\boldsymbol{C}_i)\\
E[\boldsymbol{\alpha}_{ik}\boldsymbol{\alpha}_{ik}^T|\boldsymbol{C}_i] &= \theta^2\left(\boldsymbol{I}_q + \theta^2/\sigma^2\boldsymbol{Q}_k^T\boldsymbol{Q}_k\right)^{-1} + E(\boldsymbol{\alpha}_{ik}|\boldsymbol{C}_i)E(\boldsymbol{\alpha}_{ik}|\boldsymbol{C}_i)^T
\end{aligned}$$

M-step M-step of the algorithm consists then in maximising the expectation of the complete likelihood (4) and lead to closed-form formulas. For a given iteration h, prior cluster sizes are updated in a standard way by $\widehat{\pi}_k^{[h]} = \frac{1}{n}\sum_i \tau_{ik}$. Regression coefficients and residual variance are updated with

$$\begin{aligned}
\widehat{\beta}_k^{[h]} &= \left[\sum_i \tau_{ik}(\boldsymbol{x}_i\boldsymbol{x}_i^T)\right]^{-1}\sum_i \tau_{ik}\boldsymbol{x}_i\widehat{\boldsymbol{\alpha}}_{ik}^T\\
\widehat{\theta^2}^{[h]} &= \frac{1}{nq}\sum_{i,k}\tau_{ik}(E[\boldsymbol{\alpha}_{ik}^T\boldsymbol{\alpha}_{ik}|\boldsymbol{C}_i] - 2\boldsymbol{x}_i^T\boldsymbol{\beta}_k^T E[\boldsymbol{\alpha}_{ik}|\boldsymbol{C}_i] + \boldsymbol{x}_i^T\boldsymbol{\beta}_k^T\boldsymbol{\beta}_k\boldsymbol{x}_i)
\end{aligned}$$

Then, group-specific parameters related to latent FPCA, as the principal components coefficients, functional mean coefficients and residual variance, are updated through

$$\begin{aligned}
\widehat{\boldsymbol{Q}}_k^{[h]} &= \left[\sum_i \tau_{ik}E[\boldsymbol{\alpha}_{ik}\boldsymbol{\alpha}_{ik}^T|\boldsymbol{C}_i]\right]^{-1}\sum_i \tau_{ik}(\boldsymbol{C}_i - \boldsymbol{\mu}_k)\widehat{\boldsymbol{\alpha}}_{ik}^T\\
\widehat{\boldsymbol{\mu}}_k^{[h]} &= \left[\sum_i \tau_{ik}\right]^{-1}\sum_i \tau_{ik}(\boldsymbol{C}_i - \boldsymbol{Q}_k\widehat{\boldsymbol{\alpha}}_{ik})\\
\widehat{\sigma^2}^{[h]} &= \frac{1}{np}\sum_{i,k}\tau_{ik}\left((\boldsymbol{C}_i - \boldsymbol{\mu}_k)^T(\boldsymbol{C}_i - \boldsymbol{\mu}_k) - 2(\boldsymbol{C}_i - \boldsymbol{\mu}_k)^T\boldsymbol{Q}_k E[\boldsymbol{\alpha}_{ik}|\boldsymbol{C}_i] + E[\boldsymbol{\alpha}_{ik}^T\boldsymbol{Q}_k^T\boldsymbol{Q}_k\boldsymbol{\alpha}_{ik}|\boldsymbol{C}_i]\right)
\end{aligned}$$

Finally, the model includes hyperparameters to be adjusted, namely the number of clusters K and the number of principal components q. The number of clusters is specified through the BIC criteria whereas the number of prinicpal components retained in the model is chosen through the usual empirical approach based on the ratio of the variance explained.

4 Simulations

Data generation. We generate $B = 100$ synthetic datasets with $n = 10000$ individuals under the model given by equation (3). In first attempt, we select a large number of individuals to discard the influence of the sample size. We choose $K = 3$ groups with prior sizes $(\pi_1, \pi_2, \pi_3) = (0.2, 0.35, 0.45)$. Cluster-specific means $\boldsymbol{\mu}_k = (\mu_{k,1}, \ldots, \mu_{k,p})$ for $k = 1, 2, 3$ and $p = 10$ are chosen arbitrarily with $\mu_{1,j} = (j-1)^2/20$, $\mu_{2,j} = 2 \times \cos((j-1)/2) + 1$, and $\mu_{3,j} = j$, for $j = 1, \ldots, p$. Each element of the $r \times q$ matrix β_k, with $r = q = 4$ is fixed using a random draw from a zero mean Gaussian distribution and standard deviation equal to 2. Level of variances θ^2 and σ_k^2, for $k = 1, 2, 3$ are controlled using signal-to-noise ratios (SNR), defined as the empirical signal variance divided by the variance of the noise, equal to 3 and 100 respectively.

For each group k, let $\boldsymbol{\Sigma}^k$ a $p \times p$ matrix with $\boldsymbol{\Sigma}^k_{i,i} = 1$, $\boldsymbol{\Sigma}^k_{i,i+1} = \boldsymbol{\Sigma}^k_{i+1,i} = s_1^k$ for $i = 1, \ldots, p-1$ and $\boldsymbol{\Sigma}^k_{i,i+1} = \boldsymbol{\Sigma}^k_{i+1,i} = s_2^k$ for $i = 1, \ldots, p-2$ while other coefficients of the matrix $\boldsymbol{\Sigma}^k$ are zero. Let $\tilde{\boldsymbol{Q}}_k$ be the eigen vectors of $\boldsymbol{\Sigma}^k$ and $\tilde{\boldsymbol{C}}_{k,i}$ generated given equation (3), for $i = 1, \ldots, n$. $\tilde{\boldsymbol{C}}_{k,i}$ plays the role of a tool dataset used to properly define the decomposition of the matrix $\boldsymbol{Q}_k$. Given the singular value decomposition of $\tilde{\boldsymbol{C}}_k$ written as $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^T$, we then choose $\boldsymbol{Q}_k = \boldsymbol{U}\left(\boldsymbol{\Lambda} - \sigma^2\boldsymbol{I}\right)^{1/2}$. We fix $s_1^1 = 0.7$, $s_2^1 = -0.4$, $s_1^2 = 0.7$, $s_2^2 = 0.4$ and $s_1^3 = 0.8$, $s_2^3 = 0.2$.

Initial values of the algorithm. The k-means algorithm is applied on vector C for an initial clustering of individuals in 3 groups and deduce initial values of the parameters in each cluster from empirical estimates. For the matrix $\boldsymbol{Q}_k$, we choose $\boldsymbol{U}\left(\boldsymbol{\Lambda} - \sigma^2\boldsymbol{I}\right)^{1/2}$ with $\boldsymbol{U}$ and $\boldsymbol{\Lambda}$ from being the singular value decomposition of $\boldsymbol{C}_i - \boldsymbol{\mu}_k$ for individuals i belonging to group k.

Performance criteria. Results were assessed in terms of the following criteria, for vector γ of size p_γ :

1. Empirical bias of parameters: $\frac{1}{p_\gamma}\left\|\frac{1}{B}\sum_{b=1}^{B}\widehat{\gamma}_b - \gamma\right\|_2$,
2. Standard error: $\sqrt{\frac{1}{B}\sum_{b=1}^{B}\frac{1}{p_\gamma}\left\|\widehat{\gamma}_b - \frac{1}{B}\sum_{b=1}^{B}\widehat{\gamma}_b\right\|_2^2}$.

We can precise the root-mean-square error is the sum of these two squared terms.

Results. In table 1, empirical bias of parameters and associated standard errors are presented. The different value of bias and standard errors are correct.

For more global and visual evaluation, we also compared the empirical covariance matrix of the true generated dataset $\boldsymbol{C}$ and the empirical covariance matrix of reconstituted dataset

Parameter	Group	Bias	Standard error
μ	1	0.18	0.88
	2	0.18	1.01
	3	0.07	0.74
β	1	0.61	4.17
	2	0.63	4.48
	3	0.64	4.59
σ	1	0.27	0.11
	2	0.34	0.08
	3	0.42	0.09
θ	-	0.11	0.69

Table 1: Bias and standard error of the different parameters

$\boldsymbol{C}^R$ generated with estimated parameters given equation (3). In figure 1, we display the mean of these empirical covariances over the $B = 100$ datasets. The two covariance figures look reasonably similar.

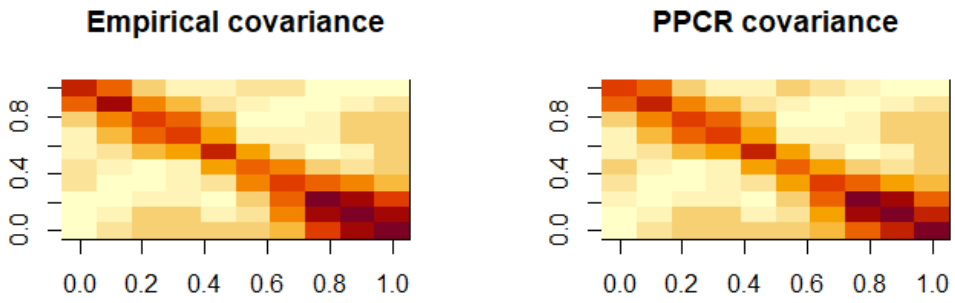

Figure 1: Left panel : Empirical covariance matrix of the true generated dataset $\boldsymbol{C}$. Right panel : the empirical covariance matrix of reconstituted dataset $\boldsymbol{C}^R$

5 Example on efforts curves

The model described above is applied to the efforts curves described in the introduction. This part is still a work in progress.

Individual efforts curves ($n = 498$) are measured in Newton on a common grid of regularly-spaced gaps $\boldsymbol{g} = \{7, 7.5, 8, \ldots, 30\}$, separating the two rigid pieces that borders the rubber seal. The observed curves are displayed in Figure 2. In addition, each curve is characterized by three geometrical parameters :

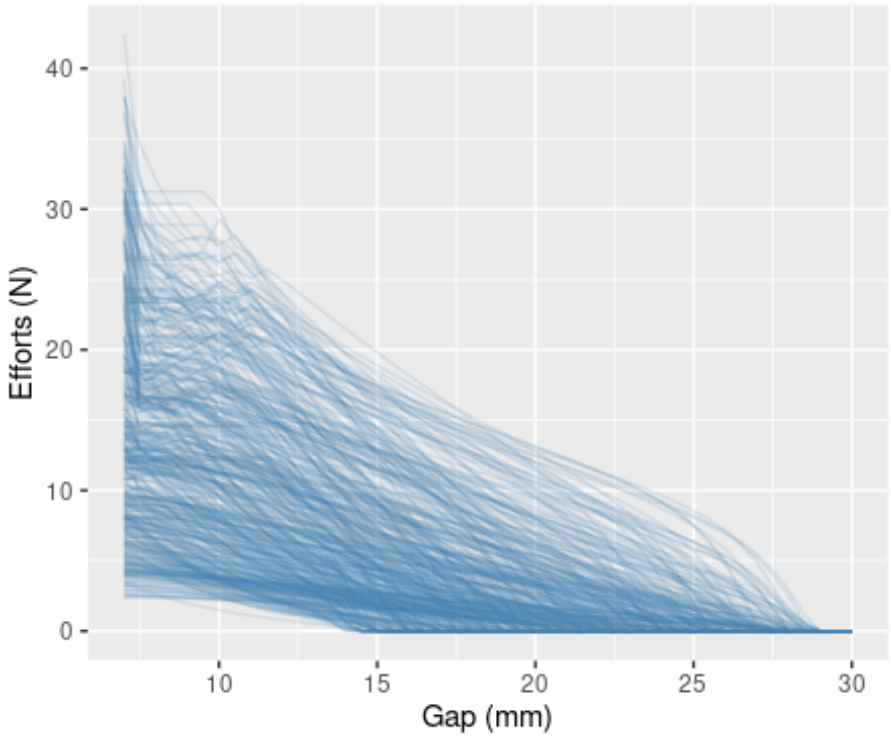

Figure 2: Observed efforts curves ($n = 498$)

- the height of the seal denoted by $(\boldsymbol{x}_i^{[H]})_{i=1,\dots,n}$,
- the thickness of the rubber denoted by $(\boldsymbol{x}_i^{[T]})_{i=1,\dots,n}$,
- a shape parameter that is an angular measure on the seal denoted by $(\boldsymbol{x}_i^{[S]})_{i=1,\dots,n}$.

The $n \times 3$ matrix of predictors $\boldsymbol{X}$ concatenate the three geometrical parameters.

First results show a meaningful partition of the effort curves (see Figure 3 on the right) : the first group (in red) is associated with curves for which a non-zero effort is observed starting from gaps comprised between 20 and 25 mm. This behaviour may be linked with a specific range of height of the seal. The third group (in blue) gathers curves that start to be non-zero roughly for gaps between 15 and 20 mm. The curves besides flattens for small gaps, indicating an upper limit for compression that could be linked with the thickness of the rubber. At a certain point, the rubber can not be more compressed. The second group contains curves with efforts measured all along the gap range and leading to mild efforts. The functional means estimated by the model are displayed in Figure 3 on the left.

Next steps to assess performance of the model includes the evaluation of the predictive performance.

$Y = Wz + \epsilon$.

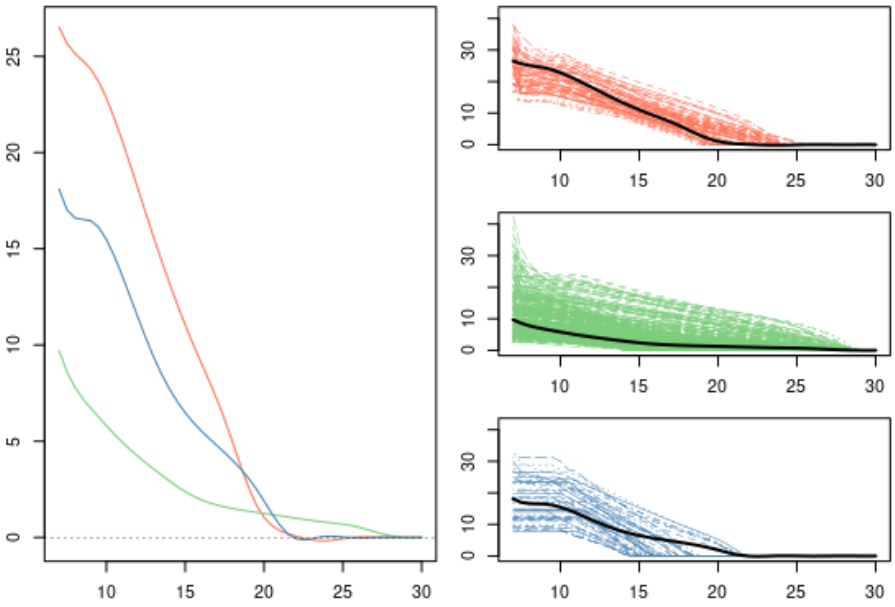

Figure 3: Left panel : Estimated group-specific functional mean. Right panel : Observed curves grouped by final class label with the functional mean estimated by the model in black solid line.

Bibliographie

Bouveyron, C. & Jacques, J. (2011), Model-based clustering of time series in group-specific functional subspaces, *Advances in Data Analysis and Classification*, 5, pp. 281-300.

Y. Chen, J. Goldsmith, T. Ogden (2016) Variable Selection in Function-on-Scalar Regression. *Stat*, 5 88-101.

Tipping, M. E. & Bishop, C. M. (1999), Probabilistic principal component analysis, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3), pp. 611-622.

Wang, J. L., Chiou, J. M. & Müller, H. G. (2016), Functional data analysis, *Annual Review of Statistics and its application*, 3, pp. 257-295.

Yao, F., Fu, Y., & Lee, T. C. (2011), Functional mixture regression, *Biostatistics*, 12(2), pp. 341-353.