Functional data analysis : modelling curves variations

Madison Giacofci
IRMAR - Université Rennes 2

Journées sur les données temporelles
INRIA Rennes - 25 mars 2019

Summary

1 Introduction : functional data

2 Nonparametric functional regression

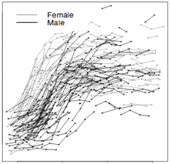
3 Modelling individual variability
- Functional Mixed-Effects model
- Estimation issue
- Model-based clustering
- Deformation models

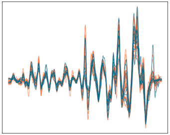
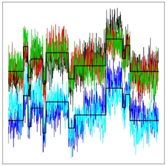
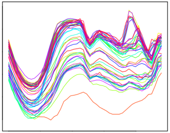
4 Conclusion and Persectives

Summary

1. Introduction : functional data
2. Nonparametric functional regression
3. Modelling individual variability
   - Functional Mixed-Effects model
   - Estimation issue
   - Model-based clustering
   - Deformation models
4. Conclusion and Persectives

- More and more research fields collect **curve-like data**
  - ▷ Growth curves, spectrometry data, genomic data, weather data...



- The term **functional data** gathers different types of data
  - Measured along time or space or any continuum
  - Sparse data
  - Dense data (high throughput)
  - With a common or not time grid

Context

- Historically, functional data appears around the 90's
- Their development has been caught between
  - Time series analysis
  - Longitudinal data analysis

- As for time series and longitudinal data, they consist of observations $(y_1^{(i)}, \ldots, y_T^{(i)})$ ordered along time for the ith individual
- Fundamental particularity :

  the ideal unit of observations are curves

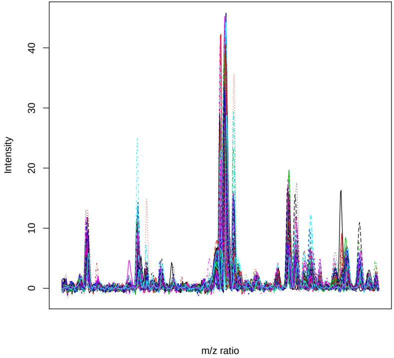
  ▷ Consider the whole curve as a single entity

  ▷ To consider functional quantities such as smoothness

Mass spectrometry data

- Related to the study of the proteome
  - ▷ To investigate biological processes involved in the development of a pathology

- Proteomic spectra generated by mass spectrometry (MALDI-TOF technology)
- Samples from 253 women under 2 conditions
  - Unaffected (91)
  - Affected by ovarian cancer (162)
- Each spectra contains 15154 ionised peptides defined by a m/z ratio.

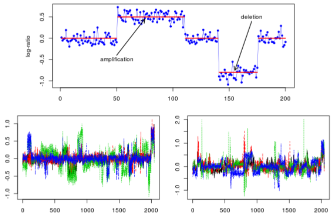


Breast cancer data
Fridlyand et al. 2006

Microarray CGH data

- Related to the study of the genome
  - ▷ To detect possible chromosomal aberrations

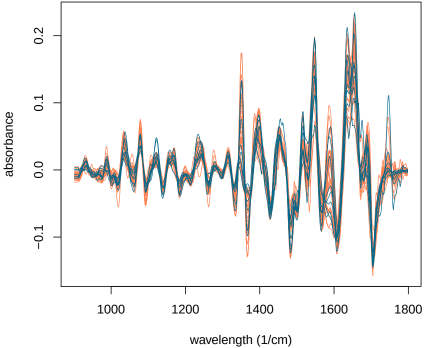
- Represent genomic profiles of breast cancer tumor cells (compared to healthy cells)
- Samples from 66 women affected
- Each spectra contains 2044 ratios of copy number measurements
- Individual spectra are piecewise constant



Mid-infrared spectrometry data

- Fast and non-invasive method to obtain the molecular composition of a biological fluid

- Ascites : abnormal presence of fluid in the abdomen
- Samples from 219 patients under 2 conditions
  - Infected liquid (46) : bad prognosis (50% mortality after 2 years)
  - Non-infected liquid (173)
- Each spectra is measured at 641 equally-spaced wavelength



Complex data

Such data are :

- Complex
- Structured
- Possibly dependent
- Intrinsically infinite dimensional

But most of the statistical questions are standard :

- Uncovering homogeneous subgroup of responses ▷ Clustering
- Recover a mean pattern ▷ Estimation/Smoothing
- Quantify variations across individuals ▷ Mixed-effects models
- Predict a response variable ▷ Supervised learning

And some others specific to the functional nature of data :

- Alignment on individual curves ▷ Deformation models
- Identify discriminative portion of a signal ▷ Domain selection
- and many others...

Summary

1 Introduction : functional data

2 Nonparametric functional regression

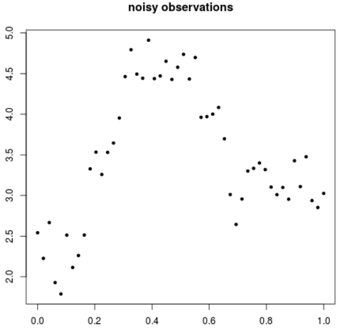
3 Modelling individual variability
- Functional Mixed-Effects model
- Estimation issue
- Model-based clustering
- Deformation models

4 Conclusion and Persectives

Functional model (Ramsay et Silverman, 1997)

The first step is to go from a discrete to a continuous representation

- Referred as smoothing in the literature
- May be included in a more complex procedure

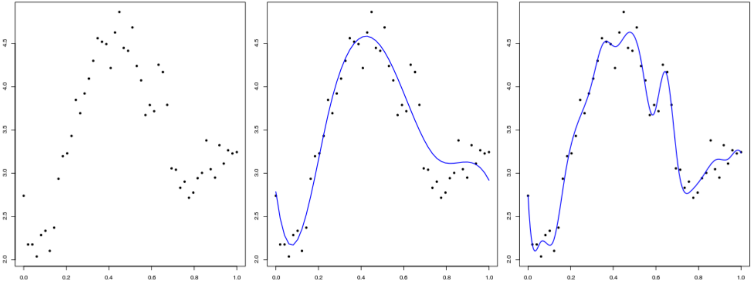


Functional model (Ramsay et Silverman, 1997)

Data are seen as curves sampled on a fine grid $(t_1, \dots, t_M)$ and corrupted by noise, such that :

$$Y(t_m) = \mu(t_m) + E(t_m), \qquad E(t_m) \sim \mathcal{N}(0, \sigma_E^2)$$

▷ Goal : Recover function μ from noisy observations



Nonparametric approach

- In a nonparametric framework, no specified forms for involved functions

  ▷ The problem lie in infinite dimensional framework

- Popular approach
  - Projection on an adapted functional basis

  ▷ Set of pre-specified basis functions $\{\phi_k, k \in \mathbb{N}\}$

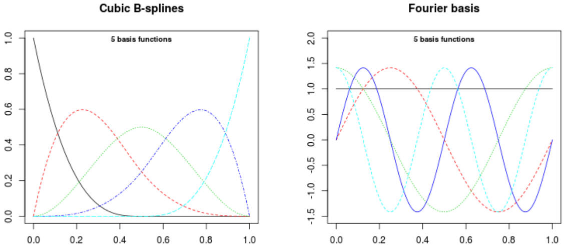
- Examples

  ▷ Fourier basis

  ▷ Splines basis

  ▷ Wavelets basis

Splines and Fourier basis



- B-splines : the most popular functional basis
  - Polynomial functions
  - Adapted to smooth data : estimation based on roughness penalty
- Fourier basis
  - Sine and cosine functions
  - Adapted to periodic data : good localisation in frequency

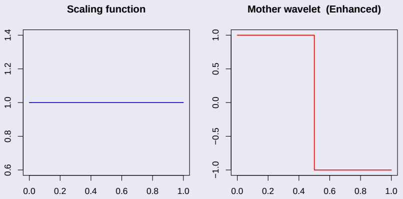
Wavelets basis

Orthonormal basis of $L^2(\mathbb{R})$ generated by dilatations and translations of a scaling function ϕ and a mother wavelet ψ such that :

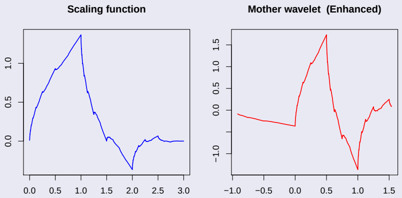
$$\{\phi_{j_0 k}(t), k = 0, \ldots, 2^{j_0} - 1;\ \ \psi_{jk}(t), j \geq j_0, k = 0, \ldots, 2^j - 1\}$$

with $\quad \phi_{jk}(t) = 2^{\frac{j}{2}}\phi(2^j t - k)$ and $\quad \psi_{jk}(t) = 2^{\frac{j}{2}}\psi(2^j t - k)$

Haar basis



Daubechies extremal phase wavelets (2 vanishing moments)



Wavelets basis

- Decomposition in the basis
  Any function $f \in L^2(\mathbb{R})$ is then expressed in the wavelet basis :

$$f(t) = \sum_{k=0}^{2^{j_0}-1} c^*_{j_0 k} \phi_{j_0 k}(t) + \sum_{j \geq j_0} \sum_{k=0}^{2^j-1} d^*_{jk} \psi_{jk}(t)$$

  where $c^*_{j_0 k} = \langle f, \phi_{j_0 k} \rangle$ and $d^*_{jk} = \langle f, \phi_{jk} \rangle$ are theoretical scaling and wavelets coefficients

Advantages and drawbacks

- Strength of wavelets
  - ▷ Fine modelling of curves with irregularities
  - ▷ Strong connexion with Besov spaces
  - ▷ Localisation in time and frequency
  - ▷ Decorrelating properties
  - ▷ Computational efficiency
  - ▷ Sparse representation of regular signals

- Drawbacks
  - ▷ Designed for equally spaced design
  - ▷ Common time grid for all individuals
  - ▷ Dyadic number of points

Discrete Wavelet Transform (DWT)

- With discretely sampled signals $\mathbf{Y} = (Y(t_1), \ldots, Y(t_M))$, we use the Discrete Wavelet Transform (Mallat, 1989) :

$$\underset{[M\times M]}{\mathbf{W}}\underset{[M\times 1]}{\mathbf{Y}} = \begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix}$$

  where $\mathbf{W}$ is an orthogonal matrix of filters (wavelet specific)

- $(\mathbf{c}, \mathbf{d})$ are empirical wavelet coefficients such that :

$$\begin{aligned} \mathbf{c} &\simeq \sqrt{M} \times \mathbf{c}^* \\ \mathbf{d} &\simeq \sqrt{M} \times \mathbf{d}^* \end{aligned}$$

Wavelet representation

- Functional nonparametric model
  In the coefficient domain, the model is written as

$$\begin{aligned} & W\mathbf{Y} = W\mu + W\mathbf{E} \\ \iff & \begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} + \varepsilon \end{aligned}$$

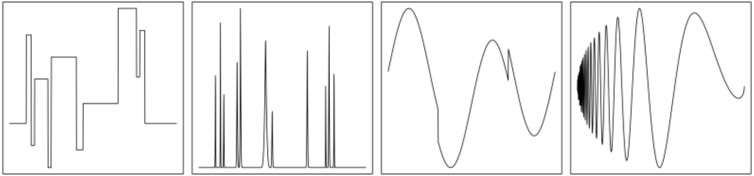
where $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$

Objective

From the linear representation in the wavelet domain, we want to find estimates of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$

Dealing with spatial heterogeneity

- Our goal is to reconstruct **spatially inhomogeneous** functions



General idea : to look at the irregularities at finer scales and stay on coarser scales for smooth parts

▷ Requires **non-linear estimates** for optimal reconstruction

- Done with **thresholding** estimators (Donoho and Johnstone, 1994)
  - ▷ Shrink, kill or keep coefficients depending on a given threshold λ to produce sparse estimates of wavelet coefficients

Thresholding estimators

General idea : Shrink, kill or keep coefficients depending on a given threshold λ

- Hard thresholding (Keep or kill rule) $d_{jk}\mathbb{1}_{\{|d_{jk}|<\lambda\}}$
- Soft thresholding (Shrink or kill rule) $\mathrm{sign}(d_{jk})(|d_{jk}|-\lambda)_+$

- Universal threshold : $\lambda = \sigma\sqrt{2\log M}$
- Robust estimate of σ based on the MAD of the coefficients at the finest resolution
- These estimators achieve a near-minimax rate of convergence (minimax within a logarithmic factor)
- Many existing variants depending on the characteristics of data

Summary

1 Introduction : functional data

2 Nonparametric functional regression

3 Modelling individual variability
- Functional Mixed-Effects model
- Estimation issue
- Model-based clustering
- Deformation models

4 Conclusion and Persectives

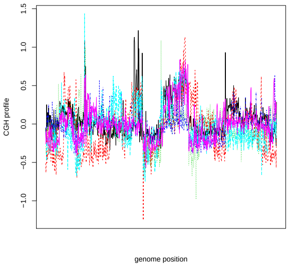
Multisample data

We now consider individual replicates, *i.e.* we observe n individual curves

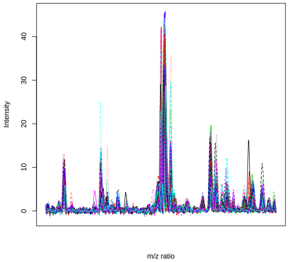
Basic nonparametric model

- Poor modelling for multisample data that can present strong individual heterogeneity

CGH data



Mass Spectrometry data



Functional mixed effect model

More appropriate framework : **Mixed-effects models**

- Modelling individuals variations around a main pattern by adding individual functional random effects

$$Y_i(t_m) = \mu(t_m) + U_i(t_m) + E_i(t_m)$$

where $U_i \sim \mathcal{N}(0, K(s,t))$ is a centered Gaussian process independant of E_i.

Wavelet representation of the model

$$\begin{bmatrix} \mathbf{c}_i \\ \mathbf{d}_i \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\nu}_i \\ \boldsymbol{\theta}_i \end{bmatrix} + \boldsymbol{\varepsilon}_i$$

where $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, \sigma_\varepsilon^2 \mathbf{I})$ and $(\boldsymbol{\nu}_i, \boldsymbol{\theta}_i) \sim \mathcal{N}(0, \boldsymbol{\Gamma})$

▷ Resumes to a linear mixed effects model on wavelet coefficients

Modelling the covariance of the random effects

- A wide litterature on functional mixed effects models
- Common essential point : modelling the matrix $\mathbf{\Gamma}$

We want to specify the covariance of the process $U_i(t)$ to :

- ▷ have a model with a "simple" structure
- ▷ make fixed and random effects lie in the same functional space (Antoniadis and Sapatinas, 2007)
  - Modelling choice justified by real data applications

Natural idea : To propose a model for $K(s,t)$ and infer conditions on matrix $\mathbf{\Gamma}$. This leads in general to difficulties :

- ▷ to control the total number of parameters
- ▷ to control trajectories regularity

Assumptions on matrix $\boldsymbol{\Gamma}$

We prefer to specify $\boldsymbol{\Gamma}$ in the wavelet domain

First assumption $\boldsymbol{\Gamma}$ is assumed to be diagonal

▷ justified by the decorrelating property of wavelets (see Johnstone and Silverman, 1997 for discussion)

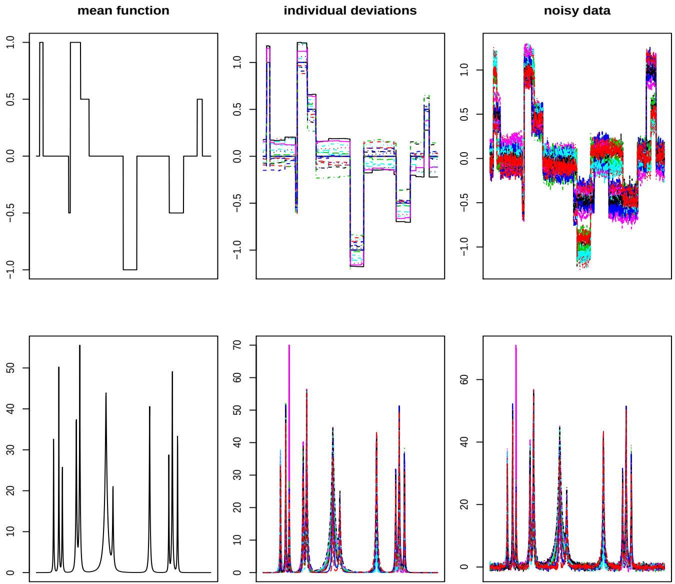
Second assumption Diagonal terms of $\boldsymbol{\Gamma}$ decrease exponentially with the scale j

$$[\boldsymbol{\Gamma}_{\boldsymbol{\theta}}]_{jk} = 2^{-j\eta}\gamma^2$$

▷ Ensure that both fixed and random effects lie in the same functional space (see theorem of Abramovich et al, 1998)

Note that γ^2 may depend on the level of decomposition and location (j, k) as γ^2_{jk}

Synthetic data



Model reinterpretation

$$\begin{bmatrix} \mathbf{c}_i \\ \mathbf{d}_i \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} + \widetilde{\varepsilon}_i \qquad \widetilde{\varepsilon}_{i,jk} \sim \mathcal{N}(0, \gamma_{jk}^2 2^{-j\eta} + \sigma_\epsilon^2)$$

individual variability + noise

▷ Nonparametric regression problem in a heteroscedastic framework with replicates

- Extension of the universal threshold to an heteroscedastic setting $\lambda = \sigma_{jk}\sqrt{2 \log M}$ (position dependant threshold)
- Estimates of parameters σ_{jk} are given by empirical estimates thanks to the N individual replicates
- The heteroscedastic thresholding enjoys a near-optimal convergence rate in the multisample setting, *i.e.* optimal within a logarithmic factor in signal size M (G., Lambert-Lacroix, Picard, 2017)

Simulation results

- $N = 100$, $M = 512$
- Control of the level of noise and individual deviations
- Competed procedures
  - Usual homoscedastic thresholding with the universal threshold [Ho]
  - Heteroscedastic thresholding with the universal threshold [He]
- Reconstruction error criteria : Mean Integrated Square Error (MISE)

Results on 200 repetitions : homoscedastic case

	SNR $= 1$		SNR$= 5$	
	Ho	He	Ho	He
Blocks	0.189 (0.016)	0.168 (0.017)	1.44e-3 (2.5e-4)	1.43e-3 (2.5e-4)
Bumps	0.736 (0.024)	0.726 (0.024)	0.045 (1.25e-3)	0.040 (1.25e-3)
Heavisine ($\times 10^{-2}$)	1.203 (0.097)	1.204 (0.104)	0.079 (0.006)	0.078 (0.006)
Doppler ($\times 10^{-4}$)	5.658 (0.246)	5.622 (0.274)	0.201 (0.011)	0.188 (0.011)

Results on 200 repetitions : heteroscedastic case

	SNR= 5			
	$\tau = 0.1$		$\tau = 1$	
	Ho	He	Ho	He
Blocks	0.186 (0.054)	0.001 (2e-4)	0.011 (0.006)	0.001 (2e-4)
Bumps	0.220 (0.029)	0.040 (0.001)	0.0573 (0.002)	0.040 (0.001)
Heavisine ($\times 10^{-2}$)	0.530 (0.016)	0.079 (0.006)	0.129 (0.008)	0.079 (0.006)
Doppler ($\times 10^{-4}$)	1.387 (0.136)	0.187 (0.117)	0.304 (0.015)	0.188 (0.010)

Curve clustering model

Natural extension : in a clustering framework, the N individuals are spread among L unknown clusters such as

$$Y_i(t_m) | \{\zeta_{i\ell} = 1\} = \mu_\ell(t_m) + U_i(t) + E_i(t_m) \qquad E(t_m) \sim \mathcal{N}(0, \sigma_E^2)$$

where $\zeta_{i\ell} = 1$ if individual i is in class ℓ

Wavelet representation of the model

- Given that $\zeta_{i\ell} = 1$,

$$\begin{bmatrix} \mathbf{c}^i \\ \mathbf{d}^i \end{bmatrix} \Bigg| \{\zeta_{i\ell} = 1\} = \begin{bmatrix} \boldsymbol{\alpha}^\ell \\ \boldsymbol{\beta}^\ell \end{bmatrix} + \begin{bmatrix} \boldsymbol{\nu}_i \\ \boldsymbol{\theta}_i \end{bmatrix} + \boldsymbol{\varepsilon}^i \; ; \quad \begin{matrix} \boldsymbol{\varepsilon}^i \sim \mathcal{N}(0, \sigma^2 \mathbf{I}) \\ (\boldsymbol{\nu}_i, \boldsymbol{\theta}_i)^T \sim \mathcal{N}(0, \boldsymbol{\Gamma}) \end{matrix}$$

- Translate into a Gaussian Mixture Model in the wavelet domain

▷ Goal : To recover individual labels $\left(\zeta_{i\ell}\right)_{i=1,\ldots,N}^{\ell=1,\ldots,L}$

Model inference

- Maximum Likelihood estimation is performed via the EM algorithm with the following unobserved variables
  - ▷ Label variables $\boldsymbol{\zeta}$
  - ▷ Random effects $(\boldsymbol{\nu}, \boldsymbol{\theta})^T$
- Dimension reduction step based on thresholding and aggregation estimators
- Model selection done via a BIC criterion

▷ Individual labels are then deduced by a *Maximum A Posteriori* (MAP) rule

Results on Mass spectrometry data

- Need important pre-treatment step (matrix effects, peaks alignment), computationally expensive (Antoniadis et. al, 2007)
- Results on a range of $M = 8192$ positions (to discard effects of matrix)

Method	FCM	FCMM	FCMM.gr	FCMM.jk
EER - global alignment	38%	24%	24%	23%
EER - group alignment	20%	21%	22%	0.4%

Conclusions

- Random effects consideration improves results for a global alignment
- Huge effect of peaks alignment
- One mismatch when variances depend on positions

  $\Rightarrow$ Suggest a sparse configuration of the random effects ?

Results on CGH data

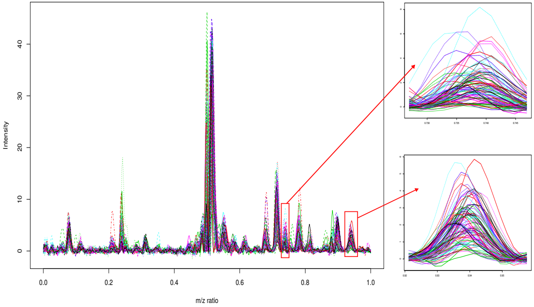
- Existing approaches based on hierarchical clustering of segmentation results (van Wieringen & van de Wiel (2008))
- Inter individual variability has never been quantified on CGH data
- On this dataset : existing clustering linked to survival data

Main conclusions

- We find more subgroups than the original study (5 vs 3)
- We retrieve the cluster associated with the best outcome (one mismatch)
- Posterior estimation of SNR and τ_U shows high level of noise and individual variability ($\approx 10^{-4}$)
  $\Rightarrow$ Find cluster with biological signifiance will require much more individuals

Back on mass spectrometry data

Among the objectives : Identifying typical main profiles associated to affected patients to enable early stage detection of a pathology.



- Irregular data
- Variations in amplitude
- Phase variations
- Error noise (electrical perturbations)

- Difference in timing can have a high impact on statistical results
- Alignment as pre-treatment step neglect its variability

Functional warped mixed model

- Individual phase deviations modeled by adding nonlinear individuals random warping functions

$$Y_i(t_{ij}) = \mu\left(w_i(t_{ij})\right) + U_i\left(w_i(t_{ij})\right) + E_{ij}, \qquad E_{ij} \sim \mathcal{N}(0, \sigma^2)$$

- $w_i(.) = w(., \boldsymbol{\theta}_i)$ are parametrized by individual random variables $\boldsymbol{\theta}_i$

Modelling of the warping functions w_i

- Monotonicity constraint : the functions w_i are nondecreasing diffeomorphisms
- [Gervini and Carter (2014)] Using Hermite splines that offer a direct link between knots and features
- [Bigot (2013)] Using Ordinary Differential Equation based on a reference function parametrized by cubic B-splines

Inference in shape invariant functional mixed models

- The model fits in the Non Linear Mixed Effects Models (NLMEM) setting (Lindstrom & Bates (1990))
- Nonobserved random effects (amplitude and phase deviations parameters) considered as hidden data
- We use EM algorithm for Maximum Likelihood estimation of the parameters

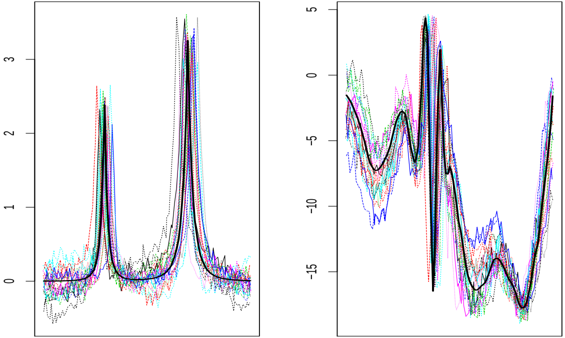
- Computation of the E-step
  - ▷ Untractable because of the nonlinearity of warping functions
- Several existing approaches to either
  - approximate the likelihood function (Lindstrom & Bates (1990))
  - stochastically approximate the expectations of interest (Kuhn & Lavielle (2005))
- We first focus on MCEM algorithm based on Monte-Carlo approximations of the desired expectations

Applications on simulated data

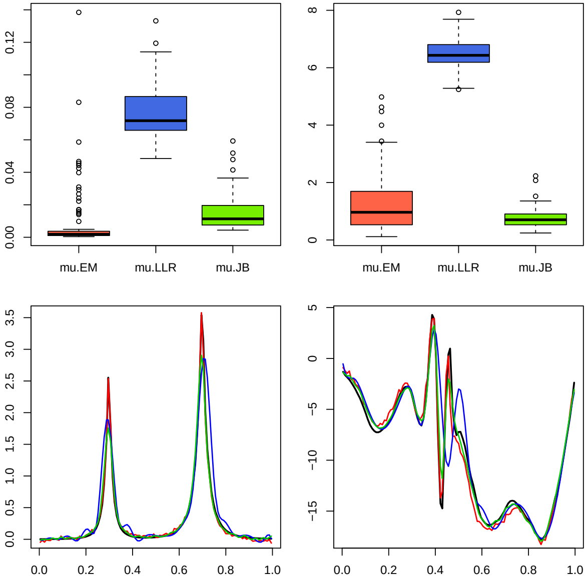
We compete three procedures

- [Bigot13]
- [Raket14]
- Wavelet-based warped mixed model

Compared w.r.t the MISE criteria (Mean Integrated Square Error) on the functional fixed effect.



MISEs results - High deviations in phase and amplitude



Summary

1 Introduction : functional data

2 Nonparametric functional regression

3 Modelling individual variability
- Functional Mixed-Effects model
- Estimation issue
- Model-based clustering
- Deformation models

4 Conclusion and Persectives

Conclusions

Conclusions

- Flexible functional modelling for irregular curves with diverse sources of variability
- Related to standard models in the coefficient domain

Remaining questions

- Supervised learning (PhD Marie Morvan in co-supervision)
- Sparse estimation of functional random effects
- Clustering in deformation models

References

Antoniadis, A. and Sapatinas, T. (2007), *Estimation and inference in functional mixed-effects models*. CSDA, 51 (10), p4793-4813.

Donoho, D.L. and Johnstone, I.M. (1994), *Ideal spatial adaptation by wavelet shrinkage*. Biometrika, 81 (3), p425-455.

Härdle, W. and Kerkyacharian, G. and Picard, D. and Tsybakov, A. (1998), *Wavelets, Approximation and Statistical Applications*. Springer.

Morris, J.S. and Carroll, R.J. (2006), *Wavelet-based functional mixed models*. JRSSB, 68 ; 179-199.

Giacofci, M., Lambert-Lacroix, S., Marot, G., Picard, F. (2013) *Wavelet-based clustering for mixed-effects functional models in high dimension*. Biometrics, 69(1) ; 31-40.

Giacofci, M., Lambert-Lacroix, S., Picard, F. (2017) *Minimax wavelet estimation in heteroscedastic nonparametric regression with repetitions*. Journal of Nonparametric Statistics, 30(1) ; 238-261.