Functional Data Analysis

Jairo Cugliari

Université Lumière Lyon 2

29 août 2022, Journées MAS, Rouen

Functional observation

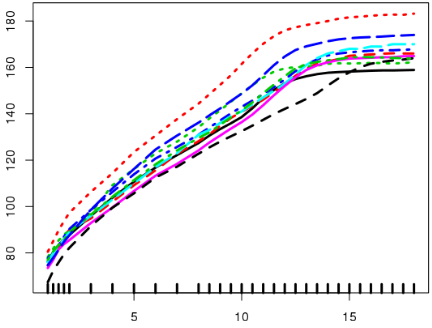
A realization of a (typically smooth) random object that takes values in an abstract function space[1]

1. From Alexandre Aue

Example 1 : Berkeley Growth Study

- Height measurement of 10 girls at 31 ages
- Ages are not equally spaced
- Uncertainty of about 3mm
- Values reflect a smooth variation in height
- Data can be considered as 10 functional observations $\texttt{Height}_i(t), i = 1, \ldots, 10$.

Tuddenham & Snyder(1954)

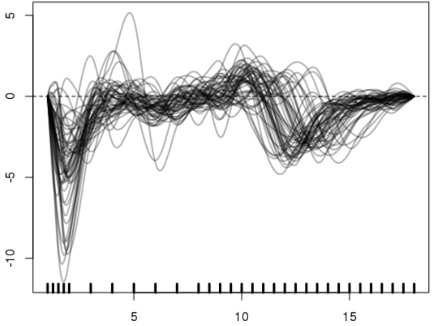


Then, we can display the acceleration curves

$$D^2\texttt{Height} = \frac{d^2\texttt{Height}}{dt^2}$$

to highlight subtle features in this data.

- Strong positive acceleration ⇒ pubertal growth
- Bump at at around 6-years old ⇒ mid-spurt



In this case, the functional representation allowed us

- to use derivatives which carry important information
- to give an explicit role to the ages

Both points are at least very difficult if we want to use multivariate data analysis.

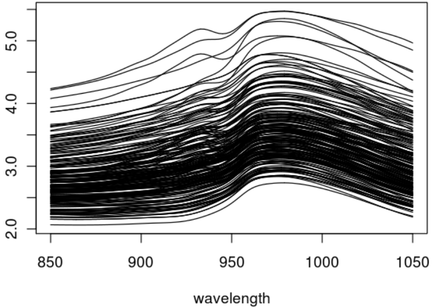
Example 2 : Tecator Infratec Food and Feed Analyzer

aim Predict the moisture, fat and protein contents of finely chopped pure meat avoiding (expensive and destructive) clinical analysis.

data p spectrometric responses associated to a finite number of wavelengths using the Near Infrared Transmission (NIT) principle.

pb $p \gg$ the number of samples

model Noiseless smooth curves defined over a range of wavelengths.



Example 3 : fda script

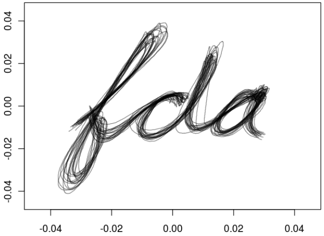
aim Analyze variability on digital signature.

data Smooth X-Y coordinates of 20 replications represented by 1401 coordinate values.

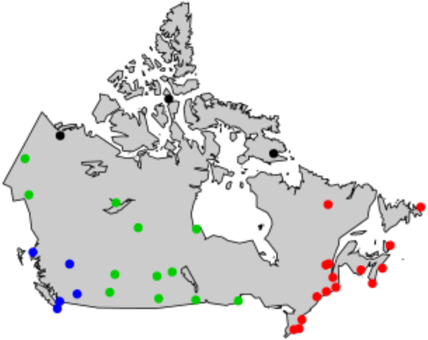
- adjusted to a common length that corresponds to 2.3 seconds
- important features in each script are aligned

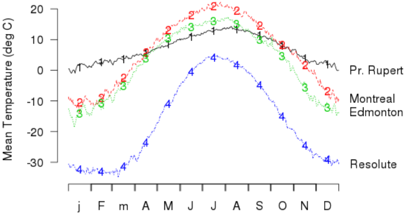
pb Curve registration, feature alignment.

model Noiseless smooth curves defined over $[0, 1]^2$



Example 4 : Canadian weather





aim Study the variability between different weather stations in Canada.

data Mean temperature measure weekly at 35 locations (grouped in 4 climate zones : Atlantic, Pacific, Continental, Arctic)

Functional Data Analysis (FDA)

FDA *refers to the statistical analysis of data samples consisting of random functions or surfaces, where each function is viewed as one sample element. Müller (2011).*

When ?

- **When the sample curves is highly regular.** Then *variables* are highly correlated which may produce numerical problems and mask relevant effects of the analysis.
- **When regular curves are sampled with noise.** Data should be denoised, preferably in coordination with the analysis technique.

Plan

Functional variable and functional dataset

- Random function : $X = \{X(t), t \in T\}$
- Realization $x = \{x(t), t \in T\}$ of X (for example $F = C[0,1]$ or $F = L_2([0,1])$)
- $X_1(t), \ldots, X_n(t)$ are iid $\Rightarrow$ Functional random sample
- In practice : serial correlation and spatial dependence are usual.

How to represent FD ?

- In practice we observe do not observe $x = \{x(t), t \in T\}$ but

$$\mathbf{x} = \{x(t_j), j = 1, \ldots, N\}, \qquad N < \infty$$

problem How to represent FD from discrete (finite dimensional) sampling discrete (possibly noisy) sampling at $\{t_k, k = 1, \ldots, N\}$.

Interpolation If the observations are assumed to be noiseless

Smoothing Smoothing, to remove noise. We can use curve estimation theory, that includes :

- Basis expansion
- Smoothing penalization
- Local regression methods
  - kernel regression
  - local polynomial regression

All these variants share the *bias-variance trade-off* and the fact that they require to choose some *smoothing parameter*.

Basis expansion

- We observe y_i that contains the target $x(t_i)$ with some noise ϵ_i

$$y_i = x(t_i) + \epsilon_i$$

- Chose a basis $\{\phi_k(t)\}_k$, popular choices are Fourier, B-splines, Wavelets, ...
- Given your basis, compute the coefficients c_k :

$$x(t) = \sum_{k=1}^{K} c_k \phi_k(t)$$

- The number of elements K is to be chosen !
  - If $K > N$ makes no sense ($K = N$ achieves perfect fit)
  - Ideal : $K << N$: interpretations are easier and computations faster.

Smoothing penalties

- We observe $\mathbf{y}$ that contains the target $\mathbf{x}$ with some noise ϵ, i.e. $\mathbf{y} = \mathbf{x} + \epsilon$. Let $\mathbf{x} = \phi\mathbf{c}$.
- We can estimate $\mathbf{c}$ using OLS but we want a smooth solution
- Penalize good fit if it produces an oscillating curve :
  - Choose a rich basis $\{\phi_k(t)\}_k$ and a smoothing penalty
  - Estimate to solve

$$\min \|y - \phi c\| + \lambda \int [Lx(t)]^2 dt$$

  - $Lx(t)$ measures the lack of smoothness (roughness) of x, some popular choices are
    - Harmonic acceleration : $Lx = \omega^2 Dx + D^3 x$
    - Curvature $Lx(t) = D^2 x(t) dt$
  - λ is a tunning parameter that increasingly penalizes roughness of the solution (it can be tunned by cross validation)

Plan

Summarize FD

- We dispose with a functional sample $X_1, \dots, X_n$.
- The mean population μ is estimated by the sample mean :

$$\hat{\mu}_n = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- The covariance operator Γ is estimated by Γ_n

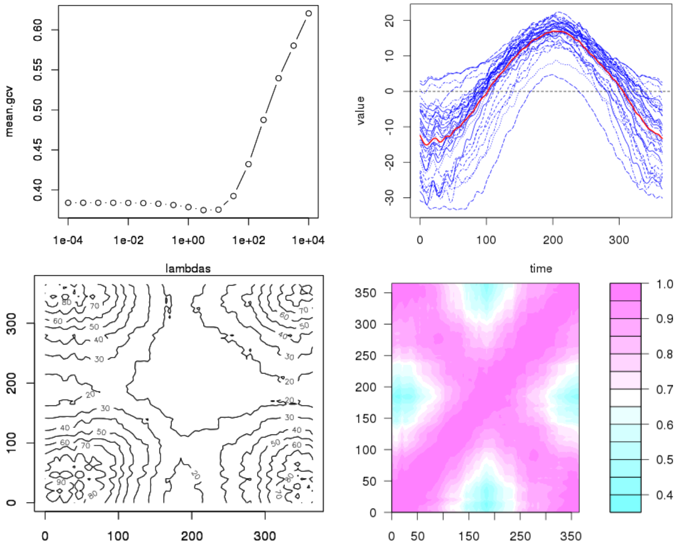
$$\hat{\Gamma}_n(h) = \frac{1}{n}\sum_{i=1}^{n} < X_i - \bar{X}_n, h > (X_i - \bar{X}_n)$$

If $F = L_2([0,1])$ we have

- Mean : $\bar{x}(t) = \frac{1}{n}\sum_{i=1}^{n} x_i(t)$
- Covariance : $\sigma(s,t) = \mathrm{cov}(x(s), x(t)) = \frac{1}{n}\sum_{i=1}^{n}(x_i(s) - \bar{x}(s))(x_i(t) - \bar{x}(t))$

Example (Canadian weather data)

Smoothing B-splines basis and second-derivative penalties (K choosen by CV)



Functional PCA

Karhunen-Loève transform of X

$$X(t) = \mu(t) + \sum_{j \geq 1} C_j f_j(t)$$

- $f_j(t)$ form an orthonormal basis of eigen-functions (principal factors) and are solutions of $\Gamma f_j = \lambda_j f_j$
- C_j are zero-mean uncorrelated random variables (principal components) with variance λ_j, $\lambda_1 \geq \lambda_2, \ldots,$

$$C_j = < X - \mu, f_j >$$

Functional PCA

- In practice $\sigma(t,s)$ is unknown and so they are f_j and λ_j.
- Solving the eigen-analysis problem needs (in general) to approximate each curve x_i and eigenfunction f_j in a basis of functions $\{\phi_k(t)\}_k$

$$x_i(t) = \sum_{k=1}^{K} \gamma_{ik}\phi_k(t), \qquad f_j(t) = \sum_{k=1}^{K} b_{jk}\phi_k(t).$$

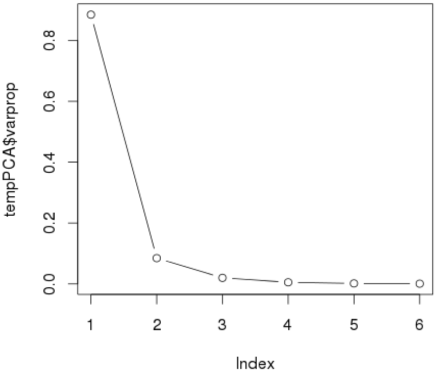
- With this approximation, C_j and λ_j are solutions of

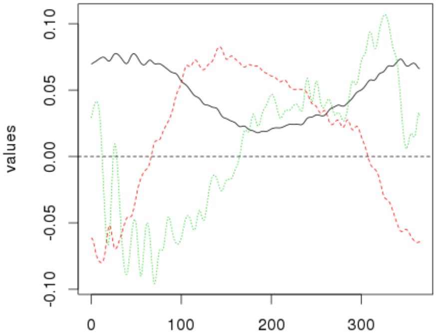
$$\frac{1}{n-1} A^{1/2} W' W A^{1/2} b_j = \lambda_j b_j$$

with

  - b_k is the vector $(b_{1k,\ldots b_{Kk}})$
  - W is matrix of centered coefficients γ_{ik}
  - A is the matrix of inner products between basis functions

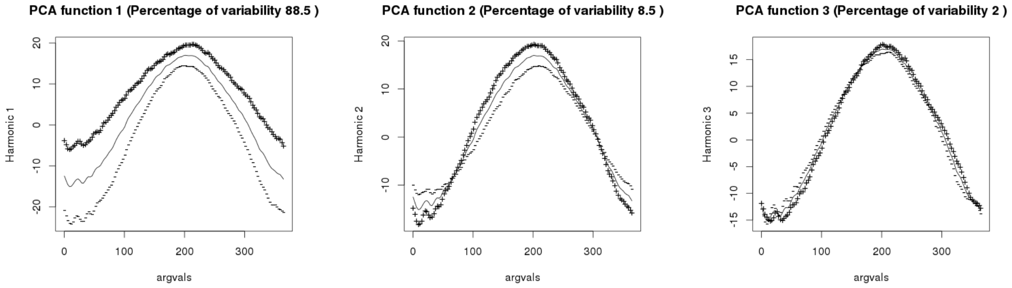
Example (Canadian weather)





- From the scree plot we choose the first 3 harmonics (as in MDA)

Example (Canadian weather)



- Plots $\bar{x}(t) \pm 2\sqrt{\lambda_k} e_k(t)$ for $k = 1, 2, 3$
- Also derivatives can be used for interpretation

Clustering functional data

- Aim : to group curves into homogeneous groups
- Different approaches :
  - Feature extraction + multivariate clustering
  - Use distances between functions : Antoinadis, Brossat, Cugliari and Poggi (2013) Clustering functional data with waveletes, IJMRA
  - Model-based clustering : J.Jacques and C.Preda (2014), Functional data clustering : a survey, Advances in Data Analysis and Classification, 8[3], 231-255

Plan

Functional linear regression

- Aim : to describe predictive relationships
- Different scenarios

$$y_i = \alpha + \mathbf{x}_i \beta + \epsilon_i$$

  - Scalar response with functional predictor
  - Functional response with scalar predictor
  - Functional response with functional predictor

Scalar response model

Use the $\{t_j\}_j$ from the sampling grid : $y_i = \alpha + \sum \beta_j x_i(t_j)$

- With increasingly finer grids, we have (in the limit)

$$y_i = \alpha + \int \beta(t) x_i(t) dt + \epsilon_i$$

- As we have infinitely many covariates, estimate β by minimizing squared error makes no sense (identification problem).
- Solution : regularization using restricted basis functions

$$\beta(t) = \sum_{k=1}^{K_\beta} b_k \psi_k(t) \Leftrightarrow \beta(t) = \psi'(t)\mathbf{b}$$

where ψ_k can be the basis used for curves smoothing or another one.

$$x_i(t) = \sum_{k=1}^{K_\beta} \gamma_{ik} \phi_k(t) \Leftrightarrow x(t) = C\Phi'(t)$$

where C is the $n \times K$ coefficient matrix.

Then

$$\mathbf{y} = \alpha + \int \hat{\beta}(t) x_i(t) dt = \alpha + \int C\Phi(t)\Psi(t)'\mathbf{b} dt + \epsilon = CJ_{\Psi,\Phi}\mathbf{b} + \epsilon$$

with $\mathbf{J}_{\Psi,\Phi} = \int \Phi(t)\Psi(t)' dt$

With the notation $\mathbf{Z} = [\mathbf{1}, \mathbf{CJ}_{\Psi,\Phi}]$ and $\xi = (\alpha, b_1, \dots, b_{K_\beta})$, the model becomes

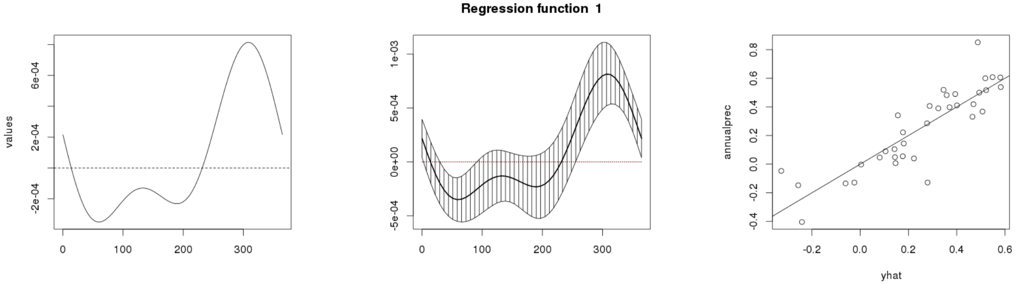
$$\mathbf{y} = \mathbf{Z}\xi + \varepsilon$$

and the OLS estimate of ξ is given by (if $K_\beta < K$),

$$\widehat{\xi} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$$

Example (Canadian weather)

Aim : predict the log annual precipitation from the temperature profile.



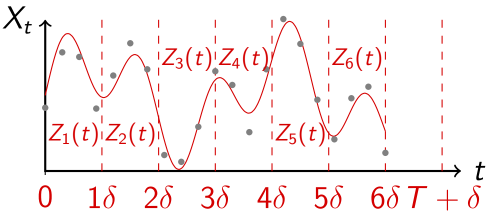
Plan

FD as slices of a continuous process

[Bosq, (1990)]

The prediction problem

- Suppose one observes a square integrable continuous-time stochastic process $X = (X(t), t \in \mathbb{R})$ over the interval $[0, T]$, $T > 0$;
- We want to predict X all over the segment $[T, T + \delta], \delta > 0$
- Divide the interval into n subintervals of equal size δ.
- Consider the functional-valued discrete time stochastic process $Z = (Z_k, k \in \mathbb{N})$, where $\mathbb{N} = \{1, 2, \dots\}$, defined by



$$Z_k(t) = X(t + (k-1)\delta)$$

$$k \in \mathbb{N} \quad \forall t \in [0, \delta)$$

If X contents a $\delta-$seasonal component, Z is particularly fruitful.

Prediction of functional time series

Let $(Z_k, k \in \mathbb{Z})$ be a stationary sequence of H-valued r.v. Given $Z_1, \ldots, Z_n$ we want to predict the future value of Z_{n+1}.

- A predictor of Z_{n+1} using $Z_1, Z_2, \ldots, Z_n$ is

$$\widetilde{Z_{n+1}} = \mathbb{E}[Z_{n+1}|Z_n, Z_{n-1}, \ldots, Z_1].$$

Autoregressive Hilbertian process of order 1

The $\mathrm{ARH}(1)$ centred process states that at each k,

$$Z_k = \rho(Z_{k-1}) + \epsilon_k \tag{1}$$

where ρ is a compact linear operator and $\{\epsilon_k\}_{k\in\mathbb{Z}}$ is an $H-$valued strong white noise.

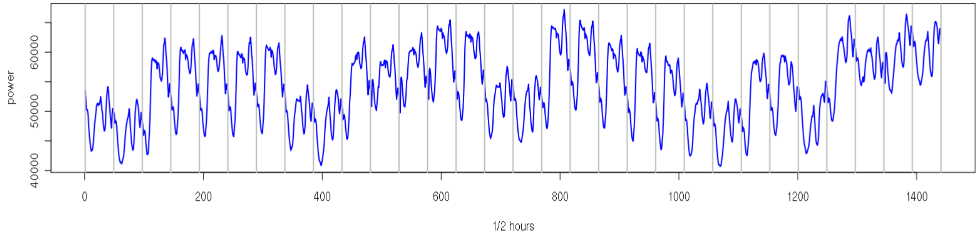
Under mild conditions, equation (1) has a unique solution which is a strictly stationary process with innovation $\{\epsilon_k\}_{k\in\mathbb{Z}}$. [Bosq, (1991)]

When Z is a zero-mean $\mathrm{ARH}(1)$ process, the best predictor of Z_{n+1} given

Overview

IDEA Similar past causes produce similar future consequences.

- We need an appropriate distance between current and past situations.



- What is a segment ?
- How do I represent segments ?
- What does similar mean ?

Approximation and details

- In practice, we don't dispose of the whole trajectory but only with a (possibly noisy) sampling at 2^J points, for some integer J.
- Each approximated segment $Z_{i,J}(t)$ is broken up into two terms :
  - a smooth approximation $\mathcal{S}_i(t)$ (lower freqs)
  - a set of details $\mathcal{D}_i(t)$ (higher freqs)

$$Z_{i,J}(t) = \underbrace{\sum_{k=0}^{2^{j_0}-1} c_{j_0,k}^{(i)} \phi_{j_0,k}(t)}_{\mathcal{S}_i(t)} + \underbrace{\sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} d_{j,k}^{(i)} \psi_{j,k}(t)}_{\mathcal{D}_i(t)}$$

- The parameter j_0 controls the separation. We set $j_0 = 0$.

$$\widetilde{z_J}(t) = c_0 \phi_{0,0}(t) + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} d_{j,k} \psi_{j,k}(t).$$

A two step prediction algorithm

Step I : Dissimilarity between segments

Search the past for segments that are similar to the last one.
For two observed series of length 2^J say Z_m and Z_l we set for each scale $j \geq j_0$:

$$\text{dist}_j(Z_m, Z_l) = \left(\sum_{k=0}^{2^j-1} (d_{j,k}^{(m)} - d_{j,k}^{(l)})^2 \right)^{1/2}$$

Then, we aggregate over the scales taking into account the number of coefficients at each scale

$$D(Z_m, Z_l) = \sum_{j=j_0}^{J-1} 2^{-j/2} \text{dist}_j(Z_m, Z_l)$$

A two step prediction algorithm

Step 2 : Kernel regression

Obtain the prediction of the scale coefficients at the finest resolution $\Xi_{n+1} = \{c_{J,k}^{(n+1)} : k = 0, 1, \ldots, 2^J - 1\}$ for Z_{n+1}

$$\widehat{\Xi_{n+1}} = \sum_{m=1}^{n-1} w_{m,n} \Xi_{m+1}, \qquad w_{m,n} = \frac{K\left(\frac{D(Z_n, Z_m)}{h_n}\right)}{\sum_{m=1}^{n-1} K\left(\frac{D(Z_n, Z_m)}{h_n}\right)}$$

Finally, the prediction of Z_{n+1} can be written

$$\widehat{Z_{n+1}(t)} = \sum_{k=0}^{2^J-1} \widehat{c_{J,k}^{(n+1)}} \phi_{J,k}(t)$$

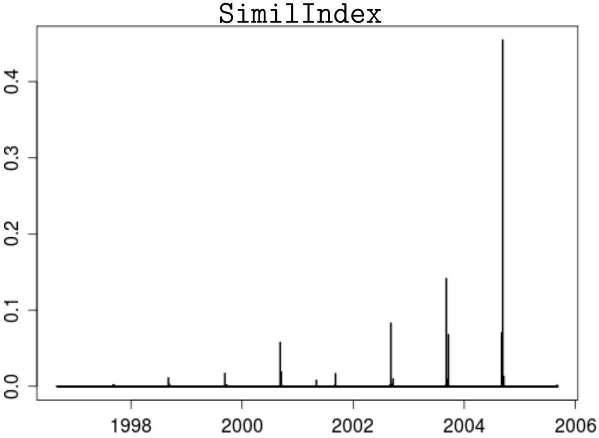
Let us predict Saturday 10 September 2005

We use Antoniadis *et al.*, (2006) prediction method with corrections to cope with non stationarity.

- Use the last observed segment ($n = 9/\text{Sept}/2005$) to look for similar segments in past.
- Construct a similarity index `SimilIndex` (using a kernel).
- Prediction can be written as

$$\widehat{\text{Load}}_{n+1}(t) = \sum_{m=1}^{n-1} \text{SimilIndex}_{m,n} \times \text{Load}_{m+1}(t)$$

- First difference correction of the approximation part.
- Use of groups to anticipate calendar transitions.



date	SimilIndex
2004-09-10	0.455
2003-09-05	0.141
2002-09-06	0.083
2004-09-03	0.070
2003-09-19	0.068
2000-09-08	0.058
2000-09-15	0.019
1999-09-10	0.017

Plan

FDA Overview

- Data samples consisting of random functions or surfaces.
- Smoothness hypothesis on the underlying stochastic processes.
- FDA provides a statistical approach to the analysis of repeatedly observed stochastic processes or data generated by such processes.
- FDA approaches and models are essentially nonparametric :
  - series expansions, penalized splines, or local polynomial smoothing,
  - functional principal component analysis
- FDA differs from time series :
  - sampling designs may be sparse and irregularly observed
  - milder hypothesis about the underlying process
- FDA differs from MDA :
  - FD are inherently infinite-dimensional
  - smoothness often is a central assumption.
  - MDA is permutation invariant
- FDA differs from smoothing :
  - smoothing typically implies one non-random object perturbed by noisy

Books

- D. Bosq (2000) *Linear Processes in Function Spaces* ISBN :978-0-387-95052-4, 283 p.
- J. Ramsay, G. Hooker & Spencer, G. (2009) *Functional Data Analysis with R and MATLAB*, ISBN :978-0-387-98185-7, 207 p.
- J. Ramsay & B.W. Silverman (2005) *Functional Data Analysis* (2nd ed.), ISBN :978-0-387-40080-8, 430 p.
- F. Ferraty & P. Vieu (2006) *Nonparametric Functional Data Analysis*, ISBN : 978-0-387-30369-7, 268 p.
- Horváth, L. & Kokoszka, P. (2013) *Inference for Functional Data with Applications* ISBN :978-1-4614-3655-3, 422p.

Software

R libraries
(`https://cran.csail.mit.edu/web/views/FunctionalData.html`)

- `fds`, H.-L. Shang & R.J. Hyndman
- `fda`, J.O. Ramsay, H. Wickham, S. Graves & G. Hooker
- `fda.usc`, M. Febrero & M. Oviedo.
- `rainbow`, H.-L. Shang & R.J. Hyndman

Python

- `scikit-fda` Grupo de Aprendizaje Automático - Universidad Autónoma de Madrid Revision

Matlab

- `fda` J.O. Ramsay, H. Wickham, S. Graves & G. Hooker (2013)
- `PACE` H.-G. Müller, *et al.* (2012)

References

- J.O. Ramsay and C. Dalzell (1991) Some tools for functional data analysis (with discussion). *Journal of the Royal Statistical Society*, Series B, 53, 365–380
- H.-G. Müller. Functional Data Analysis. In : Lovric, M. (ed.), (2010). International Encyclopedia of Statistical Science. Heidelberg : Springer Science +Business Media, LLC, reprinted and freely available at StatProb: The Encyclopedia Sponsored by Statistics and Probability Societies.
- J. Cao, J. Nielsen, J. Ramsay & F. Yao (2010) The Future of Functional Data Analysis, final rapport of the *Functional Data Analysis : Future Directions* workshop, May 2 to 7 at Banff, Canada.