Introduction to Statistical Machine Learning for Functional Data

Santiago Gallón

Departamento de Matemáticas y Estadística
Facultad de Ciencias Económicas
Universidad de Antioquia
Medellín, Colombia

Second Workshop on Big Data – CO
Facultad de Ciencias Naturales y Matemáticas – Facultad de Economía
Universidad del Rosario, May 30-31 – June 1, 2018
Bogotá – Colombia

This lecture is based primarily on:

- *Nonparametric Regression and Generalized Linear Models* by Green and Silverman (1994)
- *Smoothing Splines* by Wang (2011)
- *Functional Data Analysis* by Ramsay and Silverman (2005)
- *Elements of Statistical Learning* by Hastie et al. (2009)
- *Rainbow Plots, Bagplots, and Boxplots for Functional Data* by Hyndman and Shang (2010)
- *Visualizing and Forecasting Functional Time Series* by Shang (2010)
- *Inference for Functional Data with Applications* by Horváth and Kokoszka (2012)
- *Analysis of Variance for Functional Data* by Zhang (2013)
- *A Survey of Functional Principal Component Analysis* by Shang (2014)
- Some figures are taken from Hastie et al. (2009)

Outline I

Introduction to functional data analysis I

- Usually, the sample is a set of *finite*-dimensional elements
- In many applications, these elements are assumed as *random functions*
- This is possible due to advances of technology.
- *Sample of curves*, $Y_1(t), \ldots, Y_n(t)$, as paths of a continuous stochastic process $Y = \{Y(t),\, t \in \mathcal{T}\} \in \mathcal{F}$
- FDA: *statistical analysis of samples of curves, surfaces or anything else varying over a continuum*
- It is an important framework for *Big Data* (Hadjipantelis and Müller, 2018)

Introduction to functional data analysis II

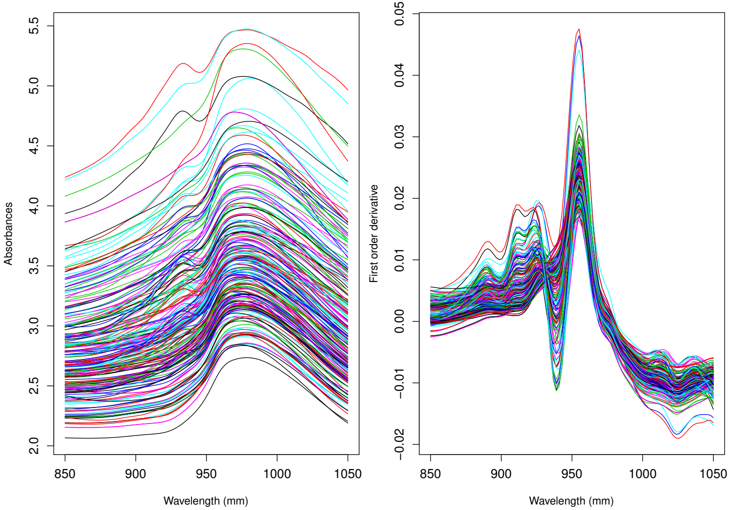


Figure 1: Spectrum of absorbances of meat samples. Ferraty and Vieu (2006).

Introduction to functional data analysis III

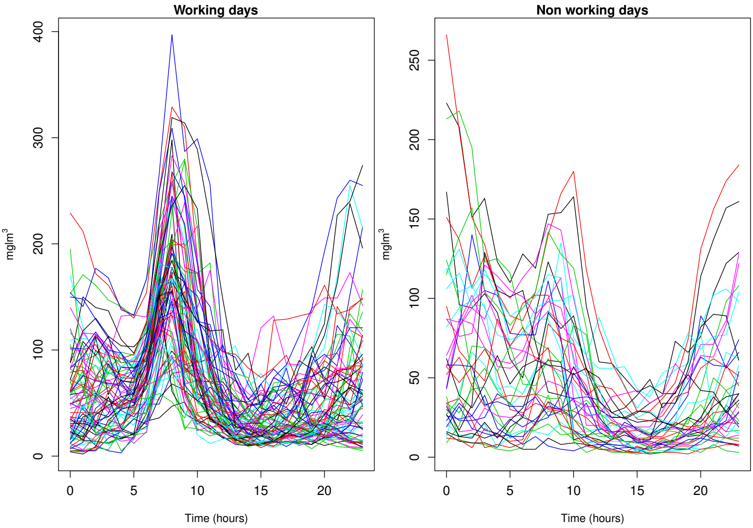


Figure 2: Hourly NOx emissions in Poblenou-Spain. Febrero et al. (2008).

Introduction to functional data analysis IV

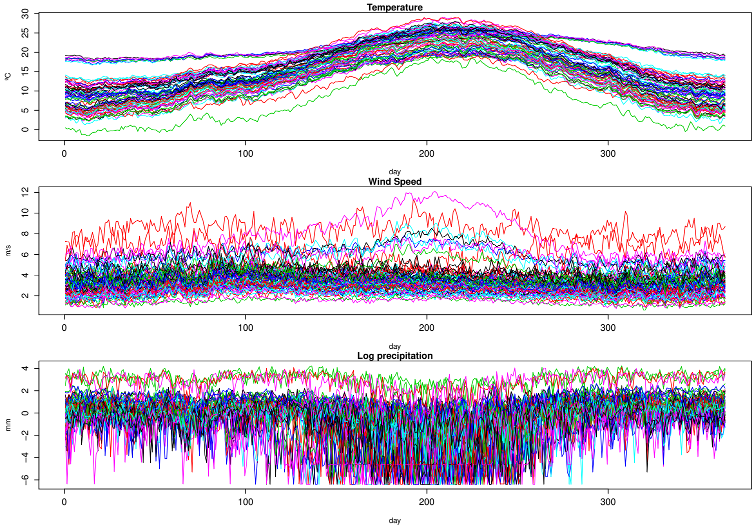


Figure 3: Spain daily weather curves, 1980-2009. AEMET.

Introduction to functional data analysis V

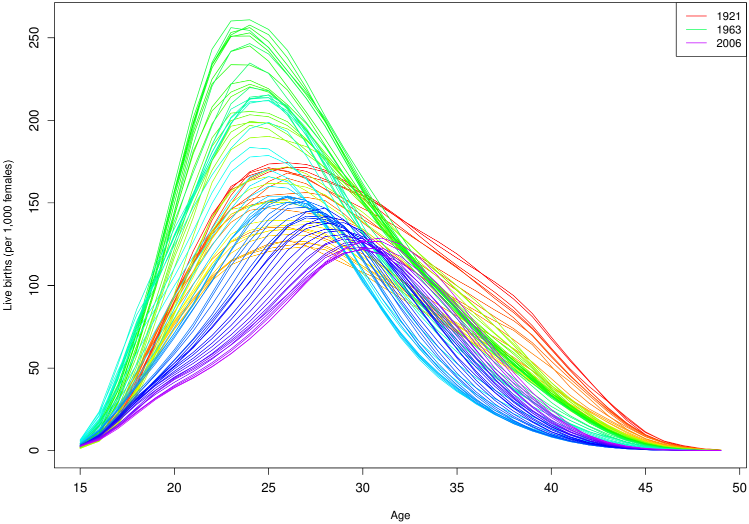


Figure 4: Australian fertility rates, 1921-2006. Hyndman and Ullah (2007).

Introduction to functional data analysis VI

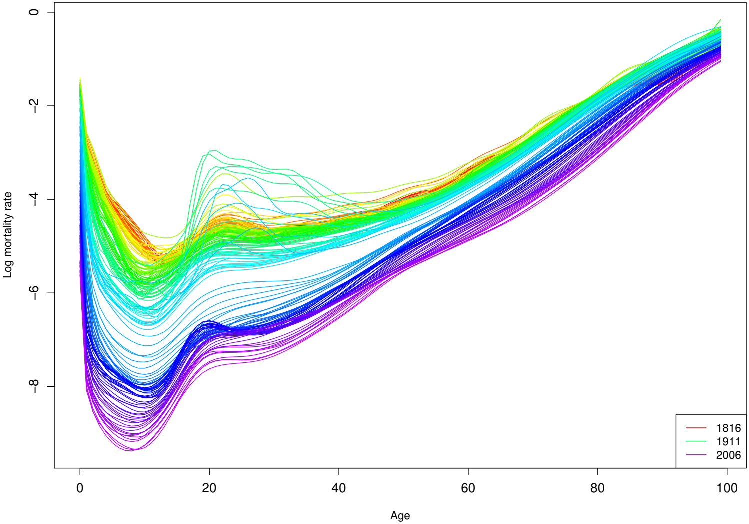


Figure 5: French male mortality rates, 1816-2006. Hyndman and Ullah (2007).

Introduction to functional data analysis VII

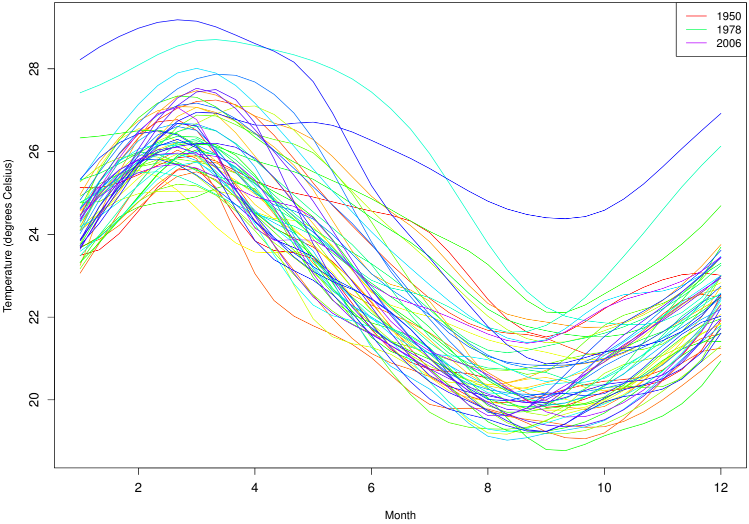


Figure 6: Monthly sea surface temperatures, Jan-1950 / Dec-2006. NOAA.

Introduction to functional data analysis VIII

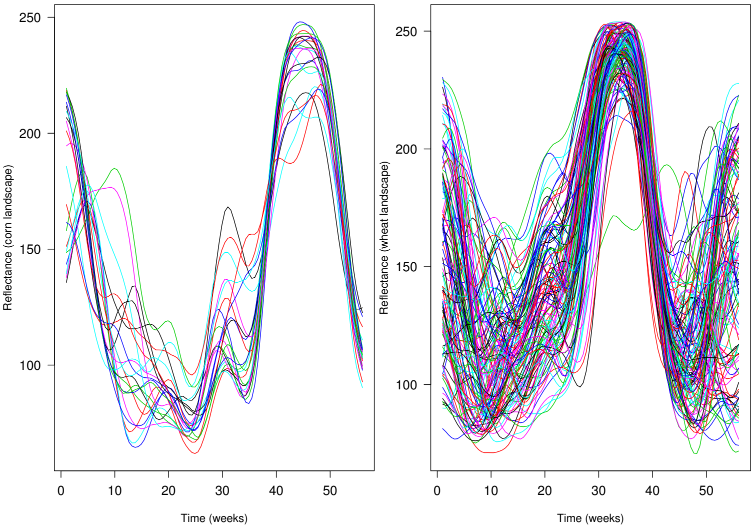


Figure 7: Landscape reflectances curves. Gallón (2013).

Introduction to functional data analysis IX

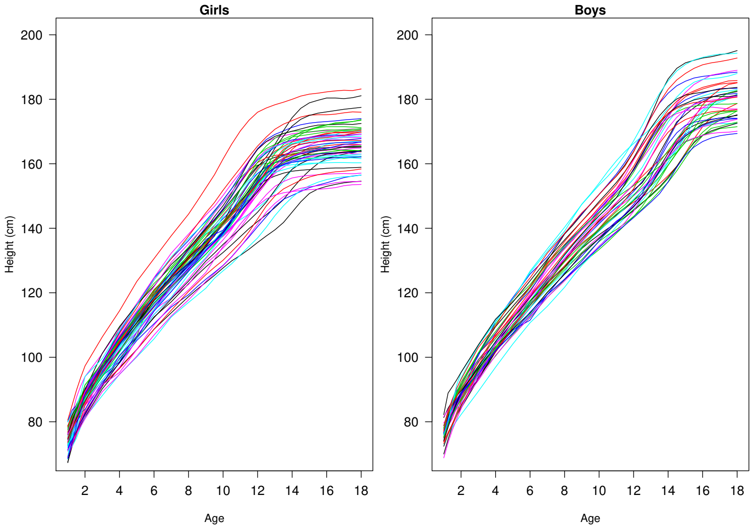


Figure 8: Heights of 54 girls and 39 boys. Ramsay and Silverman (2002).

Introduction to functional data analysis X

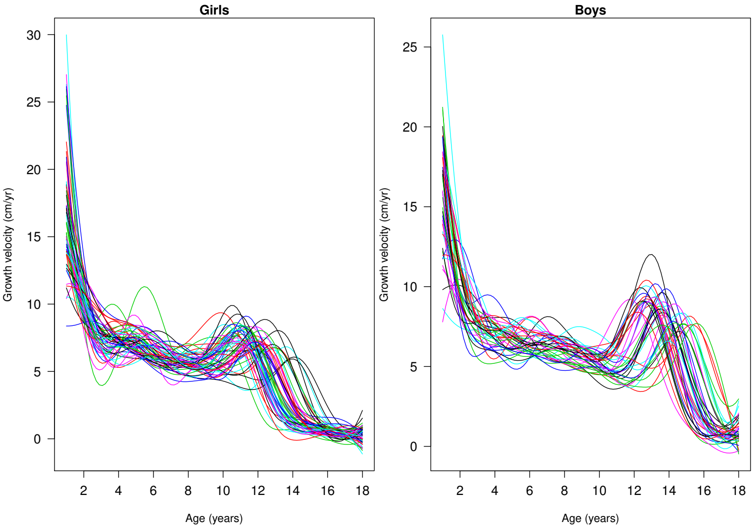


Figure 9: Child growth velocity curves. Ramsay and Silverman (2002).

Introduction to functional data analysis XI

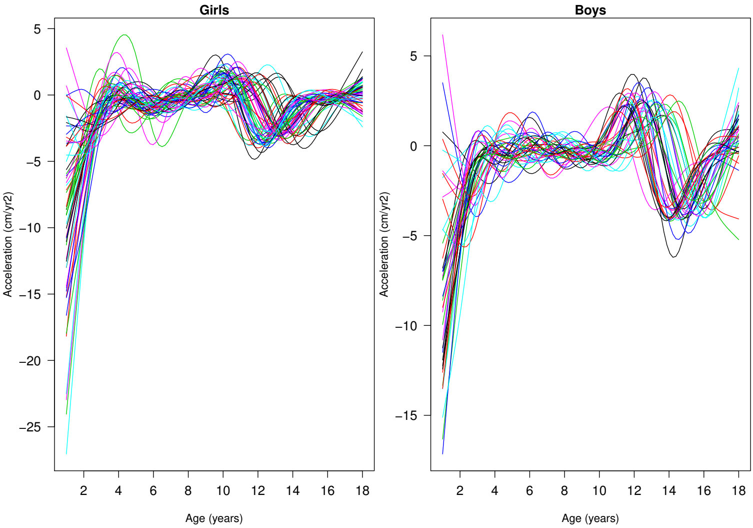


Figure 10: Child growth acceleration curves. Ramsay and Silverman (2002).

Introduction to functional data analysis XII

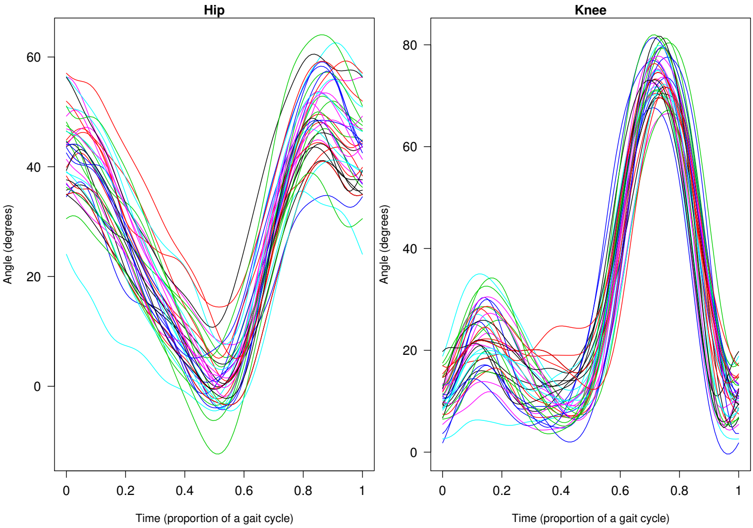


Figure 11: Angle curves through a gait cycle. Ramsay and Silverman (2002).

Introduction to functional data analysis XIII

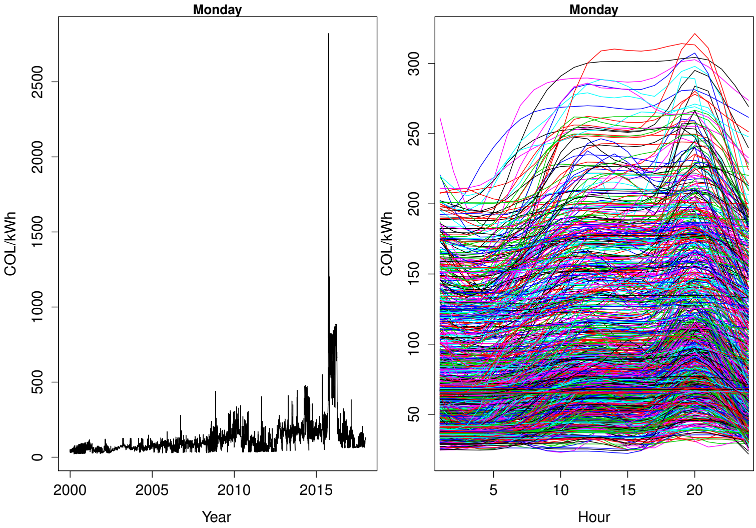


Figure 12: Hourly Colombian spot electricity price.

Introduction to functional data analysis XIV

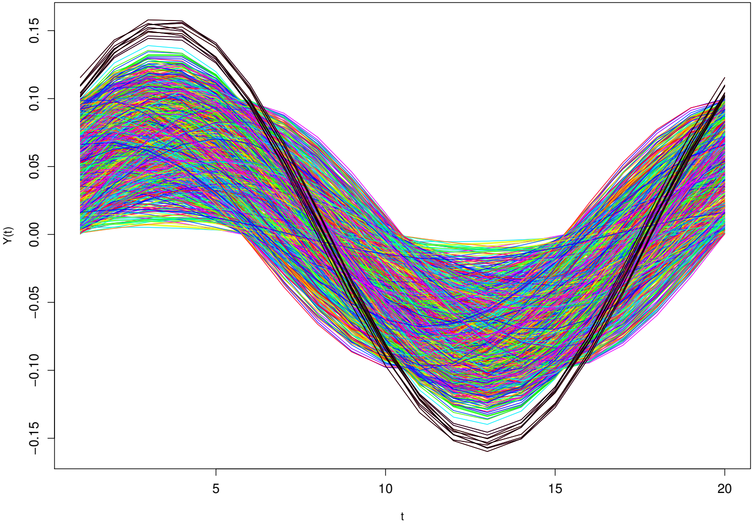


Figure 13: Simulated functional curves. Hyndman and Shang (2010).

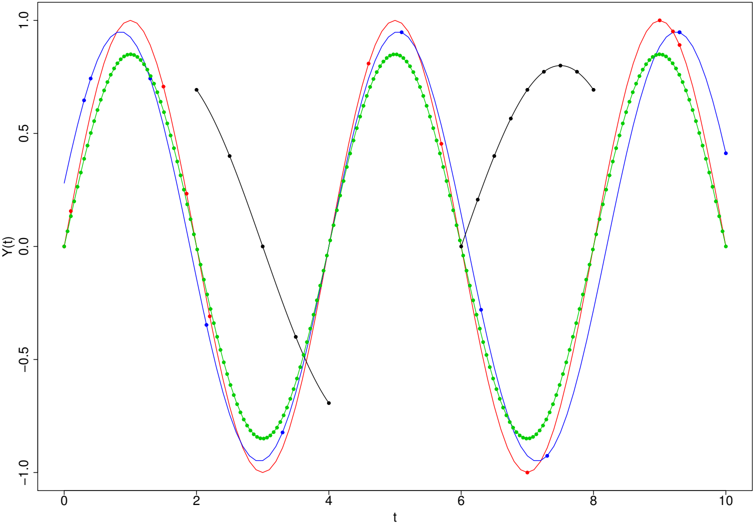
Introduction to functional data analysis XV

- Areas of application:
  - Bioinformatics (expression density curves,...)
  - Medicine (physical growth curves, heart rate curves,...)
  - Physics (impulse and spectral reflectance curves,...)
  - Economics and finance (growth curves, yield and returns curves,...)
  - Energy (price, load and demand curves)
  - Marketing (sales curves,...)
  - Astronomy (spectral curves,...)
  - Meteorology (temperature and precipitation curves,...)
  - Transport and telecommunications (velocity curves, wireless signals)
  - Archeology, ecology, geology, physiology, criminology, education,...

Introduction to functional data analysis XVI

- Functional data are mainly generated by:
  - Technological systems (GPS's, sensors, satellites, spectrograms, ...)
  - Genome projects
  - Online social networks
  - Electronic trading systems
  - Experiments and simulations
- Cases in which the functional data approach is appropriate:
  - Irregularly spaced measurements
  - Sampling time points are not the same across subjects
  - High-frequency data
  - Analysis with derivatives of the functions

Introduction to functional data analysis XVII



Introduction to functional data analysis XVIII

- Goals of FDA
  - Represent data in ways that provide further analysis
  - Study patterns and sources of variation
  - Display the data highlighting its salient structural features
  - Capture underlying dynamics through derivatives
  - ...

Introduction to functional data analysis XIX

- Important developments have been made, e.g., in *functional*
  - Clustering
  - Dimensionality reduction
  - Time series analysis
  - Regression
  - Classification
  - Outlier detection
  - Robustness
  - Variable selection
  - Testing
  - Visualization tools

Introduction to functional data analysis XX

- Main steps in FDA
  - Reconstruct functions $\{f_i\}_{i=1}^n$ by applying some curve estimation method
    * Kernel estimation
    * Local polynomial smoothing
    * Regression splines
    * Smoothing splines
    * Wavelets
  - Obtain estimators of functional parameters based on $\{\hat{f}_i\}_{i=1}^n$
  - Carry out inferences based on functional parameter estimators and $\{\hat{f}_i\}_{i=1}^n$

Notions of curve estimation I

- Observed data of a function in a functional dataset

$$\mathcal{D}_m = \{(t_j, y_j)\}_{j=1}^m$$

- Nonparametric (NP) regression model

$$y_j := y(t_j) = f(t_j) + \varepsilon_j, \qquad \varepsilon_j \overset{\text{iid}}{\sim} (0, \sigma^2), \;\; j = 1, \ldots, m$$

- Goal: Reconstruct the *unknown* function f based on $\mathcal{D}_m$.
- The NP approach makes assumptions on qualitative properties on f.
- Usually, f is assumed to be a *smooth* function.
- The idea is "*to let the data speak for themselves*".

Functional basis expansions I

- Many $\mathcal{F}$ classes admit a basis expansion

$$f(t) = \sum_{l=1}^{\infty} \theta_l \phi_l(t), \qquad \{\phi_l(t)\} \text{ basis functions}$$

- The estimation is achieved by,

$$\widehat{f}(t) = \sum_{l=1}^{L} \widehat{\theta}_l \phi_l(t) \qquad \textit{Projection estimator!}$$

- Thus, using a basis expansion for $f(t_j)$,

$$\begin{aligned} y_j &= f(t_j) + \varepsilon_j \\ &= \sum_{l=1}^{L} \theta_l \phi_l(t_j) + \varepsilon_j \\ &= \boldsymbol{\theta}^\top \boldsymbol{\phi}(t_j) + \varepsilon_j, \qquad j = 1, \ldots, m. \end{aligned}$$

Functional basis expansions II

- In matrix notation

$$
\begin{aligned}
\boldsymbol{y} &= \boldsymbol{f} + \boldsymbol{\varepsilon} \\
\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} &= \begin{pmatrix} f(t_1) \\ \vdots \\ f(t_m) \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{pmatrix} \\
&= \begin{pmatrix} \phi_1(t_1) & \cdots & \phi_L(t_1) \\ \vdots & \ddots & \vdots \\ \phi_1(t_m) & \cdots & \phi_L(t_m) \end{pmatrix} \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_L \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{pmatrix} \\
&= \boldsymbol{\Phi}\boldsymbol{\theta} + \boldsymbol{\varepsilon},
\end{aligned}
$$

where $\{\boldsymbol{\Phi}\}_{jl} = \phi_l(t_j)$.

Smoothing splines I

- Find $\hat{f} \in \mathcal{F}$ such that minimizes

$$\sum_{j=1}^{m} (y_j - f(t_j))^2 + \lambda \int \left(f^{''}(t) \right)^2 dt$$
$$= \|\boldsymbol{y} - \boldsymbol{\Phi\theta}\|_2^2 + \lambda \boldsymbol{\theta}^\top \boldsymbol{\Omega\theta},$$

  where $\{\boldsymbol{\Omega}\}_{jj'} = \int \phi_j^{''}(t)\phi_{j'}^{''}(t)dt$.

- $\lambda > 0$ controls the *fit* *vs.* *roughness* penalty trade-off.
  - ✓ $\lambda = 0$, then $\hat{f}$ interpolates the data.
  - ✓ $\lambda \to \infty$, then $\hat{f}$ converges to the least squares line.

- λ controls the amount of smoothing.

Smoothing splines II

- Solving for $\boldsymbol{\theta}$,

$$\widehat{\boldsymbol{\theta}}_\lambda = (\boldsymbol{\Phi}^\top \boldsymbol{\Phi} + \lambda \boldsymbol{\Omega})^{-1} \boldsymbol{\Phi}^\top \boldsymbol{y}.$$

- So that,

$$\begin{aligned} \widehat{\boldsymbol{f}}_\lambda = \widehat{\boldsymbol{y}} &= \boldsymbol{\Phi} \widehat{\boldsymbol{\theta}}_\lambda \\ &= \boldsymbol{\Phi} (\boldsymbol{\Phi}^\top \boldsymbol{\Phi} + \lambda \boldsymbol{\Omega})^{-1} \boldsymbol{\Phi}^\top \boldsymbol{y} \\ &= \boldsymbol{S}_\lambda \boldsymbol{y}. \end{aligned}$$

- $\hat{f}$ is a *natural cubic spline* with knots at the data points $t_1, \ldots, t_m$.
- That is, $\{\phi_l(t)\}_{l=1}^L$ are natural cubic spline basis.
- But, what is a *natural cubic spline*?

Cubic splines I

- Let $[\xi_k, \xi_{k+1})$, $k = 0, \ldots, K$ be a partition of $t \in [a, b]$

$$a = \xi_0 < \xi_1 < \cdots < \xi_K < \xi_{K+1} = b.$$

- $\xi_1 < \cdots < \xi_K$ are known as *knots*.
- A *cubic spline* is a continuous function f such that:
  - f is a cubic polynomial over $[\xi_k, \xi_{k+1})$
  - f has continuous first and second derivatives at knots $\xi_1 < \cdots < \xi_K$
- In general, an Lth-order spline is a piecewise $L-1$ degree polynomial with $L-2$ continuous derivatives at knots.

– Cubic splines correspond to $L = 4$. *The most used in practice!*

Cubic splines II

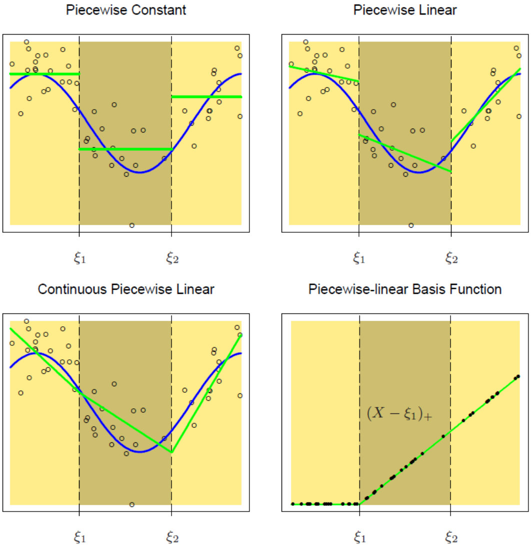


Figure 14: Piecewise-linear polynomial. Source: Hastie et al. (2009).

Cubic splines III

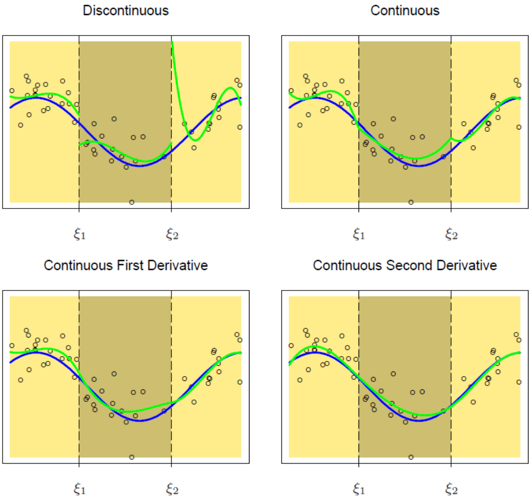


Figure 15: Piecewise-cubic polynomial. Source: Hastie et al. (2009).

Cubic splines IV

- General form of the basis expansion

$$f(t) = \sum_{l=1}^{L} \theta_l t^{l-1} + \sum_{k=1}^{K} \theta_{L+k} (t - \xi_k)_+^{L-1}, \quad t \in [a, b],$$

where

$$(t - \xi_k)_+ = \max\{t - \xi_k, 0\} = \begin{cases} t - \xi_k & \text{if } t \geq \xi_k \\ 0 & \text{if } t < \xi_k. \end{cases}$$

- For a cubic spline,

$$f(t) = \theta_1 + \theta_2 t + \theta_3 t^2 + \theta_4 t^3 + \sum_{k=1}^{K} \theta_{L+k} (t - \xi_k)_+^3$$

Cubic splines V

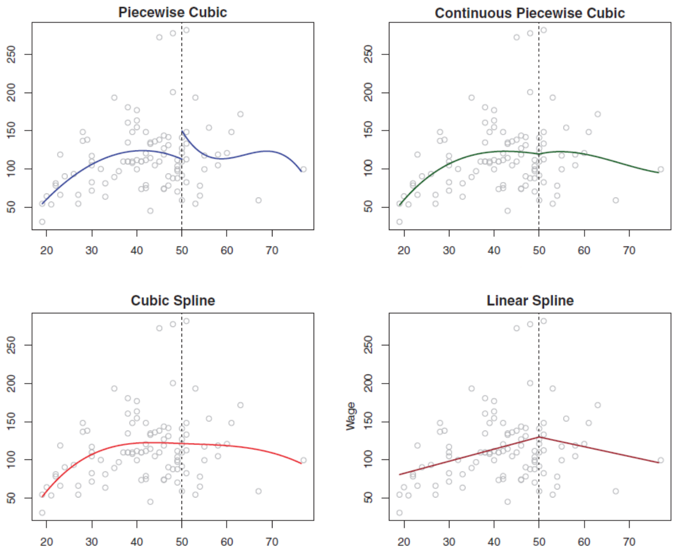


Figure 16: Smoothing Splines . Source: James et al. (2013).

Cubic splines VI

- The polynomials fit tends to be unstable near the boundaries.
- Solution: add boundary constraints to the splines.
- *Natural spline*: A spline that is linear beyond the boundary knots.
- Other kind of splines:
  - B-, M-, I-, C-, L-splines
  - Periodic splines
  - Thin-plate splines
  - Spherical splines
  - Partial splines
- R packages:
  - ✓ `splines`
  - ✓ `splines2`
  - ✓ `assist`
  - ✓ `gss`
  - ✓ `fda`

Choice of the smoothing parameter I

- λ is chosen by cross-validation

$$\mathrm{CV}(\lambda) = \frac{1}{m}\sum_{j=1}^{m}\left(y_j - \hat{f}_\lambda^{(-j)}(t_j)\right)^2 = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{y_j - \hat{f}_\lambda(t_j)}{1-\{\boldsymbol{S}_\lambda\}_{jj}}\right)^2,$$

  where $\hat{f}_\lambda^{(-j)}$ is the fit obtained by omitting the point (t_j, y_j).

- Generalized cross-validation

$$\mathrm{GCV}(\lambda) = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{y_j - \hat{f}_\lambda(t_j)}{1-m^{-1}\sum_{j=1}^{m}\{\boldsymbol{S}_\lambda\}_{jj}}\right)^2.$$

- λ is the one that $\hat{\lambda}_{\mathrm{GCV}} = \arg\min_{\lambda>0}\mathrm{GCV}(\lambda)$.

Functional principal components I

- *Covariance function* of a continuous stochastic process Y,

$$\begin{aligned} K(s,t) &:= \operatorname{Cov}\big(Y(s), Y(t)\big), \qquad\qquad s,t \in \mathcal{T} \\ &= \mathbb{E}\left[(Y(s)-\mu(s))\,(Y(t)-\mu(t))\right] \end{aligned}$$

- Under certain conditions, K induces the *kernel operator* $\mathcal{K}$,

$$(\mathcal{K}\phi)(s) = \int_{\mathcal{T}} K(s,t)\phi(t)\mathrm{d}t$$

- FPCA relies on the *spectral decomposition* of $\mathcal{K}$ (Mercer's lemma)

$$(\mathcal{K}\phi_\nu)(s) = \lambda_\nu \phi_\nu(s), \qquad s \in \mathcal{T},\ \nu \in \mathbb{N},$$

  $\{\phi_\nu\}$ is an orthonormal sequence of continuous eigenfunctions, and $\{\lambda_\nu\}$ the corresponding decreasing sequence of non-negative eigenvalues, and

$$K(s,t) = \sum_{\nu=1}^{\infty} \lambda_\nu \phi_\nu(s)\phi_\nu(t).$$

Functional principal components II

- Functional Principal Component (FPC) scores

$$\beta_\nu = \int_{\mathcal{T}} [Y(s) - \mu(s)]\, \phi_\nu(s) \mathrm{d}s,$$

  which are zero-mean uncorrelated r.v.'s with variance λ_ν.

- The Karhunen-Loève (or FPC) expansion of Y

$$Y(t) = \mu(t) + \sum_{\nu=1}^{\infty} \beta_\nu \phi_\nu(t), \quad t \in \mathcal{T}.$$

Functional principal components III

Example: $\mathcal{T} = (1, \ldots, p) \Rightarrow \boldsymbol{y} \in \mathbb{R}^p$, $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\boldsymbol{K} \in \mathbb{R}^{p \times p}$.

- Spectral decomposition

$$\boldsymbol{K}\boldsymbol{\phi}_\nu = \lambda_\nu \boldsymbol{\phi}_\nu, \quad \nu = 1, \ldots, p,$$

  or equivalently,

$$\boldsymbol{K}\boldsymbol{\Phi} = \boldsymbol{\Phi}\boldsymbol{\Lambda}, \quad \boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_p), \ \boldsymbol{\Phi}\boldsymbol{\Phi}' = \boldsymbol{I}_p = \boldsymbol{\Phi}'\boldsymbol{\Phi},$$

  where $\boldsymbol{K} = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}' = \sum_{\nu=1}^{p} \lambda_\nu \boldsymbol{\phi}_\nu \boldsymbol{\phi}_\nu'$.

- PC scores

$$\beta_\nu = (\boldsymbol{y} - \boldsymbol{\mu})' \boldsymbol{\phi}_\nu.$$

- PC's expansion

$$\boldsymbol{y} = \boldsymbol{\mu} + \sum_{\nu=1}^{p} \beta_\nu \boldsymbol{\phi}_\nu.$$

Functional data visualization I

- Useful to discover features that are not apparent with summary statistics or statistical models.
- Tools covered in this lecture (Hyndman and Shang, 2010):
  - *Rainbow plots*
  - *Bivariate and functional boxplots*
  - *Bivariate and functional highest density region (HDR) boxplots*
- Other plots:
  - Phase-plane plots
  - Rug plots
  - Singular value decomposition plots
  - ...

Rainbow plots I

- Each curve is colored with a rainbow color according to a data ordering:
  - Time observation (by default!)
  - Location depth
  - Highest density region

Rainbow plots II

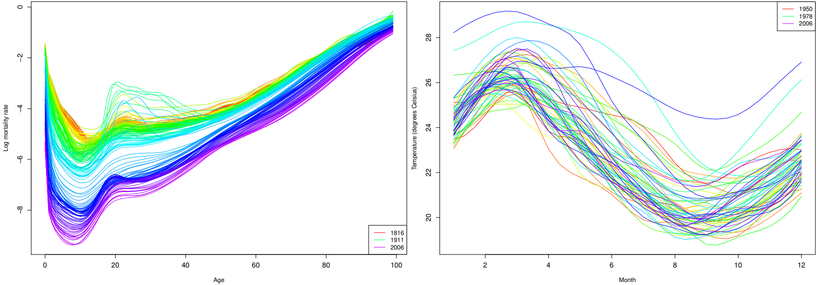


Figure 17: Rainbow plots (time ordering)

Rainbow plots III

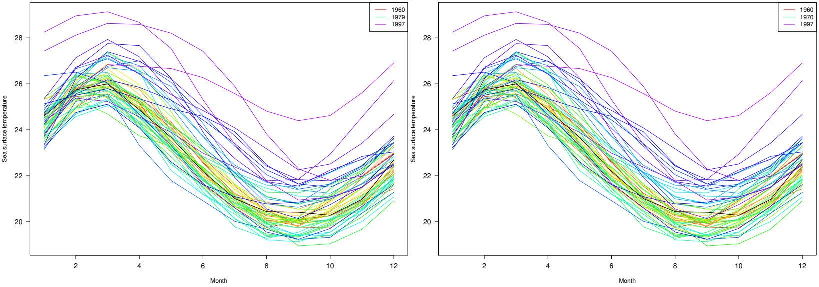
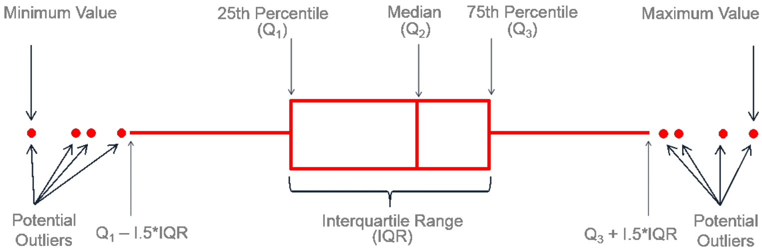


Figure 18: Rainbow plots with depth (left) and density (right) ordering

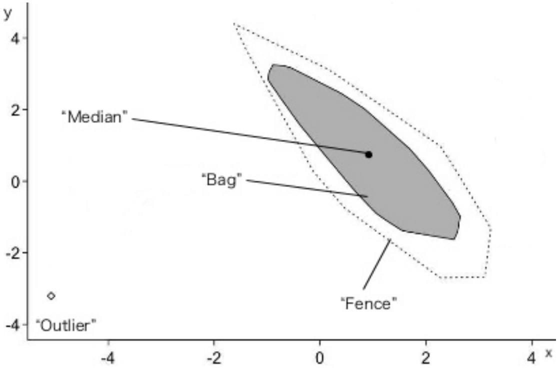
Bivariate and functional Bagplots I

- *Bagplot*: bivariate version of the univariate boxplot by Tukey (1977)
- Proposed by Rousseeuw, Ruts and Tukey (1999)



Bivariate and functional Bagplots II

- Components:
  - *Depth median*: point with highest halfspace depth
  - *Bag*: smallest depth region with 50% of points
  - *Fence*: inflated bag by a factor ρ (usually, with 99% of points)
  - *Loop*: region with points outside the bag but inside the fence
  - *Tails*: points outside the fence are flagged as outliers



Bivariate and functional Bagplots III

- Bagplot allows to visualize the data structure:
  - Location (depth median)
  - Spread (bag's size)
  - Correlation (bag's orientation)
  - Skewness (bag's shape)
  - Tails (potential outliers)

Bivariate and functional Bagplots IV

- *Bivariate bagblot:* Bagplot on the first two FPC scores
- *Functional bagblot:* Mapping of the bivariate bagplot to functional curves

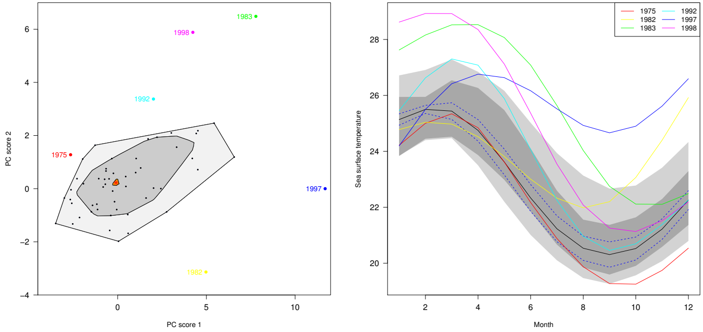


Figure 19: Bivariate and functional boxplots for sea surface temperatures.

Bivariate and functional HDR Bagplots I

- Bivariate FPC scores ordered by the *Highest Density Region* (HDR)
- Based on the bivariate kernel density estimate,

$$\hat{f}(\boldsymbol{z}) = \frac{1}{n}\sum_{i=1}^{n} K_{\lambda_i}\left(\boldsymbol{z} - \boldsymbol{Z}_i\right), \qquad \boldsymbol{Z}_i = (\beta_{i1}, \beta_{i2}),$$

  $K_{\lambda_i}(\cdot) = K(\cdot/h_i)/h_i$ is a bivariate kernel function, and λ_i the bandwidth.

- The HDR, with coverage probability $1-\alpha$,

$$R_\alpha = \left\{\boldsymbol{z} : \hat{f}(\boldsymbol{z}) \geq f_\alpha\right\},$$

  where f_α is a such that $\mathbb{P}(\boldsymbol{Z} \in R_\alpha) \geq 1-\alpha$.

  – Points within R_α have higher density estimate w.r.t. those outside R_α.

Bivariate and functional HDR Bagplots II

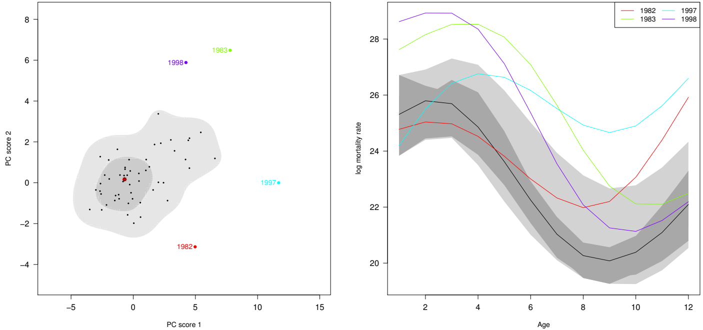


Figure 20: Bivariate and functional HDR boxplots for sea surface temperatures.

R Applications I

- R CRAN Task View: Functional Data Analysis
- Some popular packages:
  - ✓ `fda` (Ramsay et al., 2017)
  - ✓ `rainbow` (Shang and Hyndman, 2016)
  - ✓ `fda.usc` (Febrero and Oviedo de la Fuente, 2012)
  - ✓ `fdapace` (Dai et al., 2018)
- FDA in :
  - ✓ `fdasrsf`
  - ✓ `pyFDA`
- *Let's go to the tutorial session!*

References I

X. Dai, P. Hadjipantelis, K. Han, and H. Ji. *fdapace: Functional Data Analysis and Empirical Dynamics*, 2018. R package version 0.4.0.

M. Febrero and M. Oviedo de la Fuente. Statistical computing in functional data analysis: The R package fda.usc. *Journal of Statistical Software*, 51:1–28, 2012.

M. Febrero, P. Galeano, and W. González. Outlier detection in functional data by depth measures, with application to identify abnormal nox levels. *Environmetrics*, 19:331–345, 2008.

F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics. Springer–Verlag, 2006.

S. Gallón. *Template Estimation for Samples of Curves and Functional Calibration Estimation via the Method of Maximum Entropy on the Mean*. PhD thesis, Institut de Mathématiques de Toulouse, Université Toulouse III - Paul Sabatier, Toulouse, France, 2013.

P. Green and B. Silverman. *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman& Hall/CRC, Boca Raton, 1994.

P. Hadjipantelis and H. G. Müller. Functional data analysis for big data: A case study on california temperature trends. In L. H. Härdle, W. and X. Shen, editors, *Handbook of Big Data Analytics*. Springer, 2018.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, New York, 2nd. edition, 2009.

L. Horváth and P. Kokoszka. *Inference for Functional Data with Applications*. Springer, 2012.

R. Hyndman and H. Shang. Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics*, 19(1):29–45, 2010.

R. Hyndman and M. Ullah. Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, 51(10):4942–4956, 2007.

G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning with Applications in R*. Springer Series in Statistics. Springer, New York, 2013.

References II

J. O. Ramsay and B. W. Silverman. *Applied Functional Data Analysis: Methods and Case Studies*. Springer-Verlag, 2002.

J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer, New York, 2nd edition, 2005.

J. O. Ramsay, H. Wickham, S. Graves, and G. Hooker. *fda: Functional Data Analysis*, 2017. R package version 2.4.7.

H. Shang. *Visualizing and forecasting functional time series*. PhD thesis, Monash University. Department of Econometrics and Business Statistics, 2010.

H. Shang. A survey of functional principal component analysis. *AStA Advances in Statistical Analysis*, 98(2): 121–142, 2014.

H. Shang and R. Hyndman. *rainbow: Rainbow Plots, Bagplots and Boxplots for Functional Data*, 2016. R package version 3.4.

N. M. Tran. *An introduction to theoretical properties of functional principal component analysis*. PhD thesis, Department of Mathematics and Statistics, The University of Melbourne, Victoria, Australia, 2008.

J.-L. Wang, J.-M. Chiou, and H.-G. Müller. Review of functional data analysis. *Annual Review of Statistics*, 3: 257–295, 2016.

Y. Wang. *Smoothing splines: methods and applications*. CRC Press, 2011.

J. Zhang. *Analysis of Variance for Functional Data*. CRC Press, 2013.

Thanks!!!