The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

1863-14

Advanced School and Conference on Statistics and Applied Probability in Life Sciences

24 September - 12 October, 2007

Longitudinal and Functional Data Analysis
Some notions of Functional Data Analysis

Anestis Antoniadis

LJK- Universite Joseph Fourier
Grenoble, France

Longitudinal and Functional Data Analysis

Some notions of Functional Data Analysis

Anestis Antoniadis
LJK-Université Joseph Fourier

Trieste, October 2007

Introduction

Measurements treated in the FDA literature typically are recorded by high frequency automatic sensing equipment and therefore the data object is an entire curve. From a theoretical point of view, (stochastic) models and statistical analysis of a functional data set Functional Data Analysis – can be taken often one-to-one from the conventional multivariate analysis.

A first method how to deal with the functional data is to discretize them and perform a standard multivariate analysis on the resulting random vectors. The natural question is then: "what makes a data set functional?" when is it appropriate to understand a vector $\{v_j,\ j = 1, \ldots, d\}$ as a vector of discretized functional values $v_j = v(t_j)$, j =1, . . . , d?

What are functional data?

Heuristically, a data set is functional when there is a reason to assume that the discretized function v possess some features that are meaningful only in the functional context. Probably the most important example of such a feature is the smoothness.

In mathematical terms functional data are observations of some function space valued random elements.

The sample paths are usually observed on a fine discrete grid with some noise measurement process added

Data on individual i $(i = 1, \ldots, n)$:

$$X_{ij} = x_i(t_{ij}) + \epsilon(t_{ij}), \quad j = 1, \ldots, n_i$$

Why functional data?

- Real world phenomena are usually continuous at small enough time scale. Lot of interesting objects are rather functions than vectors (Finance: Yield-curves, Supply/Demand-curves; Economy: Macro-indexes; Biology: Growth-curve, etc . . .)
- Smoothness make sense in FDA
- FDA overcomes the curse of dimensionality; For smooth functions with bounded derivative the instrinsic dimension is finite and typically the practical dimension is 10-20.

Goals of FDA

- Exploratory data analysis
- Characterizing homogeneity and patterns of variability among curves, and identifying unusual ones.
- Assessing the relationships of shapes of curves to covariates.
- Prediction: Prediction of the future.

Basic Setup

A random object in FDA can be modeled as a measurable function defined on a probability space with values on some separable Hilbert space of functions.

For the purpose of FDA, the most often used function space is the space of the Lebesque integrable functions L_J^2 on some subset of $J \subset \mathbb{R}^q$ or more complicated type of function spaces – *Sobolev spaces* – spaces of smooth Lebesque integrable functions.

Typical assumptions

One usually assumes the existence of the expected value, variance and continuous covariance and correlation function of the functional random variable X and denote these by

- $\mu(t) = \mathbb{E}(X(t)), t \in J,$
- $\sigma^2(t) = \mathbb{E}\{(X(t) - \mu(t))^2\}, t \in J,$
- $\gamma(s,t) = \mathbb{E}\{(X(s) - \mu(s))(X(t) - \mu(t))\}, s,t \in J,$
- $\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\sigma^2(t)\sigma^2(s)}}, t,s \in J.$

The $\rho(s,t)$ is defined under the assumption $\sigma^2(t), \sigma^2(s) > 0$. Then $\mathbb{E}(\|X - \mu\|_H^2) = \int \gamma(t,t)dt < \infty$ and the covariance operator Γ of X is given by

$$(\Gamma v)(t) = \int \gamma(t,s)v(s)ds, \quad v \in L_J^2.$$

Empirical counterparts

Given a functional sample $\{X_i(t), t \in J\}$, $i = 1, \ldots, n$ estimates of the previous moments are constructed as straightforward generalizations of the multivariate counterparts:

- Mean $\bar{X}(t) = n^{-1} \sum_{i=1}^{n} X_i(t)$,
- Variance function $\hat{\sigma}^2(t) = (n-1)^{-1} \sum_{i=1}^{n} (X_i(t) - \bar{X}(t))^2\}$.
- Covariance function
  $\hat{\gamma}(s,t) = (n-1)^{-1} \sum_{i=1}^{n} \{(X_i(s) - \bar{X}(s))(X_i(t) - \bar{X}(t))\}$
- Correlation function $\hat{\rho}(s,t) = \frac{\hat{\gamma}(s,t)}{\sqrt{\hat{\sigma}^2(t)\hat{\sigma}^2(s)}}$.

The point-wise consistency of these estimators can be obtained using standard multivariate results. The covariance operator can be approximated by the empirical covariance operator

$$(\hat{\Gamma} v)(t) = \int \hat{\gamma}(t,s) v(s) ds.$$

Cross-correlation

Given pairs of random functional $\{(X_i(t), Y_i(t)), t \in J\}$,$i = 1, \ldots, n$ one may also compute:

- the cross-covariance
  $\hat{\text{Cov}}_{X,Y}(s,t) = (n-1)^{-1} \sum_{i=1}^{n} \{(X_i(s) - \bar{X}(s))(Y_i(t) - \bar{Y}(t))\}$
- the cross-correlation function $\hat{\text{Corr}}_{X,Y}(s,t) = \dfrac{\hat{\text{Cov}}_{X,Y}(s,t)}{\sqrt{\hat{\sigma}_X^2(t)\hat{\sigma}_Y^2(s)}}$.

Goals for This Course

Provide an overview of relevant FDA techniques and concentrate mainly on

- Some of the basics way of exploring functional data.
- An overview of functional principal components
- Some exposure to functional regression and nonparametric mixed models.

Again, as for LDA, these lectures (alone) will not train you comprehensively in the various nuances required for good applied functional data analysis.

References

FDA ideas have been around a while . . .

– Philippe Besse and Jim Ramsay worked on it early and formalized it as a stats area; currently uses exploratory approach, derives models via differential equations

seminal paper by Rice and Silverman (1991) - pca

French schools in Toulouse and Grenoble: lots of theory in function space

Müller : trying to do something model-based

big flurry of current activity linking FDA with longitudinal/mixed effects analysis (Ruppert, Wand and Carrol; Morris; Antoniadis and Sapatinas, Guo, Wu and Zhang, etc. . .)

FDA is still very much in its infancy. Areas of current interest:

– making inference in FDA

– trying to formulate model-based FDA

Books

- J. RAMSAY AND B. W. SILVERMAN (2002). *Applied Functional Data Analysis*. Springer-Verlag, Berlin, Heidelberg.
- D. RUPPERT, M.P. WAND, AND CARROLL (2003). *Semiparametric Regression*. Cambridge University Press, Cambridge.
- J. RAMSAY AND B. W. SILVERMAN (2005). *Functional Data Analysis*. Springer-Verlag, Berlin, Heidelberg, 2nd edition.
- F. FERRATY AND P. VIEU (2006). *Nonparametric Functional Data Analysis. Theory and Practice*. Springer-Verlag, Berlin, Heidelberg.
- H. WU AND J.-T. ZHANG (2006) *Nonparametric Regression methods for Longitudinal Data Analysis*, Wiley Series in Probability and Statistics, Wiley & Sons, New York.

Example: the Tecator Data

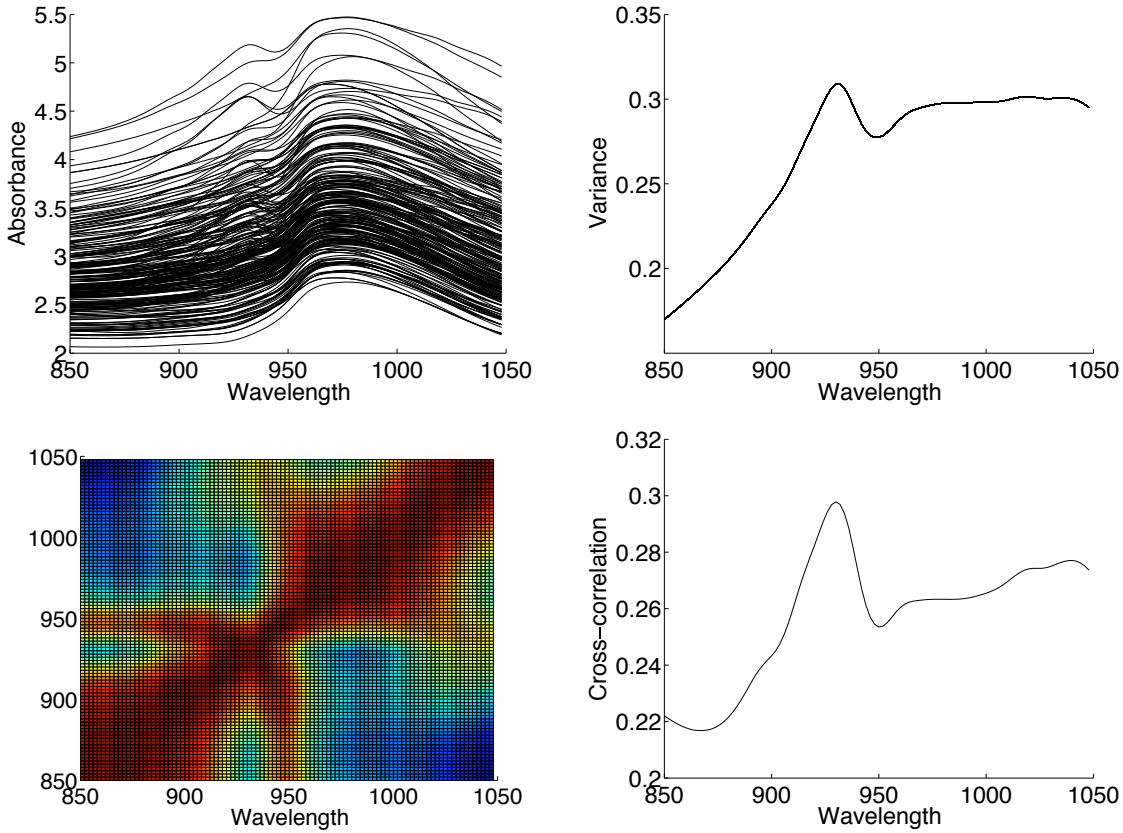
The tecator data are recorded by a Tecator near-infrared spectrometer (the Tecator Infratec Food and Feed Analyzer) which measures the spectrum of light transmitted through a sample of minced pork meat in the region 850–1050 nm.

Each sample contains finely chopped pure meat with different moisture, fat and protein contents. For each meat sample the data consists of a 100 channel spectrum of absorbances and the fat contents as output. The total number of samples is 240.

The aim is to predict the percentage of each content given the corresponding spectrometric curve X.

The cross-correlation with the output can be misleading.

Tecator Data



From left to right: the inputs, the variance function, the correlation function and the cross-correlation with the scalar output (fat content).

Functional Basis Expansion

A popular approach in FDA, motivated by the fact that one needs a flexible method for constructing approximations to the observed data than can track local curvature is the functional basis expansion technique.

Consider a functional basis on an interval J, i.e. a linear independent set of functions $\{\phi_k\}_{k=1}^{\infty}$ that spans the function space H. Prominent and popular examples of such functional bases are:

- the set of monomials $\{t^k\}_{k=0}^{\infty}$
- the Fourier basis
- Splines or B-spline bases

Basis expansion: the functional inputs $\{X_i(t)\}_{i=1}^{n}$ are approximated as (for some finite $K > 0$)

$$X_i(t) \approx \sum_{k=1}^{K} \theta_k \phi_k(t).$$

K and θ_k the shape of the approximation.

What do we want from basis functions?

- Fast computation of individual basis functions.
- Flexible so they can exhibit the required curvature where needed, but also be nearly linear when appropriate.
- Fast computation of the coefficients θ_k:
- Smooth as required since one makes lots of use of derivatives in functional data analysis
- Constrained as required, such as periodicity, positivity, monotonicity, asymptotes and etc. ...

Moreover, the approximation should somewhat denoise and also provide dimension reduction. Hence an appropriate criterion for determining the θ_k's is needed.

Fourier Basis

A well known basis for periodic functions on the interval J is the Fourier basis, defined on J by

$$\phi_0(t) = \frac{1}{\sqrt{|J|}}, \quad \phi_{2r-1}(t) = \frac{1}{\sqrt{|J|/2}} \sin(r\omega t), \quad \phi_{2r}(t) = \frac{1}{\sqrt{|J|/2}} \cos(r\omega t),$$

for $r = 1, \ldots, L/2$ where L is an even integer. The frequency ω determines the period and the length of the interval $|J| = 2\pi/\omega$.

The Fourier basis defined above is an orthonormal basis. The popularity of this basis is based partially on the possibility of fast coefficient calculation by the Fast Fourier Transformation (FFT) Algorithm ($O(n \log n)$ complexity)

Another important feature of the Fourier series is the existence of continuous derivatives of any order. A limitation is that Fourier series are only natural if $X(t)$ is periodic and extremely smooth and do not perform well for functions with strong local features, like discontinuity points in lower order derivatives.

Polynomial Basis

The polynomial basis, appropriate for non-periodic functions is defined by

$$\phi_k(t) = t^k, \quad k = 0, 1, \ldots, L-1.$$

The polynomial functions are easy to calculate, for example by a simple recursion. The calculation of derivatives is also very imple and fast.

They are okay for approximating very simple curves such as linear or slight curvature but not flexible enough to focus on specific locations . Moreover, high order polynomials become too fluctuating especially in the boundaries of J.

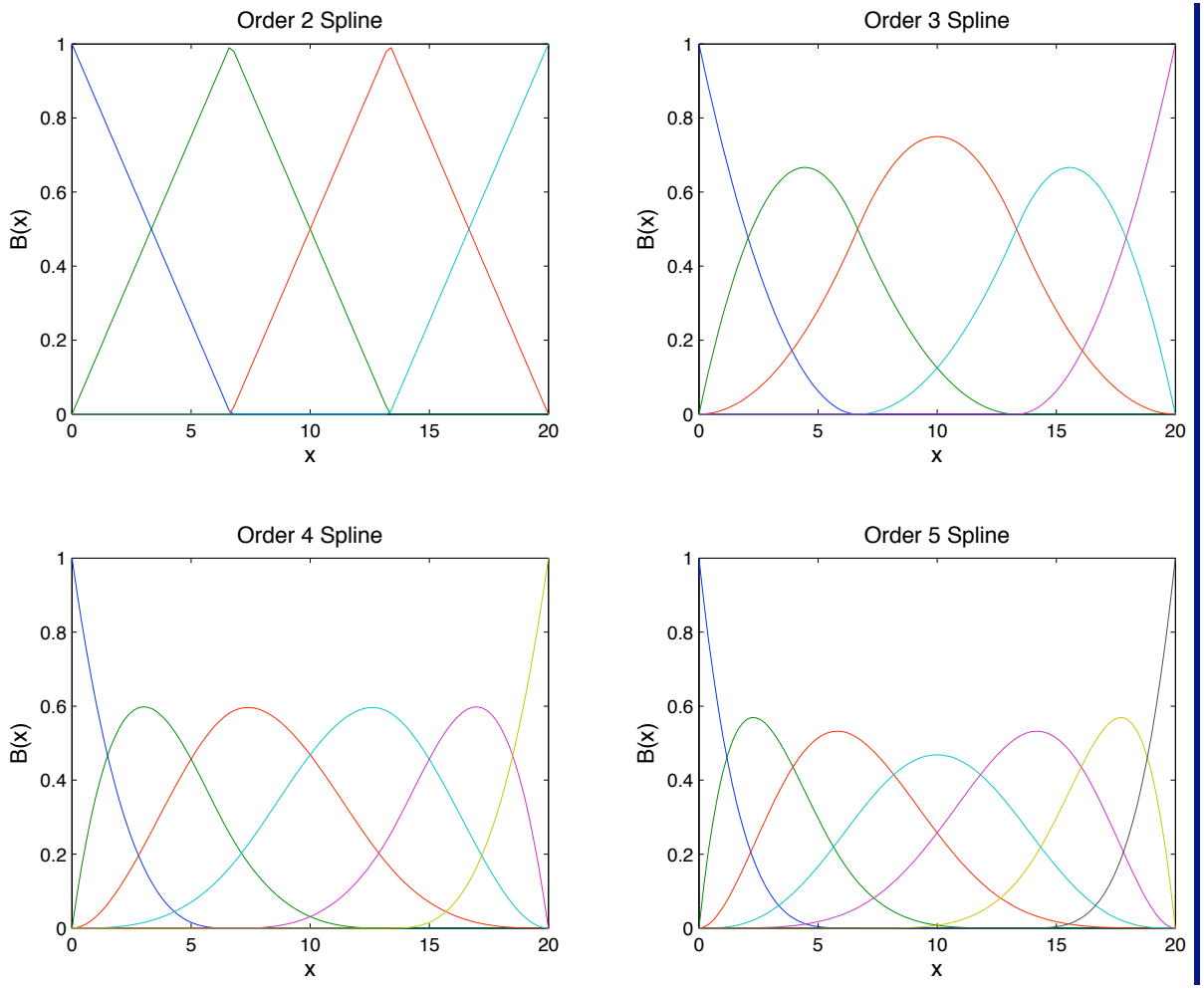
Polynomial basis as defined above is not orthogonal. However several modified types of polynomial systems exist, that are orthogonal, e.g. Legendre polynomials.

Spline Bases

Very popular functional bases for non-periodic data are spline bases. They have the potential to model sharp changes in the observed curve $X(t)$ as as well as its smooth variation. They are very flexible, relatively fast to evaluate as well as their derivatives.

- Consider a sequence of knots on the interval J, $\boldsymbol{\tau} = \{\tau_\ell\}_{\ell=0}^{L}$ with $\tau_0 = \min(J)$ and $\tau_L = \max(J)$.
- A spline with knots $\boldsymbol{\tau}$ is a piecewise polynomial of order K (degree $K-1$) smoothly connected at the knots.
- At the knots, it is required that the values of the polynomial pieces and derivatives up to $K-2$ agree.
- The number of the basis functions is uniquely defined by the spline order and the number of knots. $((K+L-1))$.
- You can have multiple knots at a point. For each additional knot, the spline function will have one less derivative at that knot.

Some B-spline bases



B-Spline bases for different orders. Note that B-splines have local support and are positive.

Approximation and Coefficient Estimation

When using a fixed number K of basis functions for approximating $X(t)$ one may write

$$X(t) = \sum_{k=1}^{K} \theta_k \phi_k(t) + \epsilon(t),$$

where ϵ models the approximation error.

The approximation looks like a linear regression model w.r.t. to the transformed variables $\phi_k(t)$ and a natural candidate for estimating the coefficient vector $\boldsymbol{\theta}$ is by minimizing some loss function.

Smoothing data by least squares

The choice of a quadratic loss function leads to the least squares criterion, i.e. fit the observed discretized curve $\mathbf{x} = \{x(t_j)\}_{j=1,\ldots,T}$ using the model

$$x(t_j) = \sum_{k=1}^{K} \theta_k \phi_k(t_j) + \epsilon_j,$$

and interpolate the data using

$$\hat{x}(t) = \sum_{k=1}^{K} \hat{\theta}_k \phi_k(t)$$

Least Squares fitting

Define by Φ the $T \times K$ matrix with values $\phi_k(t_j)$. Then ordinary least squares fits the data by minimizing with respect to $\boldsymbol{\theta}$

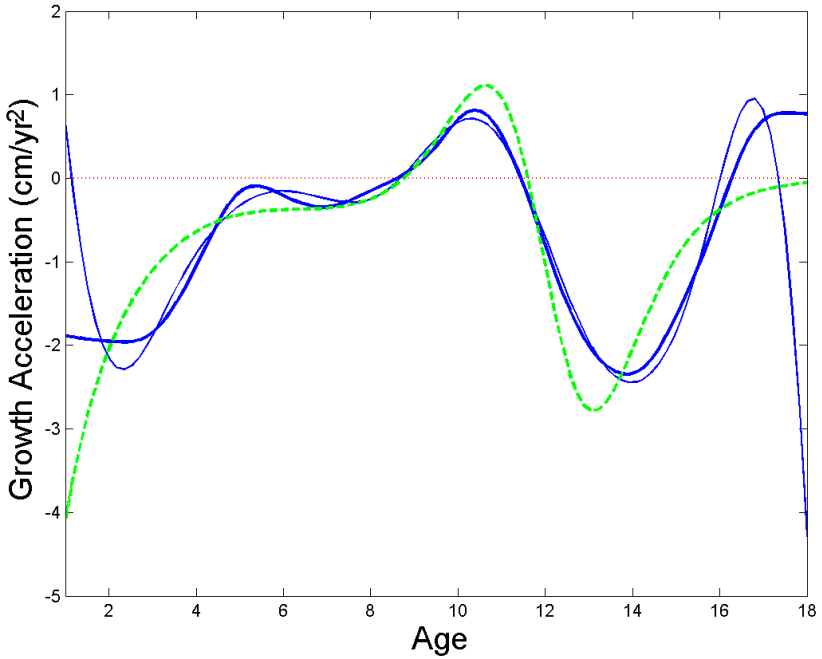
$$\|\mathbf{x} - \Phi\boldsymbol{\theta}\|^2$$

and leads to

$$\hat{\mathbf{x}} = \Phi\hat{\boldsymbol{\theta}} = \Phi(\Phi'\Phi)^{-1}\Phi'\mathbf{x}$$

This is of course appropriate for i.i.d. residuals with zero mean and constant variance. When the errors are nonstationary or autocorrelated then one uses weighted least squares.

Example: pubertal age



Two estimates for the acceleration. As one can see we get very strong boundary effects and the fit needs improvements there for the interpolation to be meaningful.

How many functions?

In the nonparametric regression, K is however unkown and needs to be estimated from the data important question for practitioners is how many functions should be used in the basis expansion.

This problem is essentially equivalent to the bandwidth-choice in the local polynomials and is related to the standard Bias/Variance dilemma:

- High K
  - High order expansion gives a better approximation to data
  - $\text{Bias}[\hat{x}(t)] = x(t) - \mathbb{E}[\hat{x}(t)]$ is small
  - Fit of noise or wrong variations
- Low K
  - Miss of important aspects of estimated function
  - $\text{Var}[\hat{x}(t)] = \mathbb{E}\{(\hat{x}(t) - \mathbb{E}[\hat{x}(t)])^2\}$ is small

One usually uses penalized least squares estimates to achieve the best MSE.

Principal Components for Functional Data

- In the multivariate framework PCA is usually used when we want to find the dominant features and modes of variation in the data, usually after subtracting the mean from each observation.
- We want to know how many of these modes of variation are required to achieve a satisfactory approximation to the original data (dimension reduction).
- It may be assumed that keeping only dominant modes will improve the signaltonoise ratio of what we keep (denoising).

The functional version of principal component analysis has much more important role than its multivariate version, in fact it is often the only way to describe and work with distribution of random functions in practice.

PCA for multivariate data

Let $\mathbf{x}$ be a p-dimensional random vector and let X be the $N \times p$ matrix whose rows form a sample of size N from $\mathbf{x}$. The aim is to find normalized weight vectors $\gamma_r \in \mathbb{R}^p$ for which the linear transformations of the p-dimensional random vector $\mathbf{x}$:

$$\beta_r = \gamma_r'(\mathbf{x} - \mathbb{E}(\mathbf{x})) = \langle \gamma_r, \mathbf{x} - \mathbb{E}(\mathbf{x}) \rangle.$$

has maximal variance subject to:

$$\gamma_\ell' \gamma_r = \langle \gamma_\ell, \gamma_r \rangle = \delta_{\ell,r}.$$

The problem is solved by the means of the Jordan spectral decomposition of the covariance matrix, the r-th principal component is the eigenvector of covariance matrix $C = \frac{1}{N} X'X$ corresponding to the r-th largest eingenvalue.

Functional PCA

In FDA the motivation for Functional Principal Components Analysis (FPCA) as the dimension reduction technique can be done via the same route: having a random function X find orthonormal weight functions γ_1, $\gamma_2, \ldots$, such that the variance of the linear transformation

$$\beta_r = \langle \gamma_r, X - \mu \rangle = \int_J \gamma_r(t)\{X(t) - \mu(t)\}dt,$$

is maximal, w.r.t. to the orthonormal weight functions γ_r, i.e. such that:

$$\|\gamma_r\|^2 = \int \gamma_r^2(t)dt = 1, \quad \langle \gamma_\ell, \gamma_r \rangle = \int \gamma_r(t)\gamma_\ell(t)dt = \delta_{\ell,r}.$$

The solution is achieved by γ_r, the normalized eigenfunction of the covariance operator Γ corresponding to the r-th largest eigenvalue λ_r, i.e. by solving the integral eigen-equation

$$\int \gamma(s,t)\gamma(t)dt = \lambda\gamma(s), \quad s \in J.$$

Karhunen-Loève Expansion

The theoretical basis for the application of the FPCA as a dimension reduction tool is given by the Karhunen-Loève Expansion (KL) of X obtained by:

$$X = \mu + \sum_{r=1}^{\infty} \beta_r \gamma_r.$$

The factor loadings

$$\beta_r = \langle \gamma_r, X - \mu \rangle$$

are uncorrelated, mean zero and with variance λ_r.

Clearly the distribution of X can be analyzed by analyzing the structure of the eigenfunctions γ_r and of the (one-dimensional) factor loading β_r.

Properties of KL-Expansion

Smoothness If X is a random function with realizations that are twice continuously differentiable with probability 1 then this will be also true for the eigenfunctions γ_r.

Best Empirical Basis The first L principal components provide a "best basis" for approximating the sample functions in terms of the integrated square error. More precisely, for any choice of L orthonormal basis functions $v_1, \ldots, v_L$ the mean integrated square error:

$$\mathbb{E}(\|(X - \mu - \sum_{r=1}^{L} \langle X - \mu, v_r \rangle \, v_r\|^2)$$

is minimized by $v_r = \gamma_r$.

The later property is of high importance in the application of the FPCA. In many important applications a small number of functional principal components will suffice to approximate random function X with a small (residual) error justifying a truncated KL-expansion.

Estimation of Functional Principal Components

The estimation of FPC is done using same arguments as in multivariate PCA: for a given sample X_i of size n generated by X an empirical analog of the KL-expansion can be constructed by using eigenvalues $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots$ and orthonormal eigenfunctions $\hat{\gamma}_1, \hat{\gamma}_2, \ldots$ of the empirical covariance operator $\hat{\Gamma}$.

Under appropriate assumptions one even has asymptotic results on $\hat{\lambda}_r$ and $\hat{\gamma}_r$ (see Dauxois, Pousse and Romain (1982)) such as, for example,

$$\|\gamma_r - \hat{\gamma}_r\| = O_P(n^{-1/2}).$$

Challenges

- How to deal with functions in practice?
- Functional values are observed just at discrete grid
- Functions may be observed with additional error
- How to do inference in practice ?

The naive method

In the standard FDA setup, one assumes that the n functions X_i are observed without additional error on a fine discrete grid of size T and the analysis proceeds "as if" the functions were directly observed. Using the discrete data

- Take the $n \times T$ data matrix X of finely sampled values of X_i
- solve the eigenvalue problem for $V = n^{-1}X'X$
- to obtain an approximate eigenfunction γ_r from the discrete values, use any convenient interpolation method

The functional basis approach

Denote a functional basis on the interval J by $\{\phi_1, \phi_2, \dots, \}$ and assume that the functions X_i are approximated by the first K basis functions of this basis:

- Express each X_i as

$$X_i(t) = \sum_{k=1}^{K} \theta_{ik}\phi_k(t) = \boldsymbol{\theta}_i'\boldsymbol{\phi}(t)$$

- In matrix form $X = \Theta\Phi$. The analysis of the functional objects is implemented through the coefficient matrix Θ. More precisely,
  - The principal components are obtained by a spectra analysis of the matrix $\frac{1}{n}\Theta'\Theta W$ where W is the $K \times K$ Gram matrix corresponding to the basis, i.e. $W_{k\ell} = \langle \phi_k, \phi_\ell \rangle$.

Easy implementation, in fact one analyzes Θ; Works fine if you use proper basis; Choice of L influences the quality of the results.

Smoothed Functional Principal Components

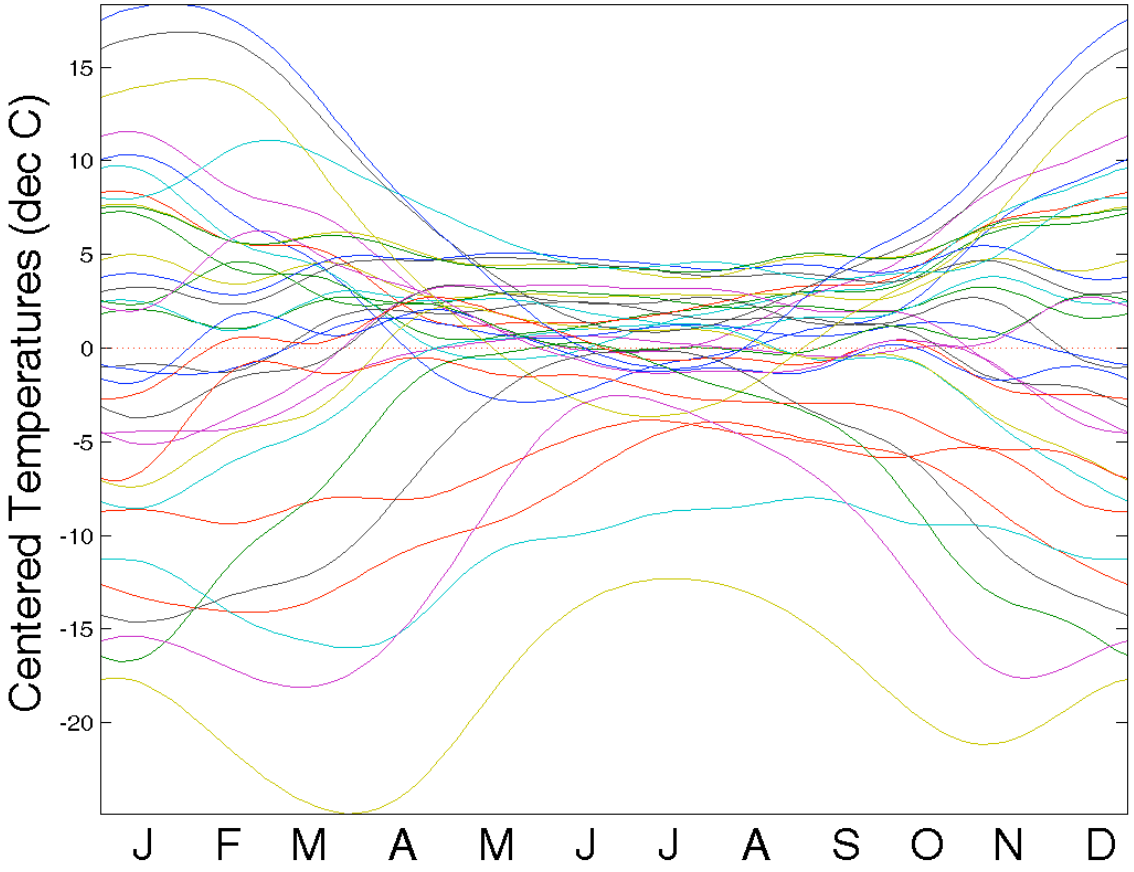
Often the functions $X_i(t)$ are observed with additional error or affected by wrong variations: combine then optimization criterion with a roughness penalty as it is done traditionally in nonparametric regression.

- use as a roughness penalty $\text{pen}(\gamma) = \|\mathcal{D}^2\gamma\|^2$
- unsmoothed PCA maximizes sample variance $\text{Var}\{\langle\gamma, X - \mu\rangle\}$
- Maximize instead the penalized variance

$$\frac{\text{Var}\{\langle\gamma, X - \mu\rangle\}}{\|\gamma\|^2 + \lambda\text{pen}(\gamma)}$$

- Smoothing parameter $\lambda \geq 0$ chosen with cross-validation
- constrains:
  * $\|\gamma_r\|^2 = 1$
  * $\int \gamma_r(t)\gamma_\ell(t)dt + \int \mathcal{D}^2\gamma_r(t)\mathcal{D}^2\gamma_\ell(t)dt = 0, \quad \ell = 1, \ldots, r-1.$

Example : A PCA of monthly temperature curves

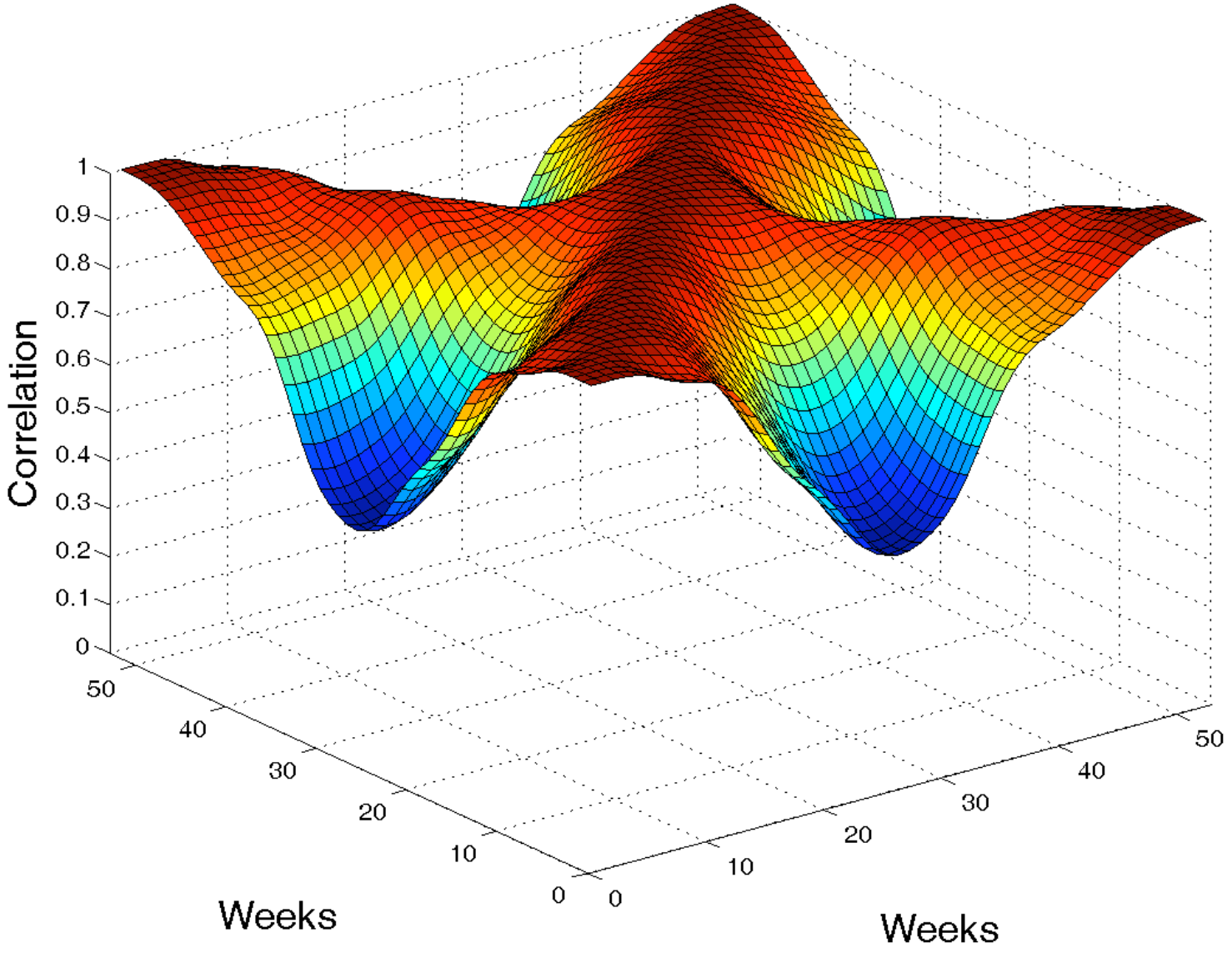


30-year centered monthly temperatures for each of 35 Canadian weather stations.

What do we see?

- some curves are high (warm) and some curves are low (cold)
- some curves have larger variation between summer and winter than others.
- How much of the variation do these two types of variation account for?

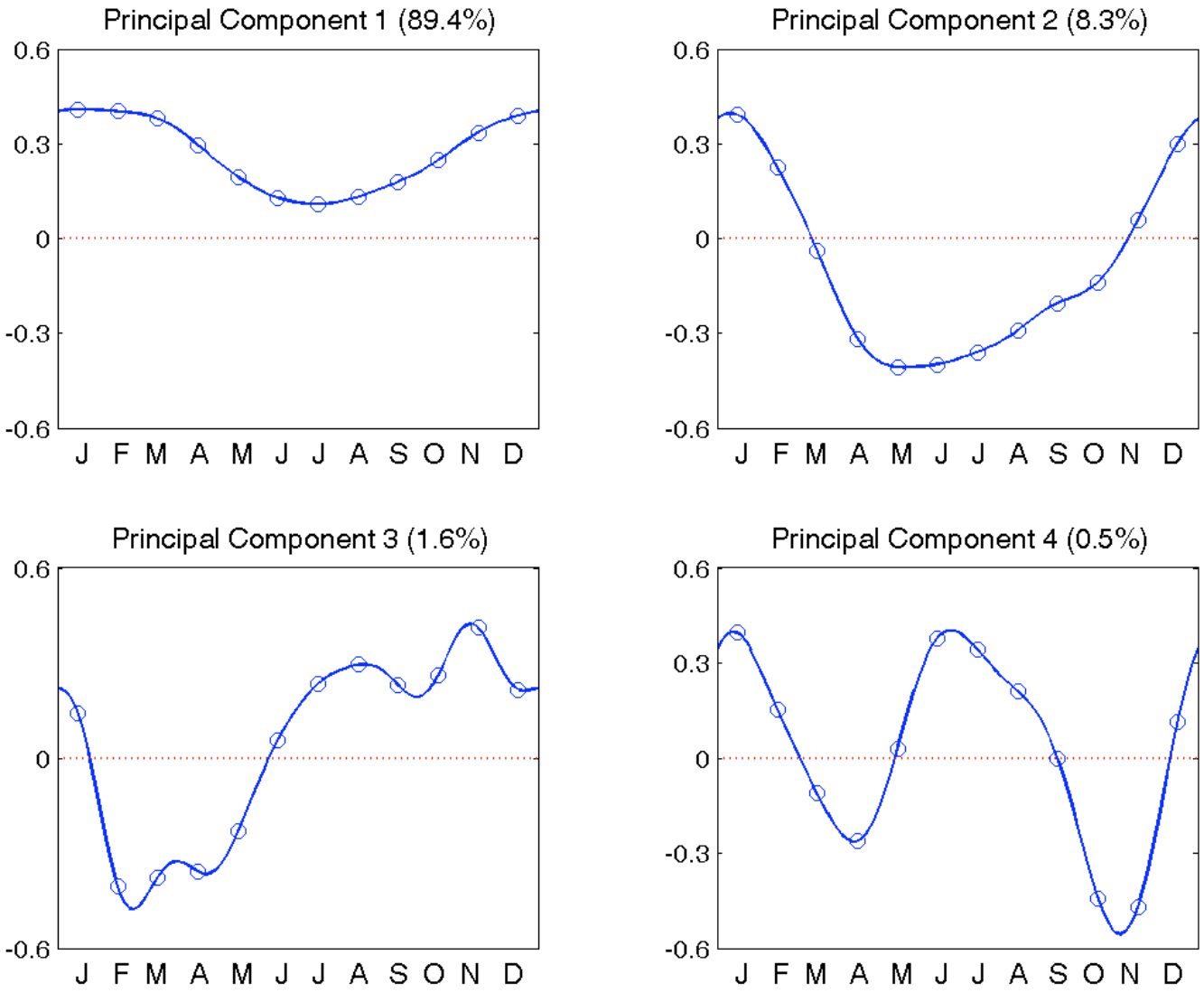
The correlation surface



What do we see?

- The diagonal ridge and the unit "wing-tips" correspond to unit correlations between temperatures at identical times.
- The ridge perpendicular to this corresponding to correlations between temperatures symmetrically placed around midsummer.
- Correlations fall off much more rapidly for times symmetric about March and September 21.

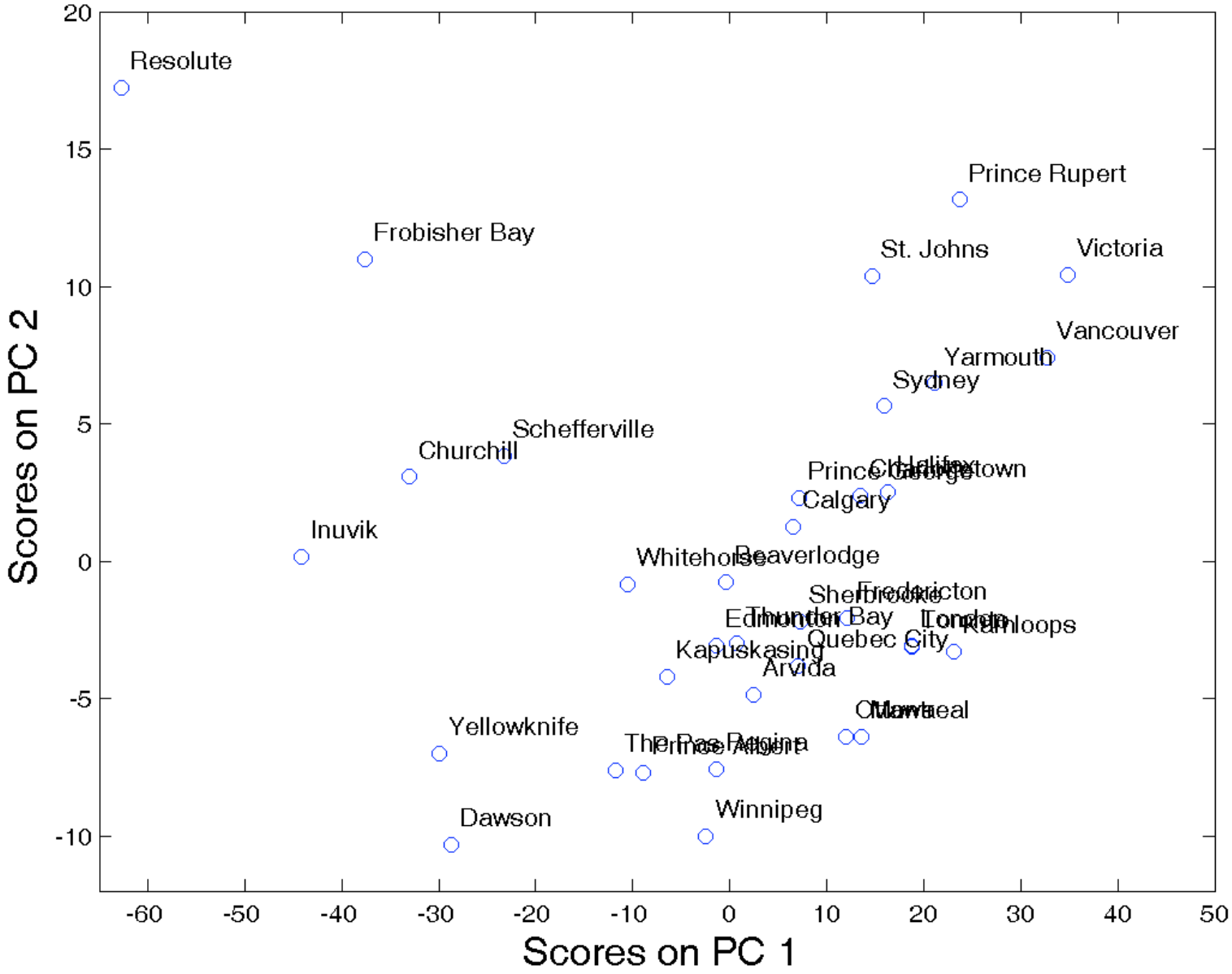
The first four principal components



What do we see?

- The two components that we saw in the centered curves account for about 98% of the variation.
- The first four components account for 99.8% of the vari- ation.
- The first four components tend to look like linear, quadratic, cubic and quartic polynomials, respectively.

The first two principal component scores



What do we see?

- Most stations are along a curved line running from lower center to top right.
- At the top end of the banana are maritime stations with less variation between winter and summer, and high average temperatures.
- At the lower end are the continental stations with large seasonal variation and lower average temperatures.
- The Arctic stations are in their own space with large seasonal variation and very low average temperatures.

Functional Data Analysis and Mixed-Effects

Basic model: Y_{ij} response of the i-th subject at time point t_{ij}

$$Y_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta}(t_{ij}) + \mathbf{Z}_{ij}\boldsymbol{\alpha}^{(i)}(t_{ij}) + \epsilon_{ij}, \quad i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, n_i$$

- $\boldsymbol{\beta}(t) = (\beta_1(t), \ldots, \beta_p(t))'$ is a $p \times 1$ vector of *fixed* functions,
- $\boldsymbol{\alpha}^{(i)}(t) = (\alpha_1^{(i)}(t), \ldots, \alpha_q^{(i)}(t))'$ is a $q \times 1$ vector of stochastically independent *random* functions that are modelled as realisations of zero-mean Gaussian processes $\boldsymbol{a}(t) = (a_1(t), \ldots, a_q(t))'$ (a $q \times 1$ collection of such independent processes) with parametrically structured covariances $\gamma_i(s, t)$,
- $\mathbf{X}_{ij} = (X_{ij}[1], \ldots, X_{ij}[p])$ and $\mathbf{Z}_{ij} = (Z_{ij}[1], \ldots, Z_{ij}[q])$ are, respectively, $1 \times p$ and $1 \times q$ design vectors that can include dummy variables as well as covariates, and
- $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ i.i.d. Gaussian independent of $\boldsymbol{a}(t)$.

Interpretation

Similar to the interpretation of linear mixed-effects models for longitudinal data

- $\mathbf{X}_{ij}\boldsymbol{\beta}(t)$ models the population-average curve profile (fixed effects),
- $\mathbf{Z}_{ij}\boldsymbol{\alpha}^{(i)}(t)$ is the i-th curve-specific deviation (also called the subject-specific deviation if each curve is from a different subject) from the population-average curve profile that accounts for correlation (random effects), and
- $\mathbf{X}_{ij}\boldsymbol{\beta}(t) + \mathbf{Z}_{ij}\boldsymbol{\alpha}^{(i)}(t)$ is the i-th curve-specific function (individual curve).
- n_i are usually quite large and equal to T with T often larger than n.

Targets of Interest

Similar to the goals of in longitudinal data analysis

- Estimate the fixed effects curves $\beta_k(t), k = 1, \ldots, p$,
- Predict the individual i-th curve-specific deviation (random effects),
- Predict the individual curves $\mathbf{X}_{ij}\boldsymbol{\beta}(t) + \mathbf{Z}_{ij}\boldsymbol{\alpha}^{(i)}(t)$.
- Estimate the covariance structure $\gamma_i(s, t)$ of the individual curves.
- Draw some inference about the fixed effects.

Example: Orthosis data

- Interesting data on human movement.
- Data : David Amarantini and Luc Martin, Laboratoire Sport et Performance Motrice, UJF.
- Underlying movement under various levels of an externally applied force to the knee.
- Seven young male volunteers wore a spring-loaded orthosis of ajustable stiffness under 4 experimental conditions:
  - Control condition (without orthosis)
  - Orthosis condition
  - Two conditions (spring1, spring2) stepping in place was perturbed by fitting a spring-loaded orthosis into the right knee.

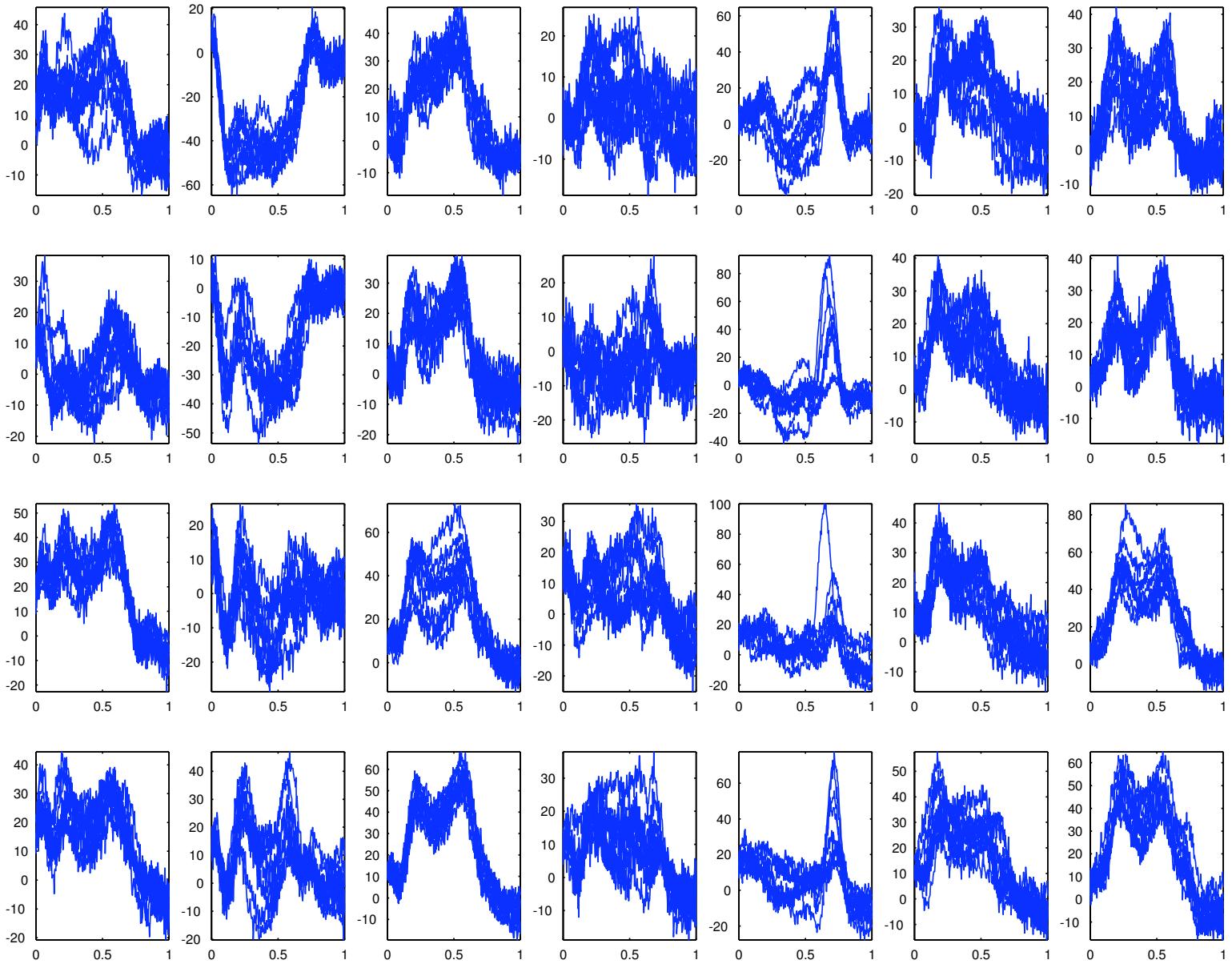
Orthosis Data (2)

The data set consists in 280 separate runs and involves the seven subjects over four described experimental conditions, replicated ten times for each subject.

Panels in rows correspond to **treatments** while the panels in columns correspond to **Subjects**. The number of time recordings T per replications is 256.

Averaging over the 10 repetitions for each subject and treating subjects as random effects, we have $n = 28$, $n_i = T = 256$, $p = 4$ and $q = 1$.

Functional Data Analysis (FDA)



Question

- Understand how a subject can cope with the external perturbation and
- quantify the ways in which the individual mean cross-sectional functions differ over the various conditions.

To do so we will not assume a parametric model neither for the fixed effects $\beta_k(t)$ nor the random effects $\alpha_\ell^{(i)}(t)$, but instead, as it was already done for the FPCA case, we will model these using a function basis expansion, using for example a regression spline basis for dimension reduction.

Basis function expansion

Denote a functional basis on the time interval J on which the data are observed by $\{\phi_1, \phi_2, \ldots, \}$, and assume that both fixed effects components $\beta_r(t), r = 1, \ldots, p$ and random effects $\alpha_\ell^{(i)}(t), \ell = 1, \ldots, q, i = 1, \ldots, n$ are approximated by their expansion in the first K basis functions.

For each $r_1 = 1, \ldots, p$, set

$$\beta_{r_1}(\mathbf{t}) = (\beta_{r_1}(t_1), \ldots, \beta_{r_1}(t_T))',$$

where

- $\mathbf{t} = (t_1, \ldots, t_T)$,
- $\boldsymbol{\phi}(t) = (\phi_1(t), \ldots, \phi_K(t))'$ and
- Φ the $T \times K$ matrix with rows the $\boldsymbol{\phi}(t_j)'$.

Then

- Express each $\beta_{r_1}, r_1 = 1, \ldots, p$ as

$$\beta_{r_1}(t) = \sum_{k=1}^{K} d_{r_1 k}\phi_k(t) = \boldsymbol{\phi}(t)'\mathbf{d}_{r_1}$$

- In matrix form $\beta_{r_1}(\mathbf{t}) = \Phi\mathbf{d}_{r_1}$, where $\mathbf{d}_{r_1}$ is the $K \times 1$ vector of the corresponding basis expansion coefficients.

- For each $i = 1, 2, \ldots, n$ and $r_2 = 1, 2, \ldots, q$, set

$$\alpha_{r_2}^{(i)}(\mathbf{t}) = \Phi\boldsymbol{\theta}_{r_2}^{(i)},$$

where $\boldsymbol{\theta}_{r_2}^{(i)}$ is the $K \times 1$ vector of the corresponding basis expansion coefficients $\{\theta_k^{(r_2,i)},\ k = 1, \ldots, K\}$.

More assumptions

Since Φ is not random the randomness of the $\alpha_{r_2}^{(i)}(\mathbf{t}), i = 1, 2, \ldots, n$, $r_2 = 1, 2, \ldots, q$ is shifted to that of the vectors $\boldsymbol{\theta}_{r_2}^{(i)}$.

Assume that $\boldsymbol{\theta}_{r_2}^{(i)}$ are independent and identically distributed as $N(0, \sigma_\theta^2 D)$ random vectors and also that $\boldsymbol{\epsilon}_i \sim N(0, \sigma_\epsilon^2 \mathbf{I}_T)$ and $\boldsymbol{\theta}_{r_2}^{(i)}$ are independent.

These assumptions implicitly restricts the time-domain covariance of the observed functional data to stationary covariance structures for the between random effects function covariance. However, even with such restrictions the model is flexible enough for often capturing key characteristics of subject specific random variations encountered in practice.

Notation

Let

- $\mathbf{Y}_i(\mathbf{t}) = (Y_{i1}, \ldots, Y_{iT})'$ and $\widetilde{\mathbf{d}} = (\mathbf{d}_1', \ldots, \mathbf{d}_p')'$.
- $\widetilde{\mathbf{X}}_i = \mathbf{X}_i \mathbf{\Phi}^{(p)}$ and $\widetilde{\mathbf{Z}}_i = \mathbf{Z}_i \mathbf{\Phi}_n^{(q)}$, where $\mathbf{X}_i = \text{diag}(\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iT})$ (each element is an appropriately constructed matrix containing dummy variables and/or covariates), where
- $\mathbf{\Phi}^{(p)} = \text{diag}(\Phi_{T\times K}, \ldots, \Phi_{T\times K})$ (p blocks), $\mathbf{Z}_i = \text{diag}(\mathbf{Z}_{i1}, \ldots, \mathbf{Z}_{iT})$,
- $\mathbf{\Phi}_n^{(q)} = \text{diag}(\mathbf{\Phi}^{(p)} \ldots \mathbf{\Phi}^{(p)})$ (n blocks).
- $\widetilde{\boldsymbol{\theta}}_i = (\boldsymbol{\theta}_{i1}, \ldots, \boldsymbol{\theta}_{iq})'$ and $\tilde{\boldsymbol{\epsilon}}_i = (\epsilon_{i1}, \ldots, \epsilon_{iT})'$.

FDA-LME

We have

$$\mathbf{Y} = \widetilde{\mathbf{X}}\widetilde{\mathbf{d}} + \widetilde{\mathbf{Z}}\widetilde{\boldsymbol{\theta}} + \tilde{\epsilon},$$

where

- $\mathbf{Y} = (\mathbf{Y}_1', \ldots, \mathbf{Y}_n')'$,
- $\widetilde{\mathbf{X}} = (\widetilde{\mathbf{X}}_1', \ldots, \widetilde{\mathbf{X}}_n')'$,
- $\widetilde{\mathbf{Z}} = (\widetilde{\mathbf{Z}}_1', \ldots, \widetilde{\mathbf{Z}}_n')'$,
- $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\theta}}_1', \ldots, \widetilde{\boldsymbol{\theta}}_n')'$
- $\widetilde{\epsilon} = (\widetilde{\epsilon}_1', \ldots, \widetilde{\epsilon}_n')'$.

Clearly a *linear mixed-effects model* where the fixed-effects are parameterized by the vector coefficients of $\beta_{r_1}(t)$ ($r_1 = 1, 2, \ldots, p$) and the random-effects are parameterized by the vector coefficients of $\alpha_{r_2}^{(i)}(t)$ ($i = 1, 2, \ldots, n$; $r_2 = 1, 2, \ldots, q$).

Quantifying Variability

$\mathbb{E}(\widetilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\epsilon}})' = (\mathbf{O}_{nKq}, \mathbf{O}_{nT})'$, where $\mathbf{O}_L$ denotes a $L \times L$ matrix with zero entries.

$\text{Var}(\widetilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\epsilon}})' = \text{diag}(\sigma_\theta^2 \boldsymbol{\Sigma}, \sigma_\epsilon^2 \mathbf{I}_{nT})$, where $\boldsymbol{\Sigma} = \text{diag}\,(D, \ldots, D)$ (n-components) and $\mathbf{I}_k$ is the $k \times k$ identity matrix.

The corresponding covariance surface for the Gaussian process modeling the random effect functions $\alpha_{r_2}^{(i)}(t)$ is given by $\sigma_\theta^2 \widetilde{\mathbf{Z}} \boldsymbol{\Sigma} \widetilde{\mathbf{Z}}'$. This matrix describes how the the functions vary one from another and the parameters of D and σ_θ^2 have a clear impact on any inference that is done but such a specification seems unavoidable since the large dimension of the covariance matrices make it infeasible to estimate them in a completely unstructured fashion.

Summary

We can now write

$$\mathbb{E}(\mathbf{Y}) = \widetilde{\mathbf{X}}\widetilde{\mathbf{d}} \qquad \text{and} \qquad \text{Var}(\mathbf{Y}) = \sigma_\epsilon^2 \mathbf{V}_\lambda,$$

where

$$\mathbf{V}_\lambda = \mathbf{I}_{nm} + \lambda \widetilde{\mathbf{Z}} \mathbf{\Sigma} \widetilde{\mathbf{Z}}^{\mathrm{T}}$$

and

$$\lambda = \sigma_\theta^2 / \sigma_\epsilon^2.$$

The parameter λ can be considered as a ratio of the *curve-to-curve variability* and the *within-curve noise*. Note that $\sigma_\theta^2 = 0$ if and only if $\lambda = 0$, and the parameter space for λ is $[0, \infty)$.

Estimation of fixed and prediction of random effects

Using classical linear mixed-effects estimation techniques, the linear mixed-effects representation is used to obtain function basis expansion estimates for both fixed-effects and random-effects.

More specifically, for the estimation of the functional fixed-effects (i.e., the population-average curve profiles), we apply the classical weighted least-squares (WLS) methodology, which is easily implemented.

On the other hand, for the estimation of the functional random-effects (i.e., the curve-specific functions), we apply RML estimation of variance components, as is commonly used in standard liner mixed-models software such as `PROC MIXED` in `SAS` and `lme()` in `S-PLUS` (see, e.g., Ngo & Wand, 2004).

Testing for Random-Effects

Testing for random-effects in the general functional mixed-effects model is equivalent to testing the following hypotheses

$$H_0 : \sigma_\theta^2 = 0 \ (\lambda = 0) \quad \text{versus} \quad H_A : \sigma_\theta^2 > 0 \ (\lambda > 0).$$

Testing is non-standard because the parameter under the null hypothesis is on the boundary of the parameter space. Crainiceanu & Ruppert (2004) and Claeskens (2004) have recently derived finite sample and asymptotic null distributions for the LR and RLR test statistics in linear mixed-effects model with one variance component.

These results can be straightforwardly adapted in the present scenario to obtain the finite sample null distributions of the corresponding LR and RLR test statistics that can be used for testing the hypotheses (see Theorem 3.1 in Antoniadis & Sapatinas, 2007).

Back to the Orthosis data

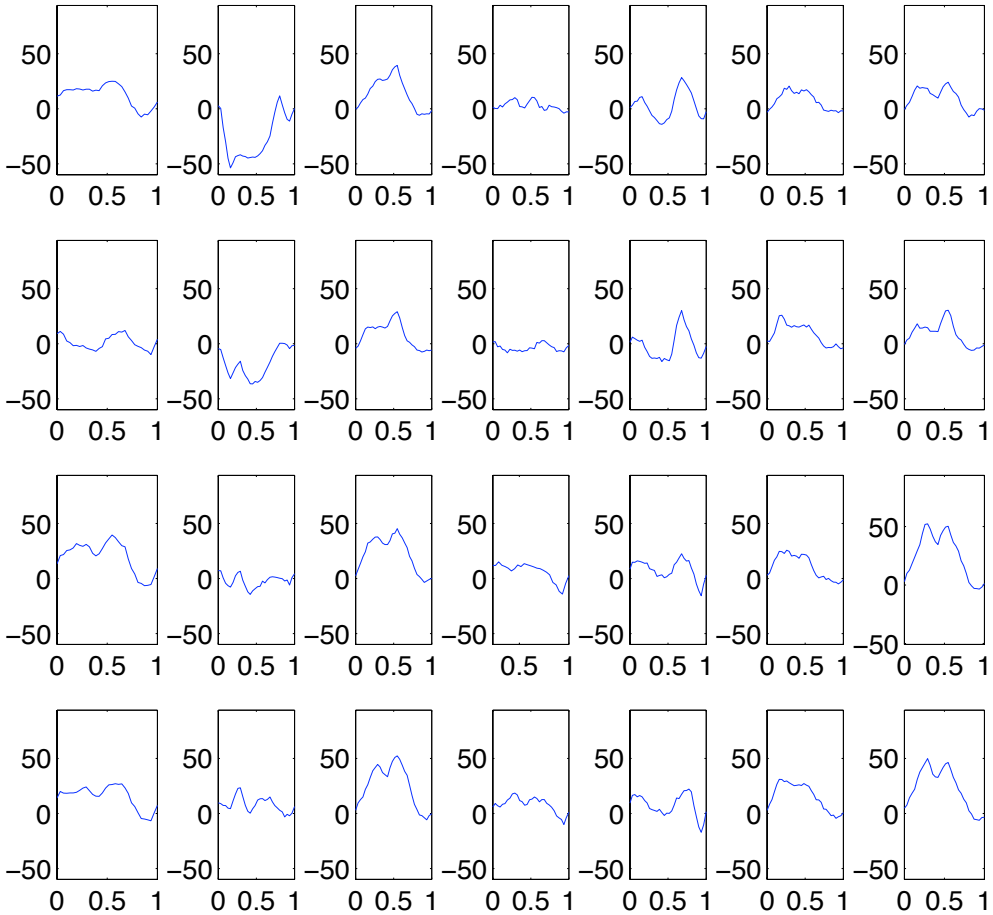
Regarding the functional random-effects, the application of the testing methodology reveals that $\hat{\sigma}_\theta = 40.3127$ and $\hat{\sigma}_\epsilon = 1.0799$ resulting in $\hat{\lambda} = 1393.5311$.

The finite sample RLR test statistic of 3.2743 (p-value of 0.0304), shows that there is significant evidence of random-effects in this case.

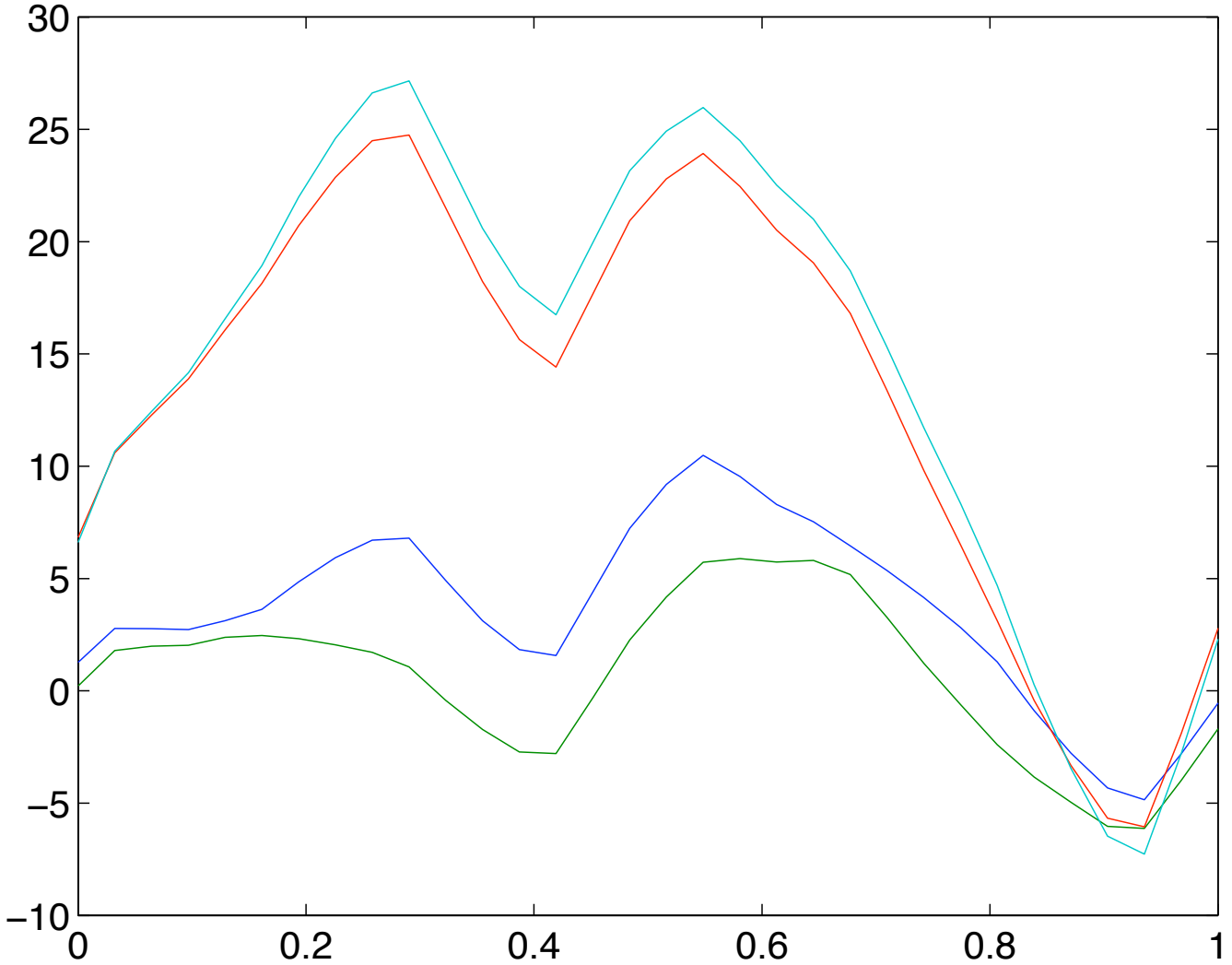
Regarding the functional fixed-effects, a Bonferroni based test statistic value of 4.8472 for Spring 1 vs Spring 2 conditions, 8.9922 for Control vs Orthosis conditions, and 48.7512 for Spring 1 + Spring 2 vs Control + Orthosis conditions.

The various fixed-effects hypotheses of a similar behaviour under the different conditions were all rejected (the overall p-values were 0.0203, 0.0012 and 0 respectively) and supports the fact that individuals adjust their posture differently under perturbations of different nature.

Individual Random-Effects EBLUP predictions



Fixed-Effects Estimates



A functional approach to time series prediction

One seeks information on the evolution of a continuous-time stochastic process $X = (X(t);\ t \in \mathbb{R})$ in the future.

Given a trajectory of X observed on the interval $[0, T]$, one would like to predict the behavior of X on the entire interval $[T, T + \delta]$, where $\delta > 0$, rather than at specific time-points.

Divide $[0, T]$ into $[i\delta, (i + 1)\delta]$, $i = 0, 1, \ldots, n - 1$ with $\delta = T/n$, and consider the discrete time function space valued process $Z = (Z_n;\ n \in \mathbb{Z})$ defined by

$$Z_n(t) = X(t + n\delta), \quad 0 \leq t \leq \delta, \quad n \in \mathbb{Z}. \tag{1}$$

This representation is especially fruitful if X processes a seasonal component with period δ.

ARH(1) Processes

See Bosq (2000) for an excellent account on the theory of linear processes in function spaces.

Let H be a (real) separable Hilbert space, with the inner product $\langle \cdot, \cdot \rangle$ and the norm $|| \cdot ||$. Typically, H is either $L^2[a,b]$) or $W_2^s[a,b]$ the Sobolev space of s-smooth functions on $[a,b] \subseteq \mathbb{R}$ ($s = 1, 2, \ldots$).

Let $\xi = (\xi_n;\ n \in \mathbb{Z})$ be a sequence of H-valued random variables defined on the same complete probability space $(\Omega, \mathcal{F}, \mathcal{P})$. We say that ξ is a (zero mean) ARH(1) process, if

$$\xi_n = \rho(\xi_{n-1}) + \epsilon_n, \quad n \in \mathbb{Z}, \tag{2}$$

where $\rho : H \mapsto H$ is a bounded linear operator and $\epsilon = (\epsilon_n;\ n \in \mathbb{Z})$ is a H-valued strong white noise. Under mild conditions, the above equation has a unique solution which is a weakly stationary process with innovation ϵ.

Let H^* be the (topological) dual of H. The covariance structure of ξ is related to two bounded linear operators from H^* to H, namely the *covariance* and *cross-covariance (of order 1) operators*. Since H^* may be identified with H (by Riesz representation), they are defined respectively by

$$f \in H \longmapsto Cf = \mathbb{E}(\langle \xi_0 \otimes \xi_0)(f))$$

and

$$f \in H \longmapsto D^*f = \mathbb{E}(\langle \xi_1 \otimes \xi_0)(f)),$$

where the tensor product (for two fixed elements in H) $u \otimes v$ is the bounded linear operator from H to H, defined by

$$x \in H \longmapsto (u \otimes v)(x) = \langle u, x \rangle \, v.$$

If $\mathbb{E}||\xi_0||^2 < \infty$, the operator C is then symmetric, positive, nuclear and, therefore, Hilbert-Schmidt. The cross-covariance (of order 1) operator D (the adjoint of D^*) defined by

$$f \in H \longmapsto Df = \mathbb{E}(\langle \xi_0 \otimes \xi_1)(f))$$

is also nuclear and, therefore, Hilbert-Schmidt, and we have the following relations

$$\begin{aligned} D &= \rho C \\ D^* &= C\rho^*, \end{aligned}$$

where ρ^* denotes the adjoint of ρ.

The operators C, D and D^* can be estimated by

$$C_n = \frac{1}{n+1}\sum_{k=0}^{n} \xi_k \otimes \xi_k, \quad D_n = \frac{1}{n}\sum_{k=0}^{n-1} \xi_k \otimes \xi_{k+1}$$

and

$$D_n^* = \frac{1}{n}\sum_{k=0}^{n-1} \xi_{k+1} \otimes \xi_k.$$

C_n, D_n and D_n^* are unbiased estimators of C, D and D^*. Also

$$||C_n - C||_{HS} \to 0 \text{ a.s. and } \mathbb{E}||C_n - C||_{HS}^2 = O\left(\frac{1}{n}\right), \text{ as} \quad n \to \infty$$

$$||D_n - D||_{HS} \to 0 \text{ a.s. and } \mathbb{E}||D_n - D||_{HS}^2 = O\left(\frac{1}{n}\right), \text{ as} \quad n \to \infty$$

and

$$||D_n^* - D^*||_{HS} \to 0 \text{ a.s. and } \mathbb{E}||D_n^* - D^*||_{HS}^2 = O\left(\frac{1}{n}\right), \text{ as} \quad n \to \infty.$$

An estimator of ρ could be based on

$$\rho = DC^{-1}. \tag{3}$$

In a finite-dimensional context, it makes sense, provided the invertibility of C. In our context, however, (3) does not make sense because C^{-1} is unbounded. Now, since C is a compact operator, by the *closed graph theorem* and the fact that the range of D^* is included in the domain of C^{-1}, the adjoint relation to (3):

$$\rho^* = C^{-1}D^* \tag{4}$$

is well-defined. If one is able to estimate ρ^* in (4), then ρ can be estimated using

$$\rho = \text{Ext}(DC^{-1}) = (C^{-1}D^*)^*, \tag{5}$$

where Ext denotes the extension to H of a bounded linear operator defined on a dense subspace of H.

Prediction

The stochastic process X is usually not centered and, therefore, is not weakly stationary. We will assume that its *mean* is a periodic function $a = (a_t;\ t \in \mathbb{R})$ with period δ and, hence, the centered stochastic process $Y = (Y_n = Z_n - a;\ n \in \mathbb{Z})$ is an ARH(1) process. This implies that the best predictor of Z_{n+1} given $Z_n, Z_{n-1}, \ldots$ is obtained by

$$\begin{aligned} \tilde{Z}_{n+1} &= \mathbb{E}(Z_{n+1} \mid Z_n, Z_{n-1}, \ldots) \\ &= a + \rho(Z_n - a), \quad n \in \mathbb{Z}. \end{aligned} \tag{6}$$

If one is able to estimate the (unknown) periodic function a, say by $\hat{a}$, and the 'prediction' operator ρ, say by $\hat{\rho}$, given $Z_0, Z_1, \ldots, Z_n$, then a statistical predictor of Z_{n+1} based on (6) is obtained by

$$\hat{\tilde{Z}}_{n+1} = \hat{a} + \hat{\rho}(Z_n - \hat{a}). \tag{7}$$

Estimation Procedures for the Prediction Operator – A Review

We assume that

$$(H_1): \quad \mathbb{E}(||Y_0||^4) < \infty \quad \text{and} \quad ||\rho^{j_0}||_{\mathcal{L}} < 1 \quad \text{for some} \quad j_0 \geq 1,$$

where $||\cdot||_{\mathcal{L}}$ stands for the supremum norm for bounded linear operators from H to H (then Y is weakly stationary with innovation ϵ). We also assume that $(H_2): \ C$ is one-to-one, otherwise ρ cannot be defined uniquely.

Let also C, D and D^* denote the covariance and cross-covariance (of order 1) operators of Y. These are respectively estimated by C_n, D_n and D_n^* with ξ_k and ξ_{k+1} respectively replaced by $Z_k - \bar{Z}_n$ and $Z_{k+1} - \bar{Z}_n$, where $\bar{Z}_n = \frac{1}{n+1}\sum_{k=0}^{n} Z_k$ is an unbiased estimator of the mean a. The eigenvalues of C and C_n are respectively denoted (in decreasing order) by $\lambda_1 \geq \lambda_2 \geq \ldots$ and $\lambda_{1,n} \geq \lambda_{2,n} \geq \ldots$ with corresponding eigenfunctions respectively denoted by $e_1, e_2, \ldots$ and $e_{1,n}, e_{2,n} \ldots$.

Continuous-Time Data

Let Π^{k_n} the projection operator onto the space spanned by the first k_n eigenfunctions of C_n. Define the class of *projection estimators* by

$$\rho_n^* = \left(\Pi^{k_n} C_n \Pi^{k_n}\right)^{-1} D_n^* \Pi^{k_n}. \tag{8}$$

The (random) operator $\left(\Pi^{k_n} C_n \Pi^{k_n}\right)^{-1}$ is defined by inverting the operator $\left(\Pi^{k_n} C_n \Pi^{k_n}\right)$ and completing it by the null operator on the subspace orthogonal to the space spanned by the first k_n eigenfunctions of C_n.

To obtain asymptotic results, it is assumed that $\left(\Pi^{k_n} C_n \Pi^{k_n}\right)$ is almost surely invertible after some $n \in \mathbb{Z}$. Under the extra assumption

$$(H_4): \qquad n\lambda_{k_n}^4 \to \infty \quad \text{and} \quad \frac{1}{n}\sum_{k=1}^{k_n} \frac{g_k}{\lambda_k^2} \to 0, \quad \text{as} \quad n \to \infty,$$

where $g_k = \max\left\{(\lambda_{k-1} - \lambda_k)^{-1}, (\lambda_k - \lambda_{k+1})^{-1}\right\}$, one can then prove that (see Mas, 2000) for any $f \in H$

$$||\rho_n^* f - \rho^* f|| \xrightarrow{p} 0, \quad \text{as} \quad n \to \infty,$$

If ρ is compact, then

$$||\rho_n^* - \rho^*||_{\mathcal{L}} \xrightarrow{p} 0, \quad \text{as} \quad n \to \infty,$$

and if ρ is Hilbert-Schmidt, then

$$||\rho_n^* - \rho^*||_{HS} \xrightarrow{p} 0, \quad \text{as} \quad n \to \infty.$$

Discrete-Time Data

In many applied contexts, Z is only observed at discrete-time points. Therefore, in order to use the previous results for continuous-time data, one first approximates the sample paths of Z and then derives appropriate approximations for ρ^*.

Hereafter, the discretization grid size m depends on n, the number of paths of an ARH(1) process, i.e. $m := m(n)$, but for notation simplicity we will omit this dependency and denote it by m.

Interpolation Estimators

Under some strict assumptions on the sample paths of Z (i.e. Holder continuity of order s, $0 < s \leq 1$), Pumo (1992) considers the processes V_i, $i = 0, \ldots, n$, defined by, for $j = 0, 1, \ldots, m$:

$$V_i(t) = Z_i\left(\frac{j}{m}\right) + \frac{t - \frac{j}{m}}{\frac{1}{m}}\left[Z_i\left(\frac{j+1}{m}\right) - Z_i\left(\frac{j}{m}\right)\right], \ \frac{j}{m} \leq t < \frac{j+1}{m},$$

and estimates C and D^* respectively by

$$C_{n,m} = \frac{1}{n+1}\sum_{i=0}^{n} V_i \otimes V_i, \text{ and } D^*_{n,m} = \frac{1}{n}\sum_{i=0}^{n-1} V_{i+1} \otimes V_i.$$

Following the steps used to obtain the projection estimators, he obtains a new class of estimators $\rho^*_{n,m}$. Using the consistency results of the projection estimators he obtains consistency results for the estimators $\rho^*_{n,m}$ when $m^{2s} > n^{1+\eta}$, for some $\eta > 0$.

Smoothing Spline Estimators

To handle the discretization problem, Besse & Cardot (1996) proposed to *simultaneously* estimate the sample paths and project the data using smoothing splines. Assuming that the predictable part of Z belong to a q-dimensional subspace H_q of smooth functions (i.e., the range of ρ is an s-smooth Sobolev space ($W_2^s[0,1]$), $s = 1, 2, \ldots$), they solve the following variational problem

$$\min_{f_k \in H_q; \dim(H_q)=q} \left\{ \frac{1}{m(n+1)} \sum_{k=0}^{n} \sum_{j=1}^{m} \left(Z_k(t_j) - f_k(t_j) \right)^2 + \lambda ||f_k^{(2)}||^2_{L^2[0,1]} \right\},$$

where $\lambda > 0$ is the *regularization parameter*. The operators C and D^* are then estimated respectively by

$$\Gamma_{n.m} = \frac{1}{n+1} \sum_{k=0}^{n} \hat{f}_k \otimes \hat{f}_k$$

and

$$\Delta^*_{n,m} = \frac{1}{n}\sum_{k=0}^{n-1} \hat{f}_{k+1} \otimes \hat{f}_k,$$

giving rise to the estimator

$$\rho^*_{n,m} = \Gamma^{-1}_{n,m}\Delta^*_{n,m},$$

where $\hat{f}_k$ is the solution of the above variational problem.

Example: El Niño-Southern Oscillation

Prediction of a climatological times series describing El Niño-Southern Oscillation (ENSO). ENSO is a natural phenomenon arising from coupled interactions between the atmosphere and the ocean in the tropical Pacific Ocean.

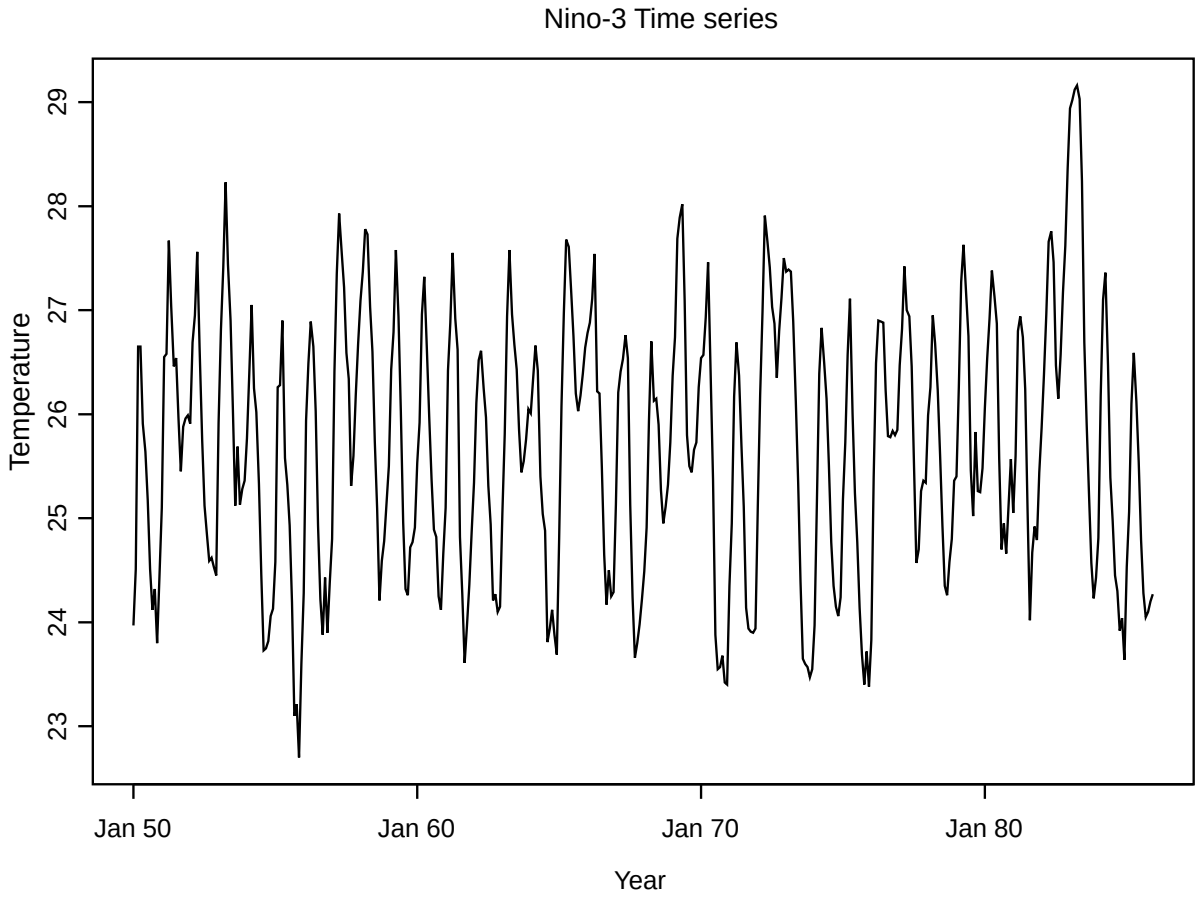
El Niño (EN) is the ocean component of ENSO.

Most of the year-to-year variability in the tropics, as well as a part of the extratropical variability over both Hemispheres, is related to this phenomenon. Southern Oscillation (SO) is the atmospheric counterpart of ENSO.

An useful index of El Niño variability is provided by the sea surface temperatures averaged over the Niño-3 domain. Monthly mean values have been obtained from January 1950 to December 1996 from gridded analyses made at the U.S. National Centers for Environmental Prediction (see Smith, Reynolds, Livezey & Stokes, 1996).

The time series of this EN index is depicted in the Figure that follows.

The monthly mean Niño-3 surface temperature index

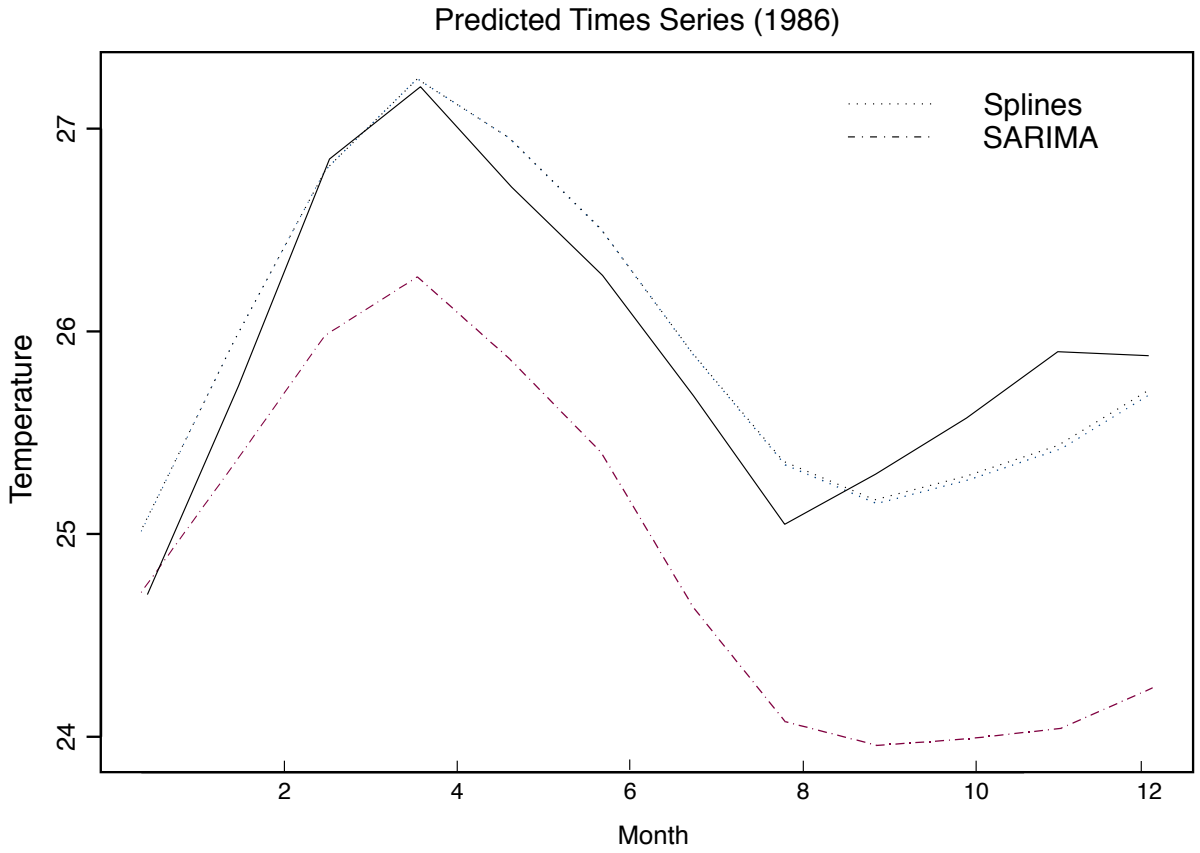


The monthly mean Niño-3 surface temperature index in (deg C) which provides a contracted description of ENSO.

The first 36 years, from 1950 to 1985, were considered as a learning sample. The discretization size grid for each sample path, is equal to $m = 12$.

We have used the smoothing spline interpolation estimator with smoothing parameters chosen optimally by a cross-validation criterion ($q = 2$).

To complete the comparison, a suitable ARIMA model, including 12 month seasonality, has also been adjusted to the times series from January 1950 to December 1985. The most parsimonious SARIMA model, validated through a portmanteau test for serial correlation of the fitted residuals, was driven by the parameters $(0, 1, 1) \times (1, 0, 1)_{12}$.



The raw Niño-3 surface temperature during 1986 and its prediction.

THANK YOU!