



POLITECNICO
MILANO 1863



Statistical methods of data science

An introduction to Functional Data Analysis

Alessandra Menafoglio^{1*}

¹MOX, Department of Mathematics, Politecnico di Milano

*alessandra.menafoglio@polimi.it

3. FDA & Dimensionality reduction in Hilbert spaces

Agenda

3. FDA & Dimensionality reduction in Hilbert spaces

- 3.1. Functional Principal Components in Hilbert spaces
- 3.2. Examples in L^2
- 3.3. Examples in Bayes spaces

Recall: Principal Component Analysis

Problem: Principal Component Analysis

Given a dataset of n zero-mean multivariate observations in \mathbb{R}^p , X_1, \dots, X_N , find the directions of maximum variability of the dataset, i.e., those maximizing

$$\text{Var}(\mathbf{a}' X) \text{ subject to } \mathbf{a}' \mathbf{a} = 1$$

Maximum variance

$$\mathbf{a}'_k \mathbf{a} = 0, \quad j < k$$

Orthonormality

- We can re-write the problem as maximizing

$$\frac{1}{N} \sum_{i=1}^N (\mathbf{a}' X_i)^2 \text{ subject to } \mathbf{a}' \mathbf{a} = 1$$
$$\mathbf{a}'_k \mathbf{a} = 0, \quad j < k$$

or, equivalently

$$\frac{1}{N} \sum_{i=1}^N \langle \mathbf{a}, X_i \rangle^2 \text{ subject to } \|\mathbf{a}\| = 1$$
$$\langle \mathbf{a}_k, \mathbf{a} \rangle = 0, \quad j < k$$

Maximum sample variance

Orthonormality in \mathbb{R}^p

Recall: Principal Component Analysis

Problem: Principal Component Analysis

Given a dataset of n zero-mean multivariate observations in \mathbb{R}^p , X_1, \dots, X_N , find the directions of maximum variability of the dataset, i.e., those maximizing

$$\text{Var}(\mathbf{a}' X) \text{ subject to } \mathbf{a}' \mathbf{a} = 1$$

$$\mathbf{a}'_k \mathbf{a} = 0, \quad j < k$$

Maximum variance

Orthonormality

Solution

Call S the sample covariance matrix of X_1, \dots, X_N . Then, the principal components e_1, \dots, e_k are found as the eigenvectors of the matrix S , i.e., they solve the eigen-equations

$$S e_k = \lambda_k e_k$$

The eigenvalue λ_k associated with the eigenvector e_k represents the variability along the direction e_k .

Note. We call *score* x_{ik} the projection of the observation X_i along the direction e_k , i.e.,

$$x_{ik} = \langle X_i, e_k \rangle = X'_i e_k$$

Recall: Principal Component Analysis

Problem: Principal Component Analysis

Given a dataset of n zero-mean multivariate observations in \mathbb{R}^p , X_1, \dots, X_N , find the directions of maximum variability of the dataset, i.e., those maximizing

$$\text{Var}(\mathbf{a}' X) \text{ subject to } \mathbf{a}' \mathbf{a} = 1$$

$$\mathbf{a}'_k \mathbf{a} = 0, \quad j < k$$

Maximum variance

Orthonormality

Solution

Call S the sample covariance matrix of X_1, \dots, X_N . Then, the principal components e_1, \dots, e_k are found as the eigenvectors of the matrix S , i.e., they solve the eigen-equations

$$Se_k = \lambda_k e_k$$

The eigenvalue λ_k associated with the eigenvector e_k represents the variability along the direction e_k .

Note. We call *score* x_{ik} the projection of the observation X_i along the direction e_k , i.e.,

$$x_{ik} = \langle X_i, e_k \rangle = X'_i e_k$$

Can we do the same in *any* Hilbert space,
using its inner product?

3. FDA & Dimensionality reduction in Hilbert spaces

- 3.1. Functional Principal Components in Hilbert spaces
- 3.2. Examples in L^2
- 3.3. Examples in Bayes spaces

Functional Principal Component Analysis

Problem statement

Problem: Functional Principal Component Analysis

Given a dataset of n zero-mean functional observations in H , X_1, \dots, X_N , find the directions of maximum variability (in H) of the dataset, i.e., those maximizing

$$\frac{1}{N} \sum_{i=1}^N \langle \xi, X_i \rangle_H^2 \text{ subject to } \|\xi\|_H = 1$$
$$\langle \xi_k, \xi_j \rangle_H = 0, \quad j < k$$

Maximum sample variance

Orthonormality in H

- We look for an orthonormal basis in H maximizing the *variability* of the corresponding projections
- Indeed, $\langle \xi, X_i \rangle_H$ is the projection of X_i «along the direction ξ » (i.e., a «direction» in H).

Note that $\langle \xi, X_i \rangle_H$ is a scalar, hence $\frac{1}{N} \sum_{i=1}^N \langle \xi, X_i \rangle_H^2$ is a sample variance in the usual sense.

Note. If the data are not zero-mean, they can be centered by subtracting the (sample) mean.

Functional Principal Component Analysis

Problem statement

Problem: Functional Principal Component Analysis

Given a dataset of n zero-mean functional observations in H , X_1, \dots, X_N , find the directions of maximum variability (in H) of the dataset, i.e., those maximizing

$$\frac{1}{N} \sum_{i=1}^N \langle \xi, X_i \rangle_H^2 \text{ subject to } \begin{aligned} \|\xi\|_H &= 1 \\ \langle \xi_k, \xi_j \rangle_H &= 0, \quad j < k \end{aligned}$$

Maximum sample variance

Orthonormality in H

- Similar as for multivariate principal component analysis, functional principal components are related with the eigen-decomposition of the functional counterpart of the (sample) covariance matrix
- Recall that the **sample covariance operator** is defined as

$$Sx = \frac{1}{N} \sum_{i=1}^N \langle X_i, x \rangle X_i, \quad x \in H$$

In L^2 it is equivalently defined as

$$[Sx](t) = \int_T \widehat{c}(s, t)x(s)d(s), \quad x \in L^2 \text{ with } \widehat{c}(s, t) = \frac{1}{N} \sum_{i=1}^N X(s)X(t)$$

Note. We could have used the version divided by $(N-1)$, if preferred.

Functional Principal Component Analysis

Problem statement

Problem: Functional Principal Component Analysis

Given a dataset of n zero-mean functional observations in H , X_1, \dots, X_N , find the directions of maximum variability (in H) of the dataset, i.e., those maximizing

$$\frac{1}{N} \sum_{i=1}^N \langle \xi, X_i \rangle_H^2 \text{ subject to } \|\xi\|_H = 1$$
$$\langle \xi_k, \xi_j \rangle_H = 0, \quad j < k$$

Maximum sample variance

Orthonormality in H

Solution

Call S the sample covariance operator of X_1, \dots, X_N . Then, the functional principal components ξ_1, \dots, ξ_{N-1} are found as the eigenfunctions of the operator S , i.e., they solve the eigen-equations

$$S\xi_k = \lambda_k \xi_k$$

The eigenvalue λ_k associated with the eigenfunction ξ_k represents the variability along the direction ξ_k .

We call *functional score* x_{ik} the projection of the observation X_i along the direction ξ_k , i.e.,

$$x_{ik} = \langle X_i, e_k \rangle$$

Functional Principal Component Analysis

Dimensionality reduction and Interpretation of the results

- To reduce the dimensionality of the dataset one can proceed as in the multivariate setting, e.g., by looking for an elbow in the cumulative percentage of total variance explained by the first p functional principal components.

$$CPV(p) = \frac{\sum_{k=1}^p \hat{\lambda}_k}{\sum_{k=1}^N \hat{\lambda}_k}.$$

- Other useful plots are the boxplots of the scores along the first p directions, to investigate the possible presence (and influence) of outliers on the results
- Interpretation of the loadings can be performed by:
 - Plotting the loadings themselves (only for expert users)
 - Plotting the mean +/- the eigenfunctions multiplied by a proper constant, e.g., the std. along the component, which corresponds to the sqrt of the eigenvalue:

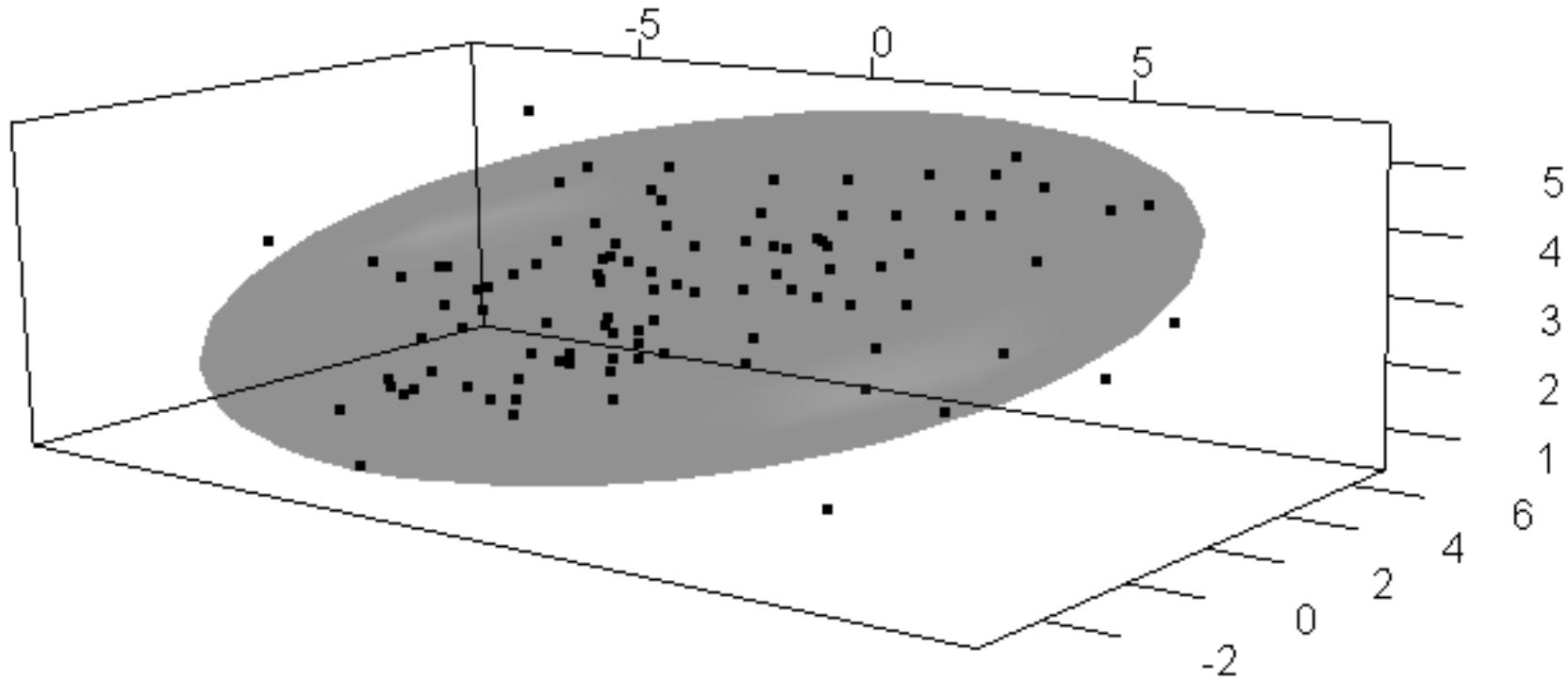
$$\bar{X} \pm \sqrt{\lambda_k} \xi_k$$

- Plotting the projection of the dataset along each component or along the first p components

$$\frac{\bar{X} + x_{ik} \xi_k}{\bar{X} + \sum_{k=1}^p x_{ik} \xi_k}$$

Functional Principal Component Analysis

FPCA as space of best approximation

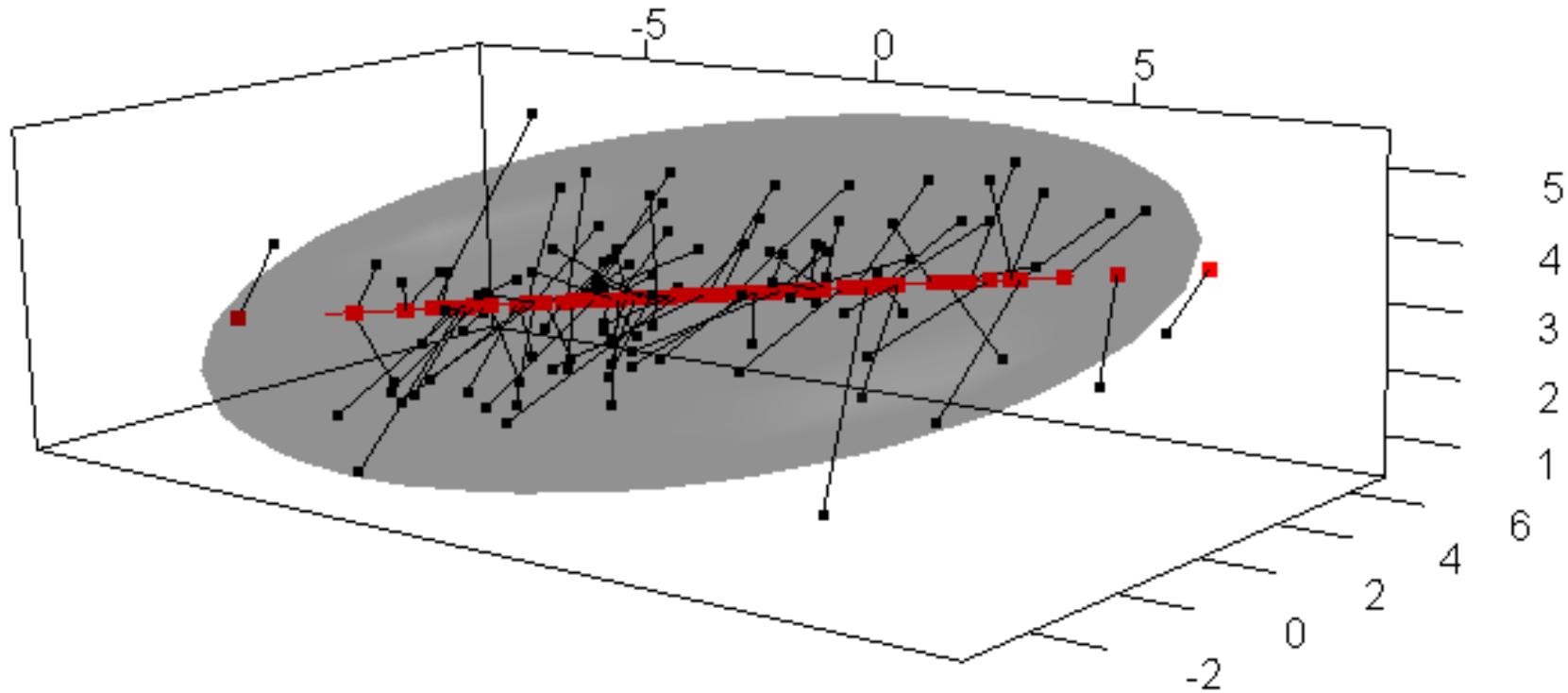


Problem: find the space of dimension k that best approximate the data in the mean square sense

If $k=0$: sample mean

Functional Principal Component Analysis

FPCA as space of best approximation

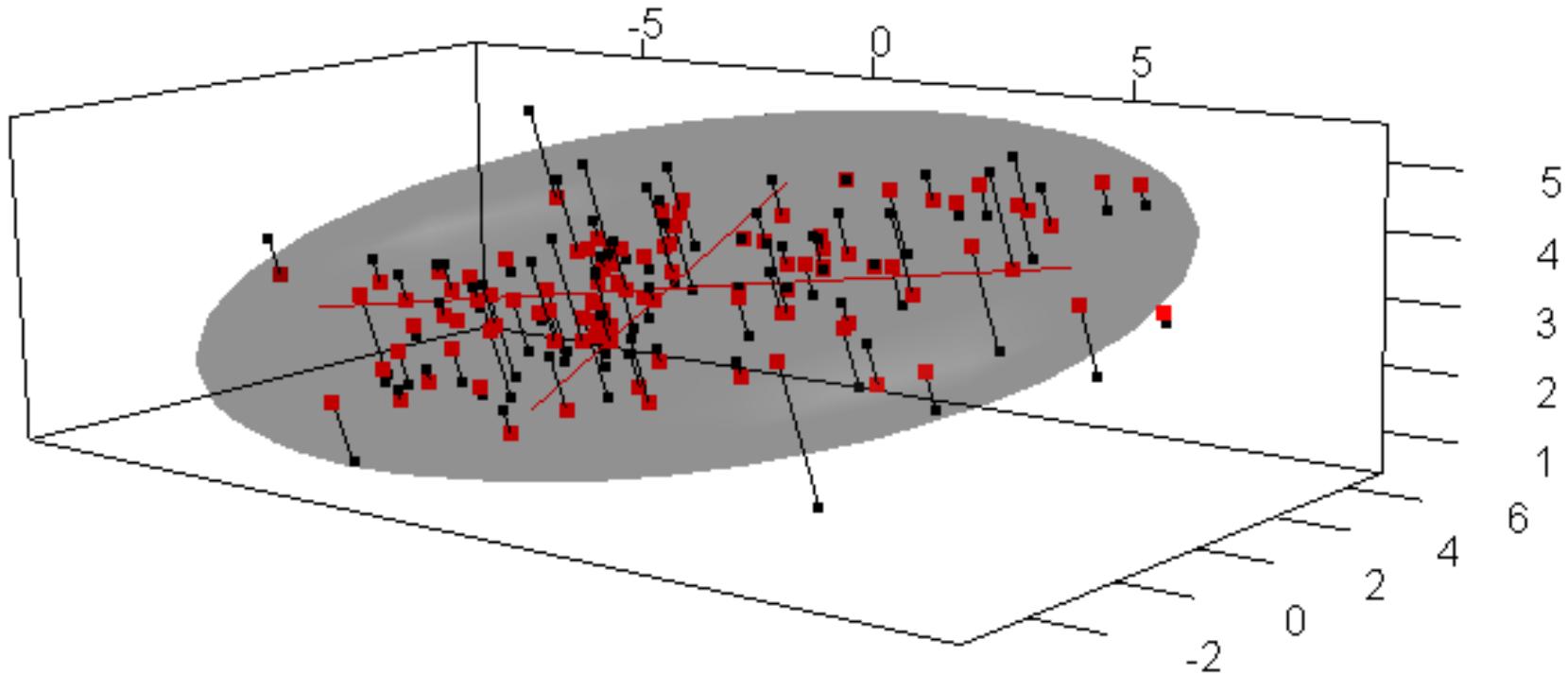


Problem: find the space of dimension k that best approximate the data in the mean square sense

If $k=1$: linear space generated by the first FPC

Functional Principal Component Analysis

FPCA as space of best approximation



Problem: find the space of dimension k that best approximate the data in the mean square sense

If $k=2$: linear space generated by the first two FPCs

...

Agenda

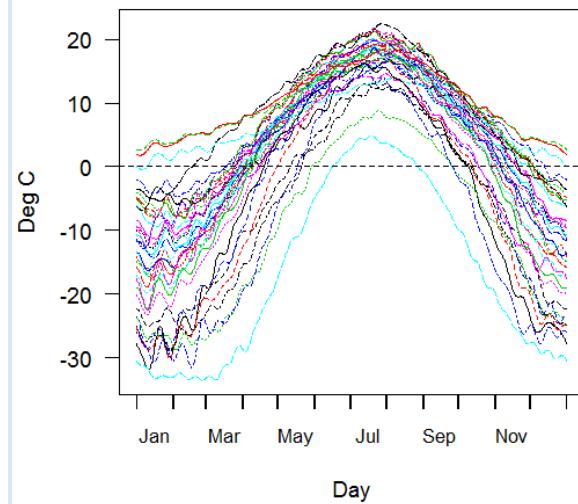
3. FDA & Dimensionality reduction in Hilbert spaces

- 3.1. Functional Principal Components in Hilbert spaces
- 3.2. Examples in L^2
- 3.3. Examples in Bayes spaces

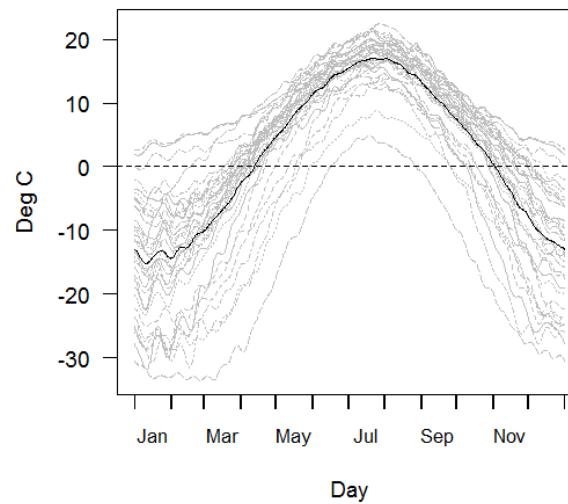
3.2 Examples in L^2

Dataset of Canadian temperatures

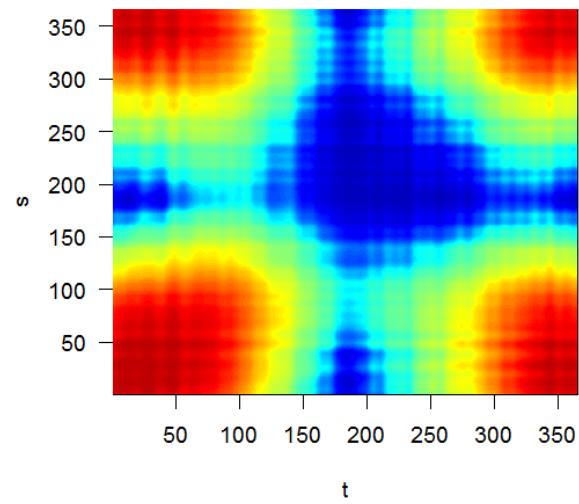
Example: Dataset of Temperatures in Canada (35 observations)



Functional dataset



Sample mean



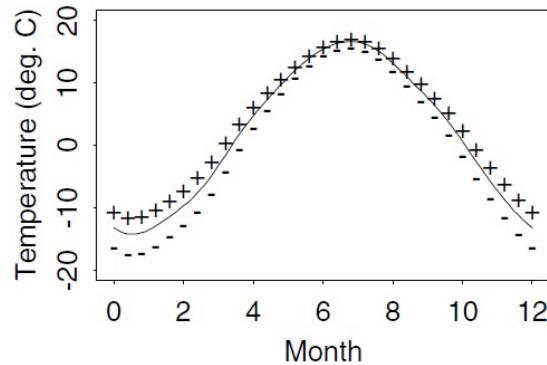
Sample covariance kernel

3.2 Examples in L^2

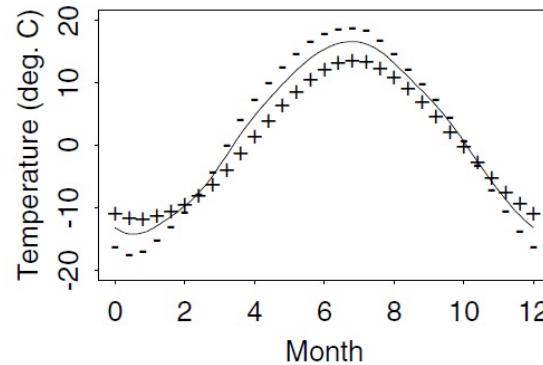
Dataset of Canadian temperatures

Example: Dataset of Temperatures in Canada (35 observations)

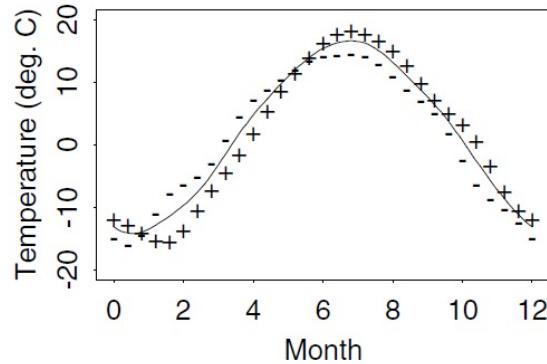
PC 1 (89.3%)



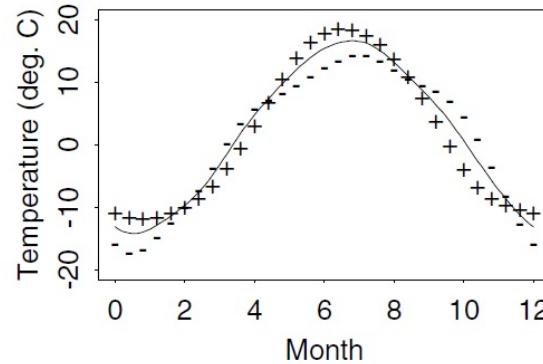
PC 2 (8.3%)



PC 3 (1.6%)

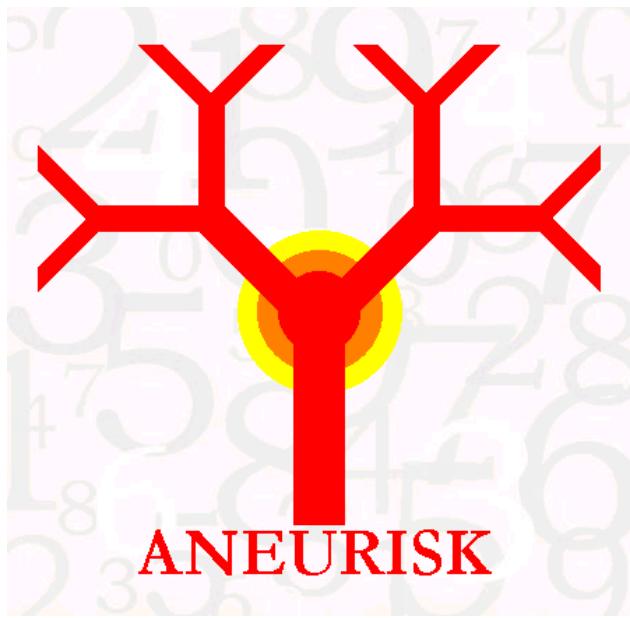


PC 4 (0.5%)



3.2 Examples in L²

Aneurisk project



SIEMENS



Statistics
Computer fluid dynamics



Structure Mechanics



Neuroradiology



Image Reconstruction

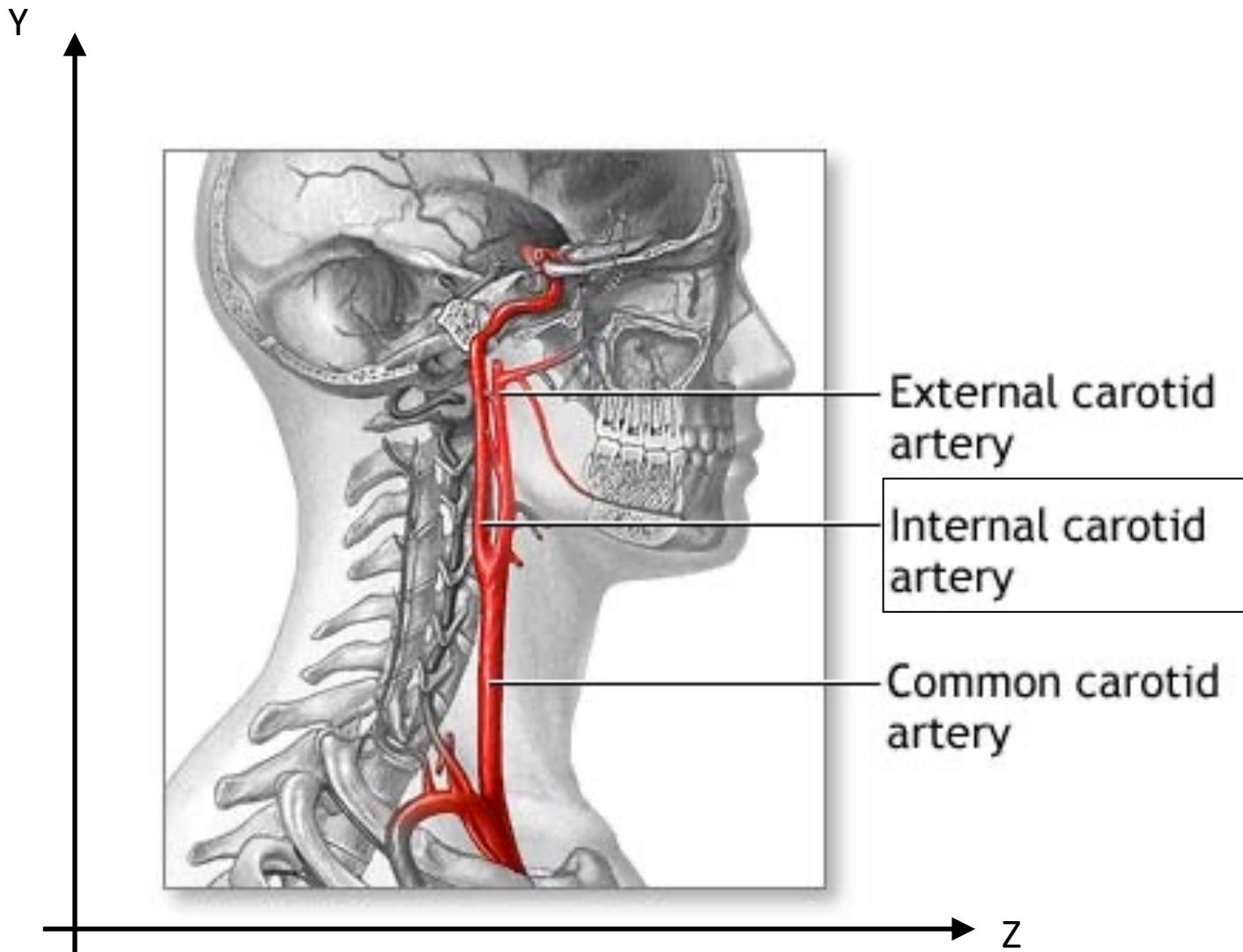


Neurosurgery



3.2 Examples in L²

Aneurisk project

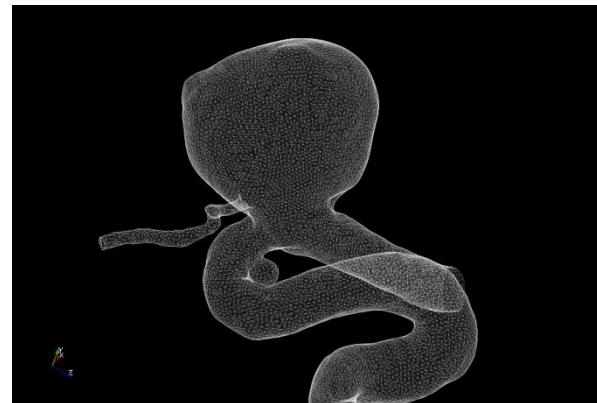
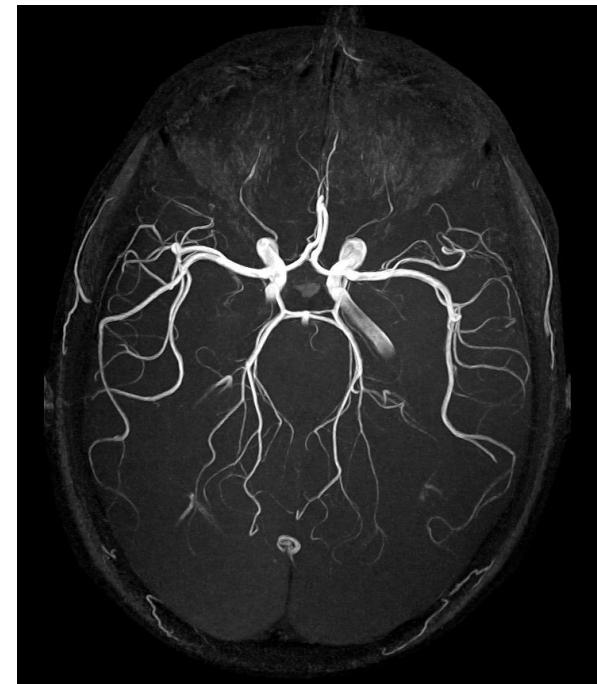


3.2 Examples in L²

Aneurisk project

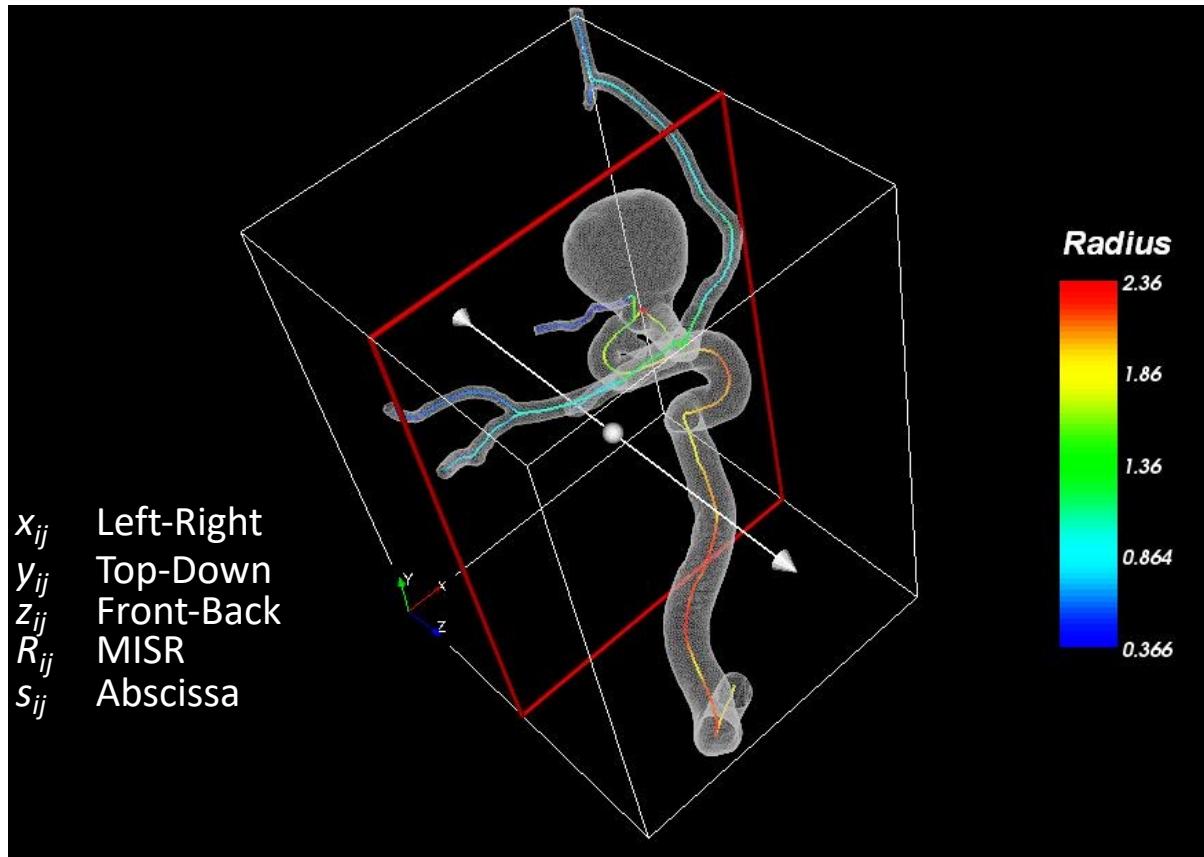
Cerebral aneurysms: malformations of cerebral arteries, in particular of arteries belonging to or connected to the Circle of Willis.

- Incidence rate of cerebral aneurysms:
1/20 people
- Incidence rate of ruptured cerebral aneurysms per year:
1/10000 people per year
- Mortality due to a ruptured aneurysm:
> 50%: Out of 9 patients with a ruptured aneurysm:
 - 3 are expected to die before arriving at the hospital
 - 2 to die after having arrived at the hospital
 - 2 to survive with permanent cerebral damages
 - 2 to survive without permanent cerebral damages



3.2 Examples in L²

Aneurisk project – Data collection



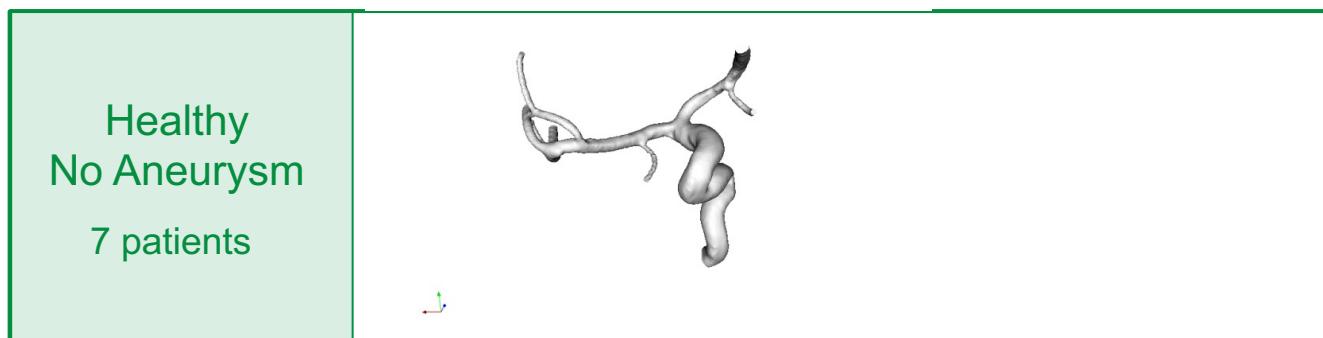
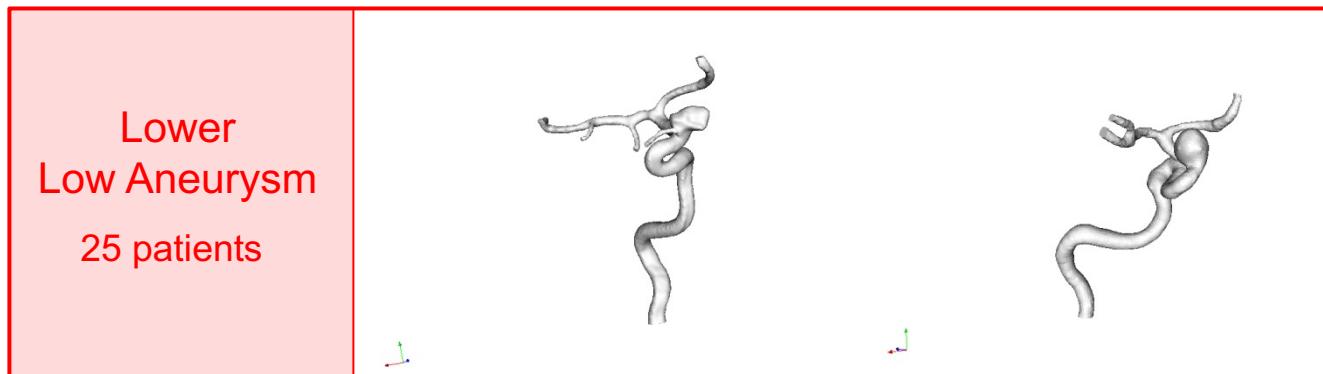
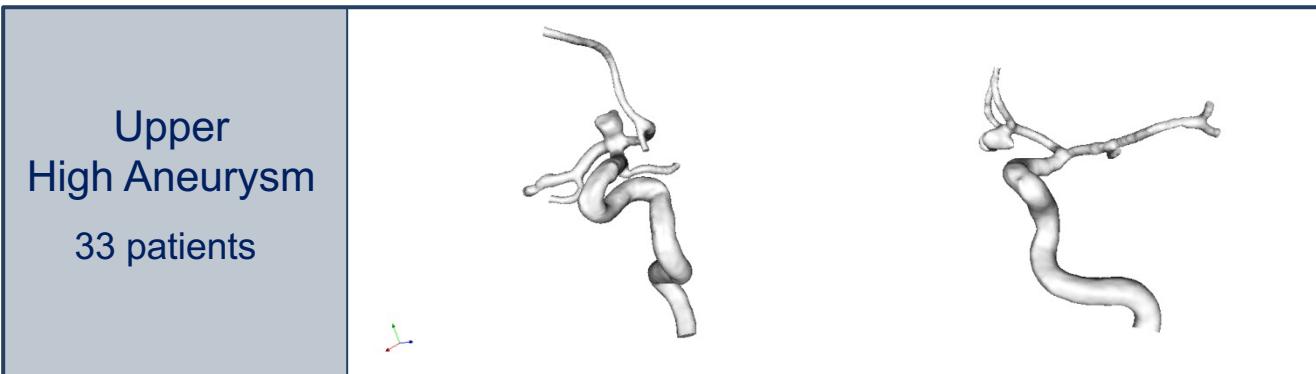
Observational Study conducted at Ospedale Ca' Granda Niguarda – Milano relative to 65 patients hospitalized from September 2002 to October 2005.

Courtesy of S. Vantini

3.2 Examples in L²

Aneurisk project – Patient classification

Courtesy of S. Vantini

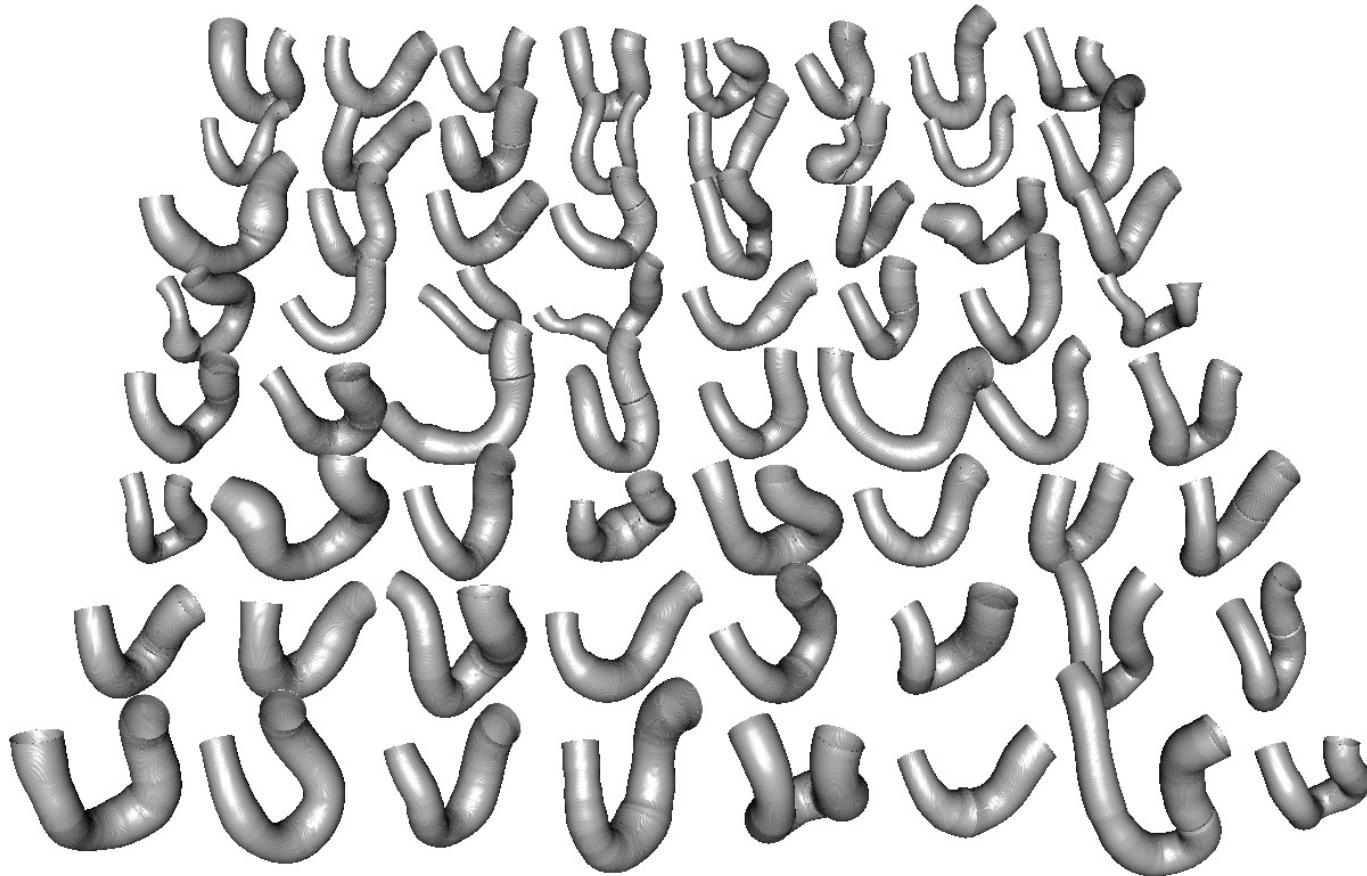


3.2 Examples in L²

Aneurisk project – Dataset

Courtesy of S. Vantini

The sample of 65 ICA: each patient is represented by the centerline and radius of ICA



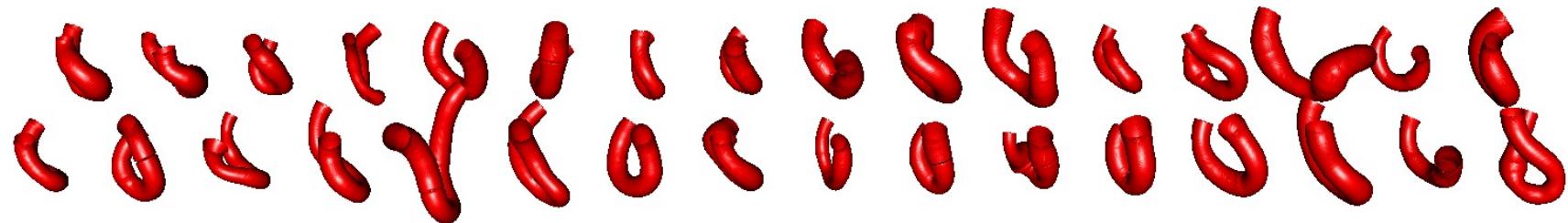
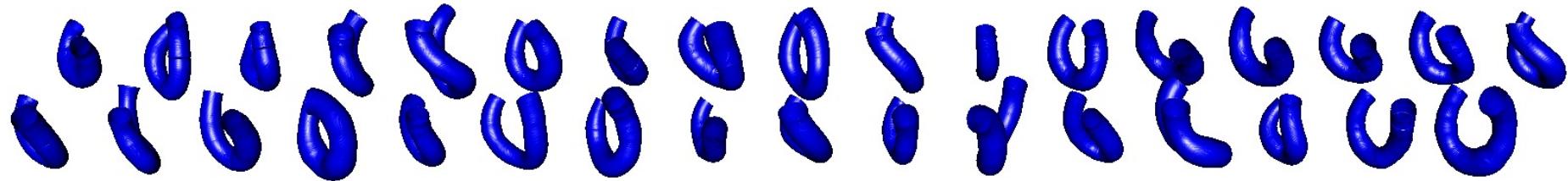
3.2 Examples in L²

Aneurisk project - Dataset

Courtesy of S. Vantini

Aneurysm at or after ICA bifurcation

Upper Group: 33



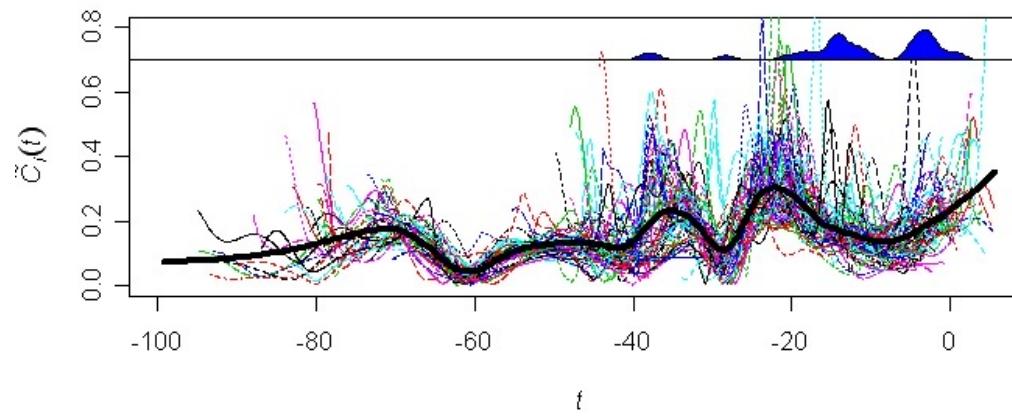
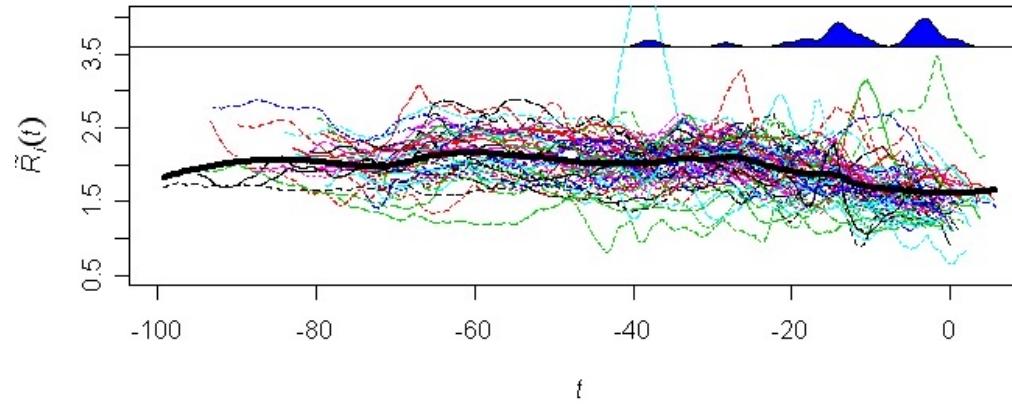
Lower Group: 32

Aneurysm before ICA bifurcation or no aneurysm

3.2 Examples in L^2

Aneurisk dataset – Radius and curvature

Courtesy of S. Vantini

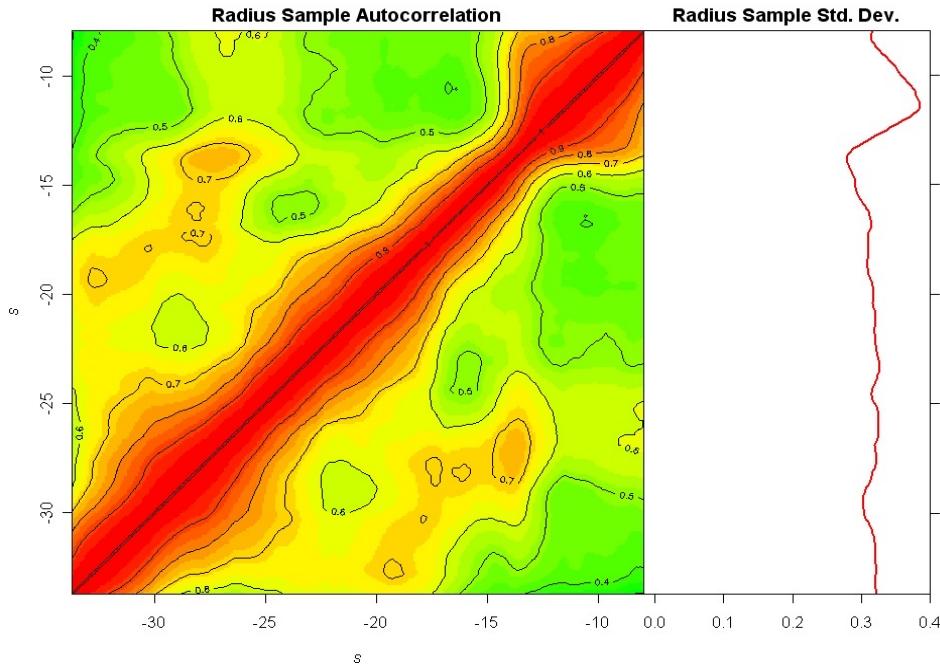


3.2 Examples in L^2

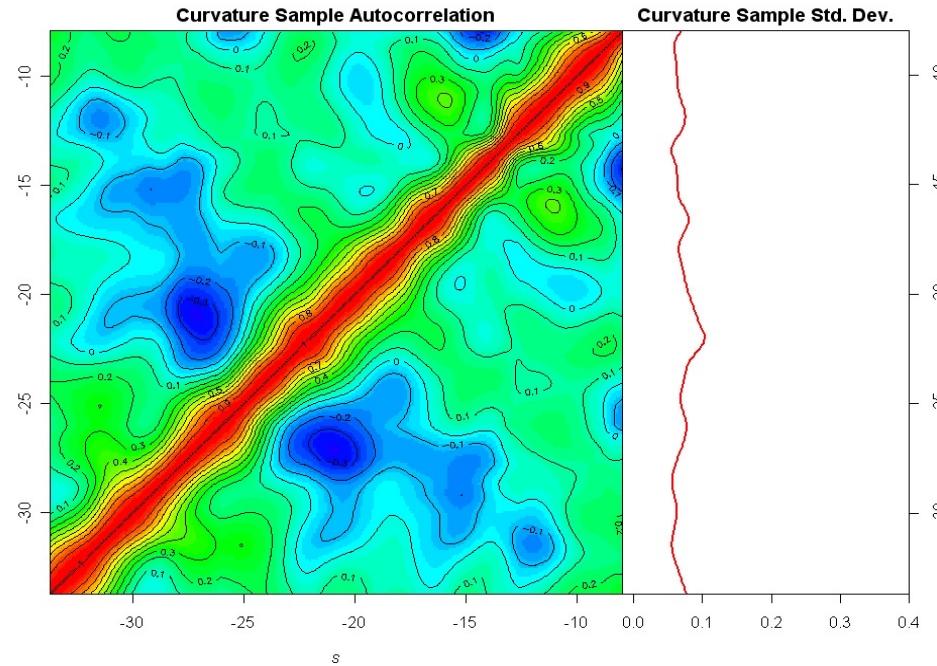
Aneurisk dataset – Radius and curvature

Courtesy of S. Vantini

**Sample Autocorrelation Function
and Std. Dev. for Radius Profiles**



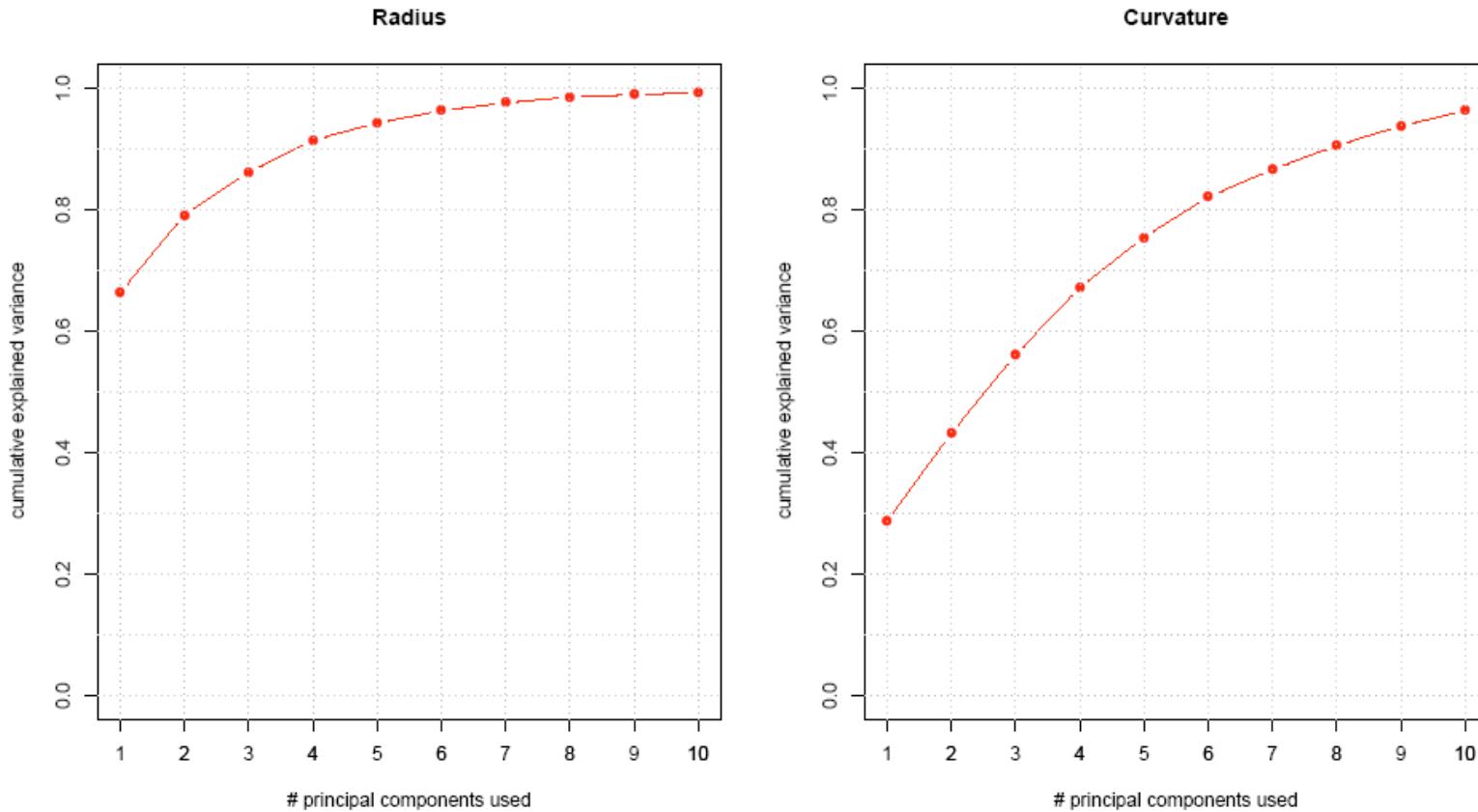
**Sample Autocorrelation Function
and Std. Dev. for Curvature Profiles**



3.2 Examples in L^2

Aneurisk dataset – FPCA of radius and curvature

Courtesy of S. Vantini

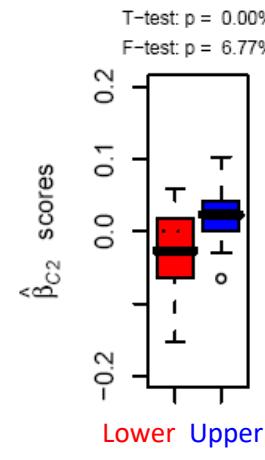
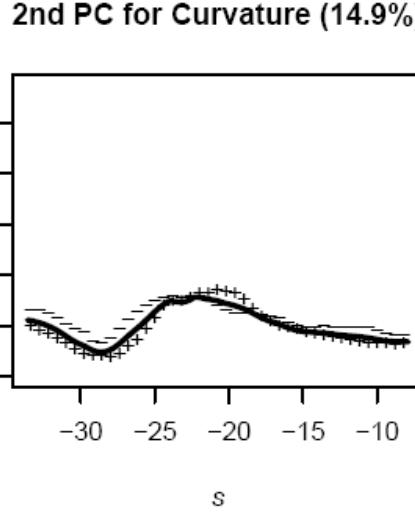
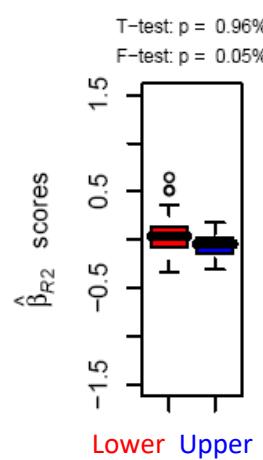
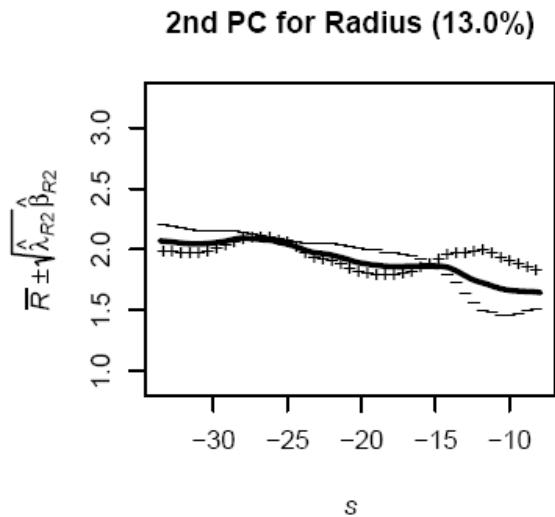
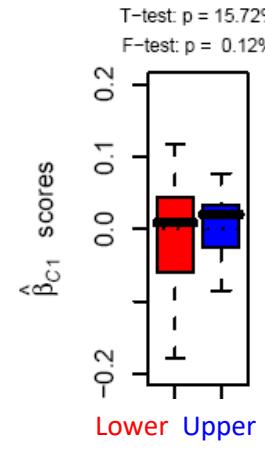
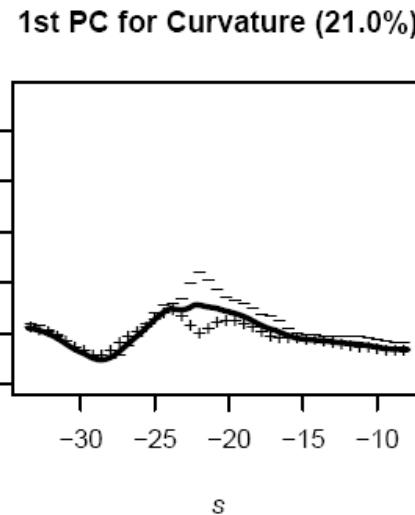
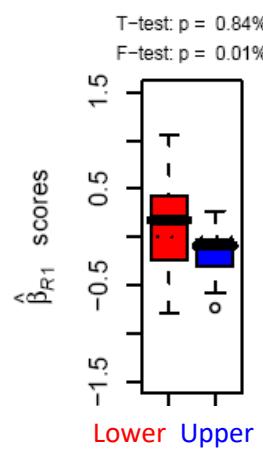
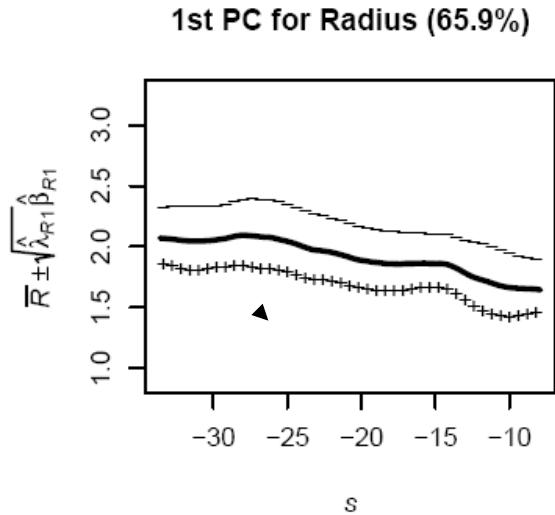


Courtesy of S. Vantini

3.2 Examples in L^2

Aneurisk dataset – FPCA of radius and curvature

Courtesy of S. Vantini

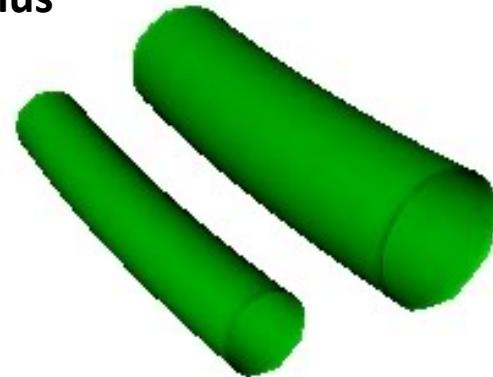


3.2 Examples in L^2

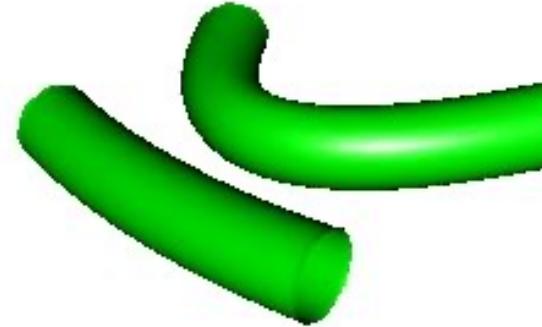
Aneurisk dataset – FPCA of radius and curvature

Courtesy of S. Vantini

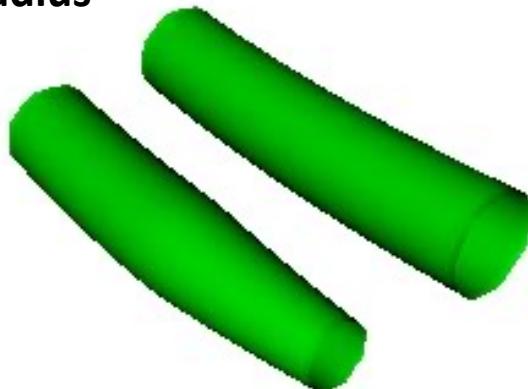
1st PC
Radius



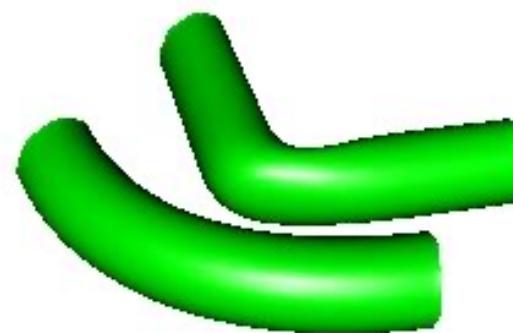
1st PC
Curvature



2nd PC
Radius



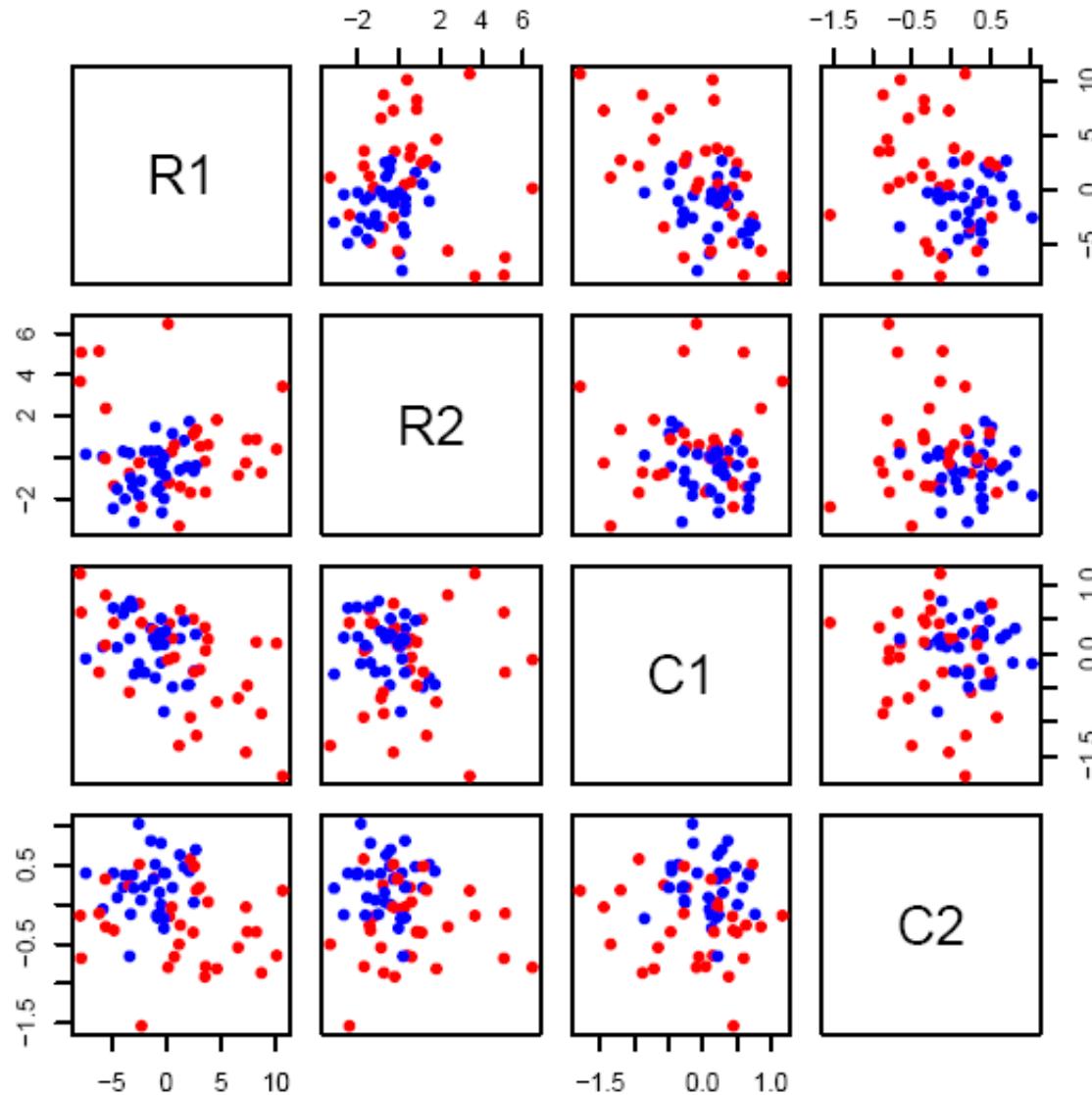
2nd PC
Curvature



3.2 Examples in L^2

Aneurisk dataset – FPCA of radius and curvature

Courtesy of S. Vantini



Agenda

3. FDA & Dimensionality reduction in Hilbert spaces

- 3.1. Functional Principal Components in Hilbert spaces
- 3.2. Examples in L^2
- 3.3. Examples in Bayes spaces

3.3. Examples in Bayes spaces

Simplicial Functional Principal Components

- We have given the formulation of the FPCA in a very general framework. Let's focus on the very peculiar case of the Bayes space B^2
- Data in B^2 are (equivalence classes of) functions with square-integrable logarithm. Among these, we find probability density functions
- *Recall*

B^2 : space of density functions on a close interval I , with \log in L^2

- Equivalence relation: f, g are equivalent if they are proportional (*scale invariance*)
- Sum (perturbation): $(f \oplus g)(t) = \frac{f(t)g(t)}{\int_I f(s)g(s) ds}$,
- Product by a constant (powering): $(\alpha \odot f)(t) = \frac{f(t)^\alpha}{\int_I f(s)^\alpha ds}, \quad t \in I.$
- Inner product: $\langle f, g \rangle_{\mathcal{B}} = \frac{1}{2\eta} \int_I \int_I \ln \frac{f(t)}{f(s)} \ln \frac{g(t)}{g(s)} dt ds$
- Norm: $\|f\|_{\mathcal{B}} = \left[\frac{1}{2\eta} \int_I \int_I \ln^2 \frac{f(t)}{f(s)} dt ds \right]^{1/2}$

3.3. Examples in Bayes spaces

Why choosing a Bayes space to embed continuous PDFs?

- Let's focus on the main features of **continuous PDFs**. Amongst the properties of these data, we mention:
 - Constraints**: positive functions integrating to 1
 - Scale invariance**: the constant 1 describing the integral constraint is rather a convention than a relevant information
 - Relative scale**: increase in probability is relative to the probability itself, e.g., 0.05 vs 0.1 (2 multiple) is different from 0.5 vs 0.55 (1.1 multiple)
- Continuous** probability density functions can be modeled as **functional compositions**: positive functional data that convey only relative information; log-ratios between “infinitesimal parts” provide the only relevant information

$$x : \mathcal{T} \rightarrow \mathbb{R} \quad x(t) > 0, \int_{\mathcal{T}} x(t) dt = const$$

- Multivariate counterpart**: discrete probability density functions, modeled as **multivariate compositions**

3.3. Examples in Bayes spaces

Bayes space geometry and compositional data

Let's give a closer look to the Bayes space geometry

B²: space of density functions on a close interval I, with log in L²

- Equivalence relation: f, g are equivalent if they are proportional

Scale
invariance

- Sum (perturbation): $(f \oplus g)(t) = \frac{f(t)g(t)}{\int_I f(s)g(s) ds}$,

Bayes
theorem

- Product by a constant (powering): $(\alpha \odot f)(t) = \frac{f(t)^\alpha}{\int_I f(s)^\alpha ds}, \quad t \in I.$

- Inner product: $\langle f, g \rangle_{\mathcal{B}} = \frac{1}{2\eta} \int_I \int_I \ln \frac{f(t)}{f(s)} \ln \frac{g(t)}{g(s)} dt ds$

Log-ratios
between
infinitesimal
parts

- Norm: $\|f\|_{\mathcal{B}} = \left[\frac{1}{2\eta} \int_I \int_I \ln^2 \frac{f(t)}{f(s)} dt ds \right]^{1/2}$

This is precisely the Aitchison geometry for multivariate compositions, but in infinite-dimension!

3.3. Examples in Bayes spaces

Bayes space geometry and compositional data

- We have said that the **centered log-ratio transformation** is an isometric isomorphism between B^2 and L^2

$$\text{clr}(f)(t) = f_c(t) = \ln f(t) - \frac{1}{\eta} \int_I \ln f(s) ds.$$

- In particular, we have

$$\text{clr}(f \oplus g)(t) = f_c(t) + g_c(t), \quad \text{clr}(\alpha \odot f)(t) = \alpha \cdot f_c(t),$$

$$\langle f, g \rangle_{\mathcal{B}} = \langle f_c, g_c \rangle_2 = \int_I f_c(t) g_c(t) dt.$$

- We can use the clr transformation to perform computation in the Bayes space using convenient formulations in L^2 .
 - *Example:* the sample mean function in L^2 can be computed point-wise, in B^2 it has to be computed as Fréchet sample mean. We can thus map the data in L^2 , make computations on clr-transformed data and map back the results in B^2 .

$$\bar{X} = \text{clr}^{-1} \left(\frac{1}{N} \sum_{i=1}^N \text{clr}(X_i) \right)$$

Note: similarly, we can compute the covariance function

3.3. Examples in Bayes spaces

Bayes space geometry and clr-transformation

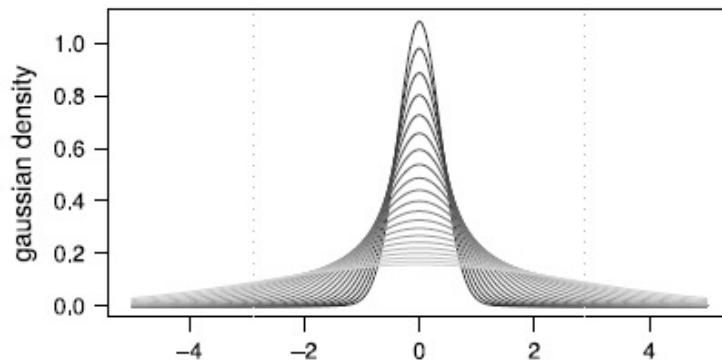
Example 1: Dataset of Gaussian distributions

Consider a Gaussian distribution with zero-mean and variance $\sigma_i = \exp(-1 + (i - 1)/10)$ for $i = 1, \dots, 21$

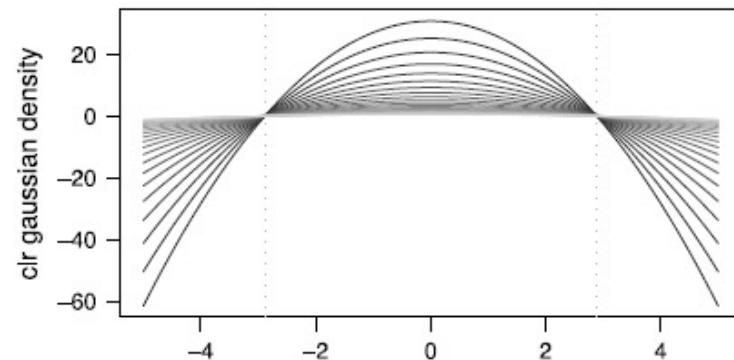
$$f(t; \sigma_i) =_{\mathcal{B}} \exp \left\{ -\frac{t^2}{2\sigma_i^2} \right\}, \quad t \in I.$$

It's clr can be explicitly computed as

$$f_c(t; \sigma_i) = -\frac{t^2}{2\sigma_i^2} + \frac{25}{6\sigma_i^2}, \quad t \in I.$$



(a) Density functions.



(b) Clr-transforms.

3.3. Examples in Bayes spaces

Bayes space geometry and clr-transformation

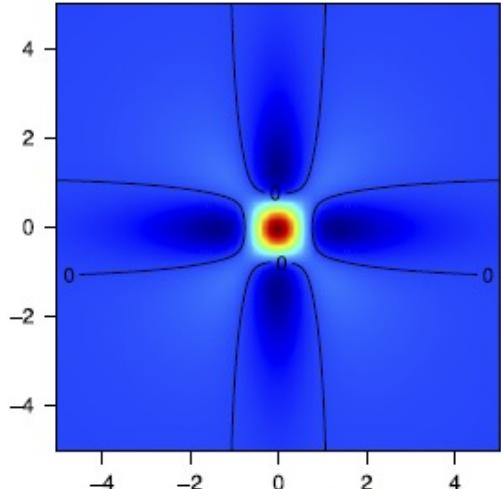
Example 1: Dataset of Gaussian distributions

Consider a Gaussian distribution with zero-mean and variance $\sigma_i = \exp(-1 + (i - 1)/10)$ for $i = 1, \dots, 21$

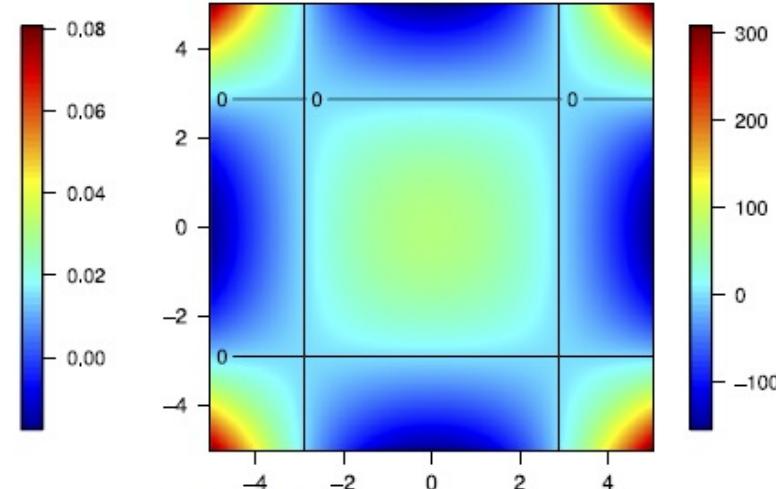
$$f(t; \sigma_i) =_{\mathcal{B}} \exp \left\{ -\frac{t^2}{2\sigma_i^2} \right\}, \quad t \in I.$$

It's clr can be explicitly computed as

$$f_c(t; \sigma_i) = -\frac{t^2}{2\sigma_i^2} + \frac{25}{6\sigma_i^2}, \quad t \in I.$$



(c) Covariance function in L^2 .



(d) Covariance function in \mathcal{B}^2 .

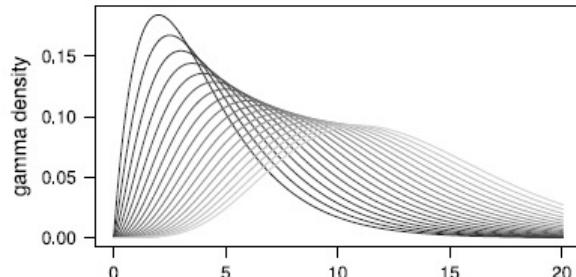
3.3. Examples in Bayes spaces

Bayes space geometry and clr-transformation

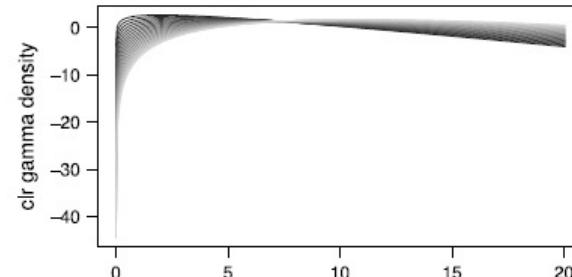
Example 2: Dataset of Gamma distributions

Consider a Gamma distribution with scale parameter $\theta = 2$ and shape parameter $\kappa_i = 2 + (i - 1)/4$ for $i = 0, \dots, 19$, $I = [e^{-7}, e^3]$:

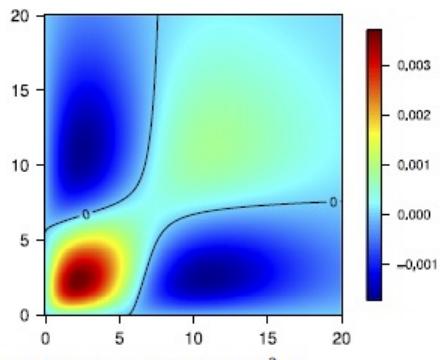
$$f(t; \kappa_i) = {}_{\mathcal{B}} t^{\kappa_i-1} \exp \left\{ -\frac{t}{2} \right\}, \quad t \in I.$$



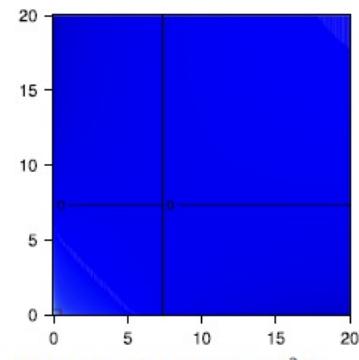
(a) Density functions.



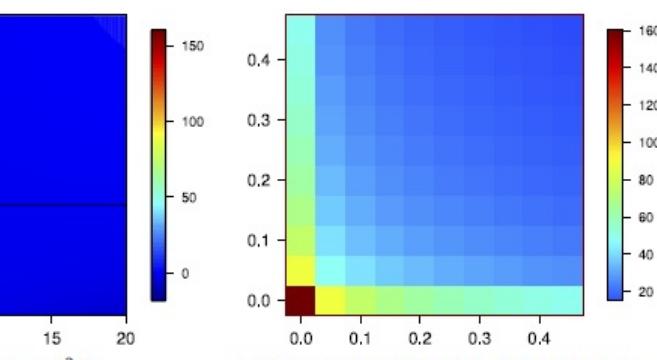
(b) Clr-transforms.



(c) Covariance function in $L^2(I)$.



(d) Covariance function in $\mathcal{B}^2(I)$.



(e) A zoom into the top-left corner of the covariance function in $\mathcal{B}^2(I)$.

3.3. Examples in Bayes spaces

Bayes space geometry and exponential families

- More in general, we can define the extended exponential family on I as the collection of distributions having density expressible as

$$f(t, \alpha) =_{\mathcal{B}} g(t) \cdot \exp \left\{ \sum_{j=1}^k \vartheta_j(\alpha) T_j(t) \right\}, \quad t \in I,$$

- We can re-write the previous expression in the geometry of the Bayes space

$$f(t, \alpha) =_{\mathcal{B}} g(t) \oplus \bigoplus_{j=1}^k [\vartheta_j(\alpha) \odot \exp\{T_j(t)\}], \quad t \in I,$$

Note: in the Bayes space B^2 , the **exponential family is an affine space**. In particular, sum of PDFs in the exponential family is still in the exponential family.

From the **Bayesian perspective**: this is coherent with the fact that *distributions in the exponential family are conjugate with themselves*, and perturbation in the Bayes space is actually a Bayesian update of information through the Bayes theorem.

3.3. Examples in Bayes spaces

Simplicial Functional Principal Components

Problem: Functional Principal Component Analysis

Given a dataset of n zero-mean functional observations in H , X_1, \dots, X_N , find the directions of maximum variability (in H) of the dataset, i.e., those maximizing

$$\frac{1}{N} \sum_{i=1}^N \langle \xi, X_i \rangle_{\mathcal{B}}^2 \text{ subject to } \|\xi\|_{\mathcal{B}} = 1$$
$$\langle \xi_k, \xi_j \rangle_{\mathcal{B}} = 0, \quad j < k$$

Maximum sample variance

Orthonormality in \mathcal{B}^2

Solution

Call S the sample covariance operator of X_1, \dots, X_N . Then, the functional principal components ξ_1, \dots, ξ_{N-1} are found as the eigenfunctions of the operator S , i.e., they solve the eigen-equations

$$S\xi_k = \lambda_k \xi_k$$

The eigenvalue λ_k associated with the eigenfunction ξ_k represents the variability along the direction ξ_k .

We call *functional score* x_{ik} the projection of the observation X_i along the direction ξ_k , i.e.,

$$x_{ik} = \langle X_i, e_k \rangle$$

3.3. Examples in Bayes spaces

Clr and Simplicial Functional Principal Components

One can prove that the problem can be equivalently re-stated in terms of clr transformation in L^2 , with an additional zero-integral constraint

Equivalent formulation of SFPCA

Problem: Given a dataset of n zero-mean functional observations in B^2 , X_1, \dots, X_N , find the directions of maximum variability (in B^2) of the dataset, i.e. ξ_1, \dots, ξ_{N-1} , with $\xi_i = \text{clr}^{-1}(\nu_i)$ and ν_i maximizing

$$\frac{1}{N} \sum_{i=1}^N \langle \text{clr}(X_i), \nu \rangle_2^2 \text{ subject to } \|\nu\|_2 = 1; \quad \langle \nu, \nu_k \rangle_2 = 0, \quad k < j; \quad \int_I \nu = 0,$$

In practice, to perform SFPCA we can:

1. Transform the data to clr
2. Estimate the covariance operator from clr transformed data (in L^2) via the sample estimator
3. Compute the eigenfunctions of the sample covariance operator
4. Back-transform the results to get the eigenfunction in B^2

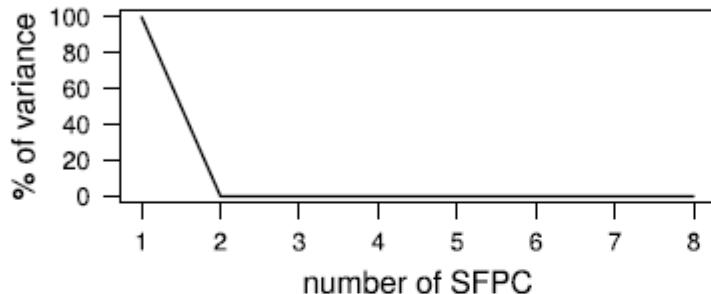
Note. The functional scores in B^2 and L^2 coincide, as well as the eigenvalues

3.3. Examples in Bayes spaces

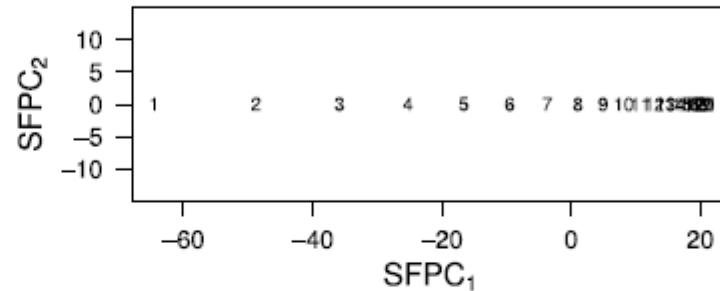
An Example with Gaussian distributions

Example 1: Dataset of Gaussian distributions

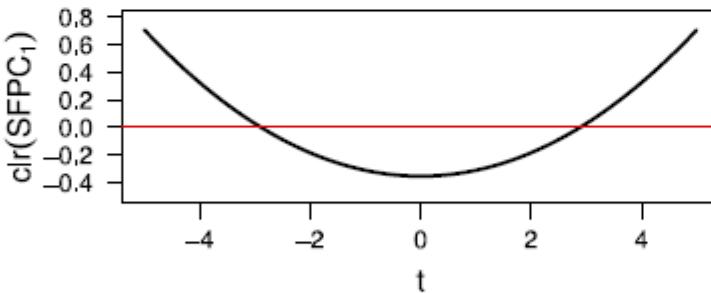
Note. The Gaussian distribution with zero-mean belongs to a **1-parametric exponential family**



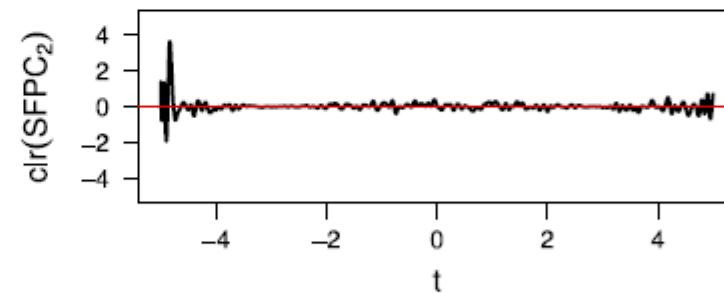
(a) Explained variance.



(b) Scores along SFPC₁ and SFPC₂.



(c) SFPC₁ (100% of variability).



(d) SFPC₂ (0% of variability).

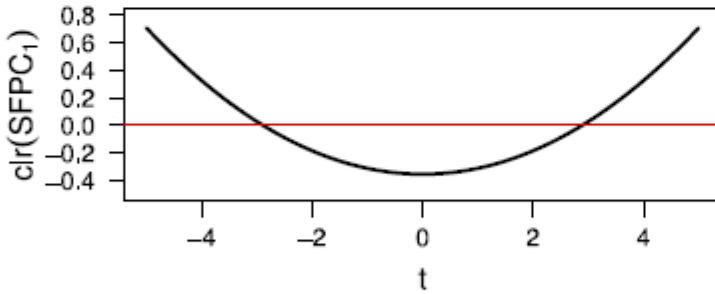
3.3. Examples in Bayes spaces

An Example with Gaussian distributions

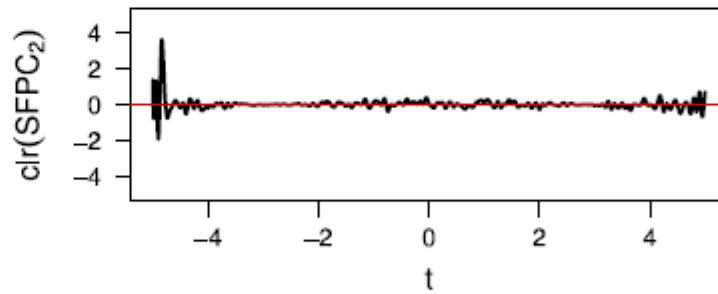
Example 1: Dataset of Gaussian distributions

Note. The Gaussian distribution with zero-mean belongs to a **1-parametric exponential family**

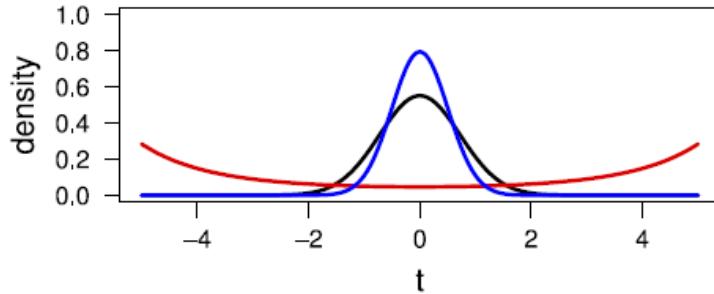
Interpretation of the SFPC (simplicial functional principal components)



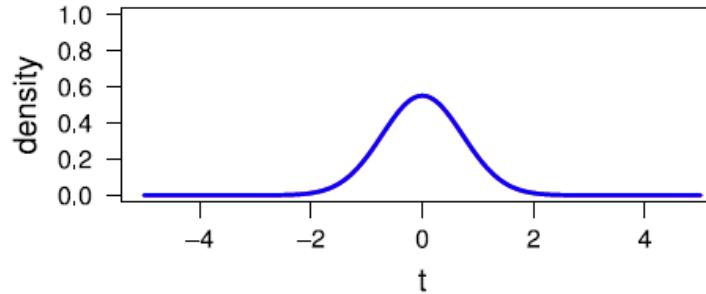
(c) SFPC₁ (100% of variability).



(d) SFPC₂ (0% of variability).



(e) Mean $\pm \sqrt{\lambda_1}$ SFPC₁.



(f) Mean $\pm \sqrt{\lambda_2}$ SFPC₂.

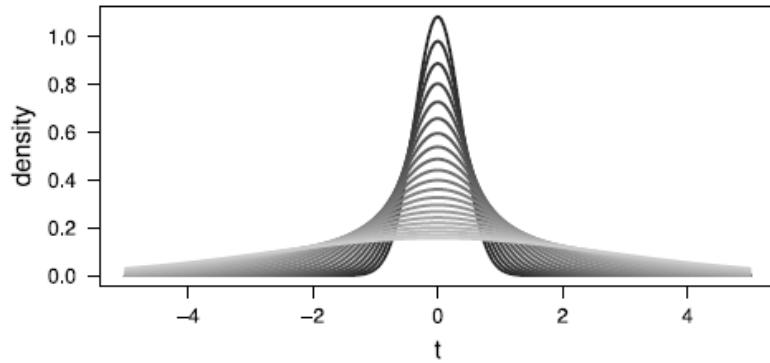
3.3. Examples in Bayes spaces

An Example with Gaussian distributions

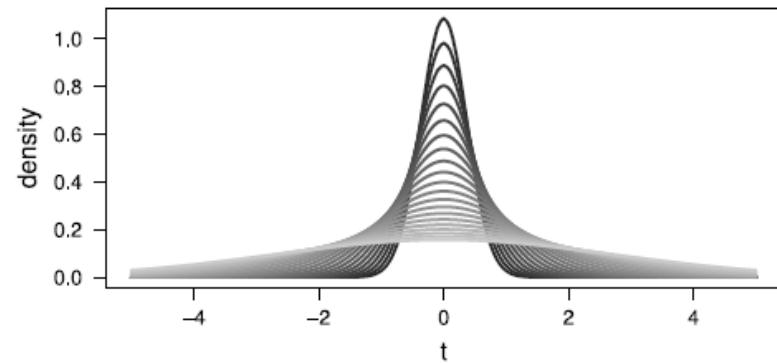
Example 1: Dataset of Gaussian distributions

Note: The Gaussian distribution with zero-mean belongs to a 1-parametric exponential family

Dimensionality reduction: exact approximation with 1 SFPC



(i) Original densities.



(j) Approximated densities (via SFPC₁).

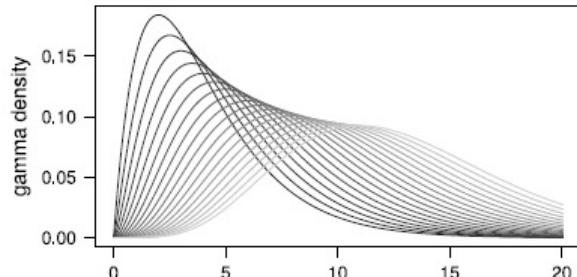
3.3. Examples in Bayes spaces

An Example with Gamma distribution

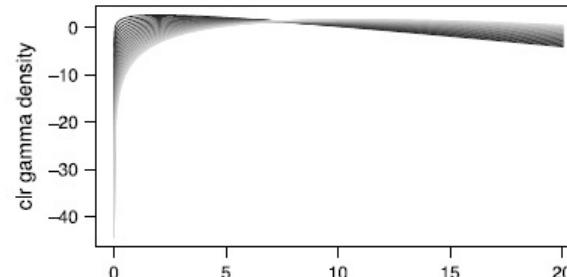
Example 2: Dataset of Gamma distributions

Consider a Gamma distribution with scale parameter $\theta = 2$ and shape parameter $\kappa_i = 2 + (i - 1)/4$ for $i = 0, \dots, 19$, $I = [e^{-7}, e^3]$:

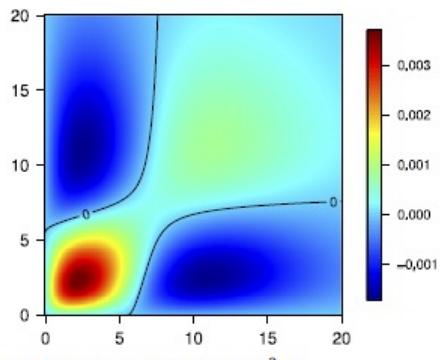
$$f(t; \kappa_i) = {}_{\mathcal{B}} t^{\kappa_i-1} \exp \left\{ -\frac{t}{2} \right\}, \quad t \in I.$$



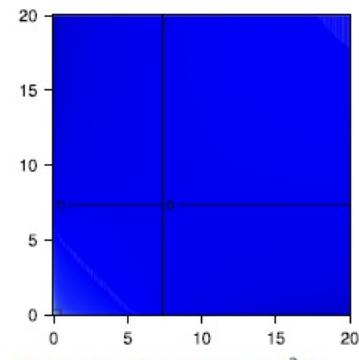
(a) Density functions.



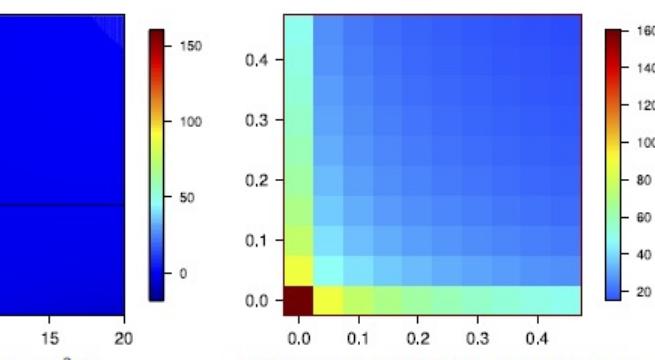
(b) Clr-transforms.



(c) Covariance function in $L^2(I)$.



(d) Covariance function in $\mathcal{B}^2(I)$.



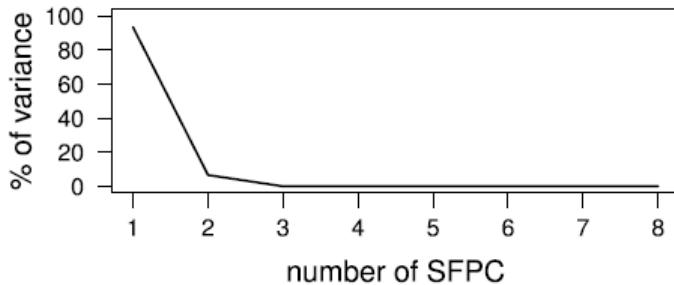
(e) A zoom into the top-left corner of the covariance function in $\mathcal{B}^2(I)$.

3.3. Examples in Bayes spaces

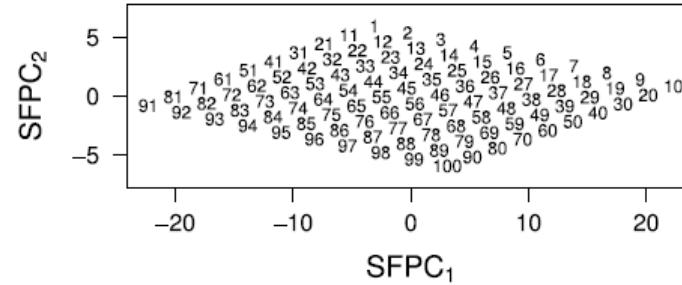
An Example with Gamma distribution

Example 2: Dataset of Gamma distributions

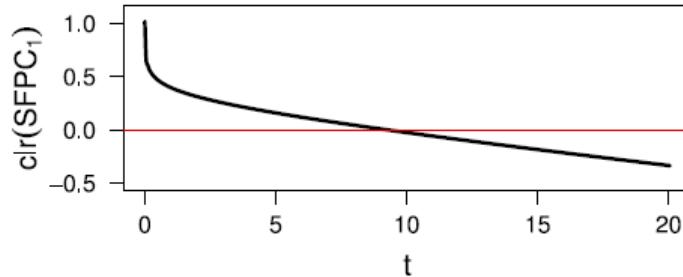
Note. The Gaussian distribution with belongs to a **2-parametric** exponential family.



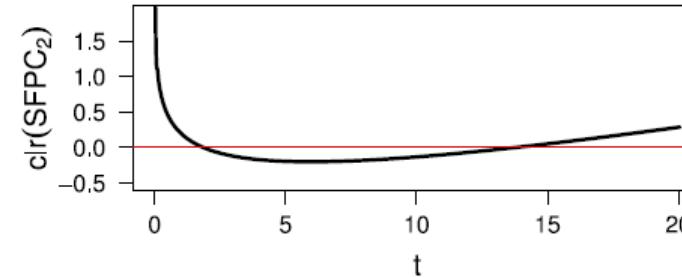
(a) Explained variance.



(b) Scores along SFPC₁ and SFPC₂.



(c) SFPC₁ (93.4% of variability).



(d) SFPC₂ (6.6% of variability).

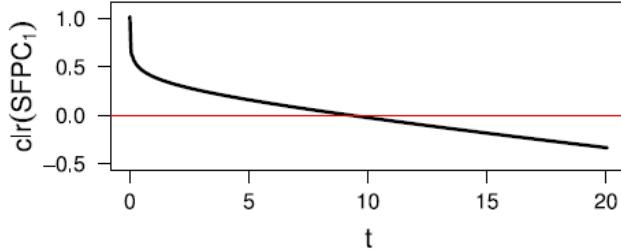
3.3. Examples in Bayes spaces

An Example with Gamma distribution

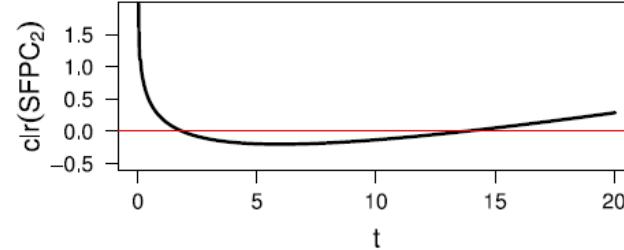
Example 2: Dataset of Gamma distributions

Note. The Gaussian distribution with belongs to a **2-parametric** exponential family.

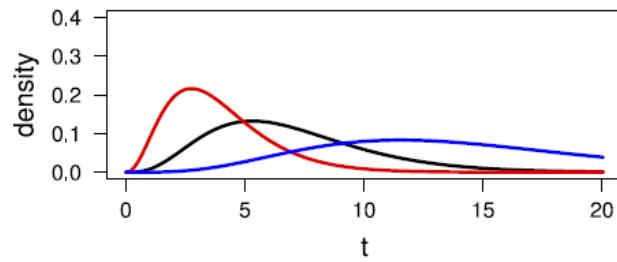
Interpretation of the SFPC (simplicial functional principal components)



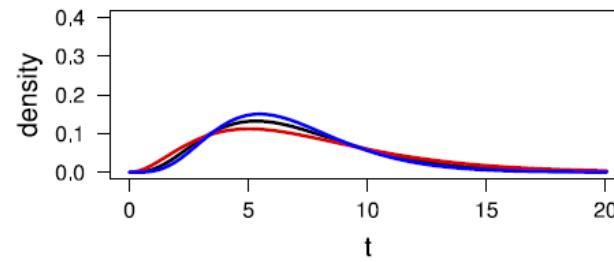
(c) SFPC₁ (93.4% of variability).



(d) SFPC₂ (6.6% of variability).



(e) Mean $\pm \sqrt{\lambda_1}$ SFPC₁.



(f) Mean $\pm \sqrt{\lambda_2}$ SFPC₂.

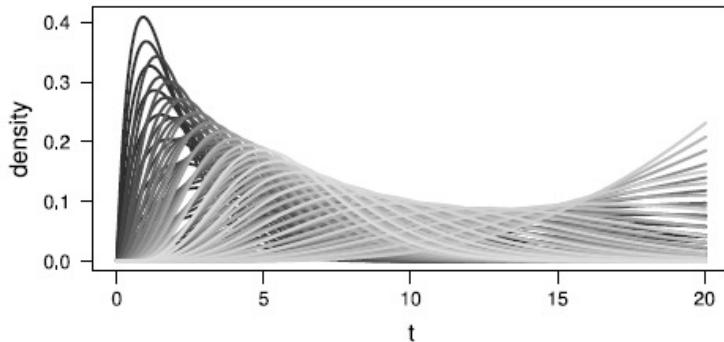
3.3. Examples in Bayes spaces

An Example with Gamma distribution

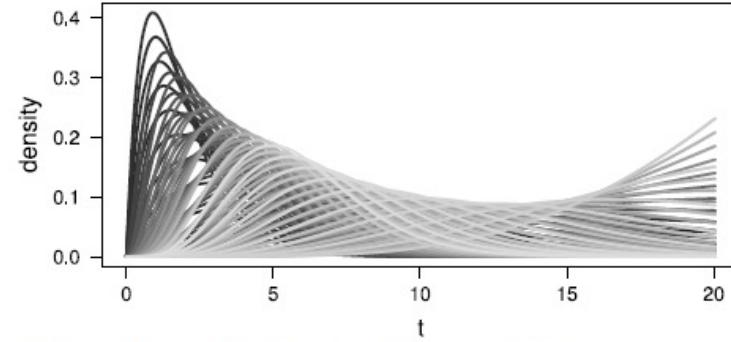
Example 2: Dataset of Gamma distributions

Note. The Gaussian distribution with belongs to a **2-parametric** exponential family.

Dimensionality reduction: exact approximation with 1 SFPC



(i) Original densities.



(j) Approximated densities (via SFPC₁ and SFPC₂).

More in general: if densities belong to a k -parametric exponential family, with k_0 parameters that vary linearly independently, SPFCA singles out k_0 SFPCs associated with non-null eigenvalues

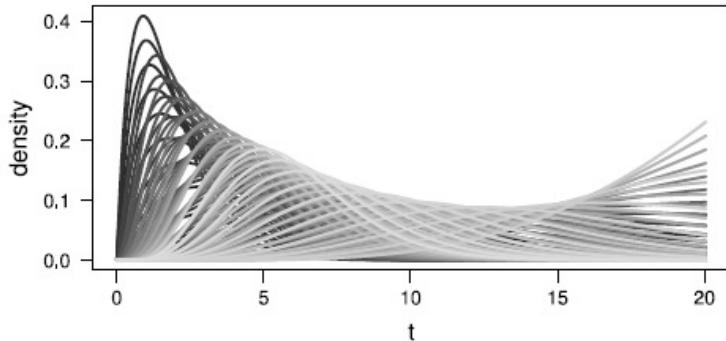
3.3. Examples in Bayes spaces

An Example with Gamma distribution

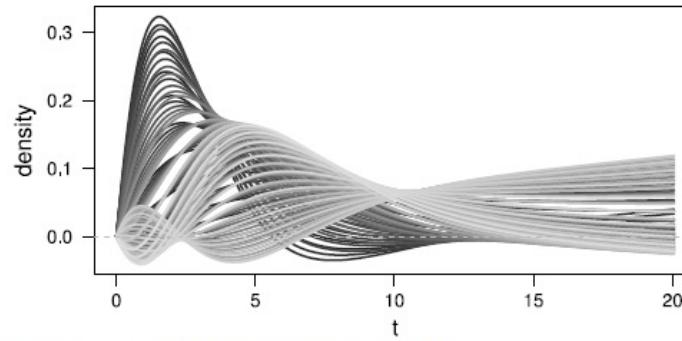
Example 2: Dataset of Gamma distributions

Note. The Gaussian distribution with belongs to a **2-parametric** exponential family.

What if we forget about the data features and work in L2 (i.e., inappropriate Hilbert space embedding)?



(i) Original densities.



(h) Approximated densities (via FPC₁ and FPC₂).

We can't guarantee that the data characteristic (especially their constraints) are respected if we don't pay attention to choose a (Hilbert) space that properly captures the data features

3.2 Examples in B²

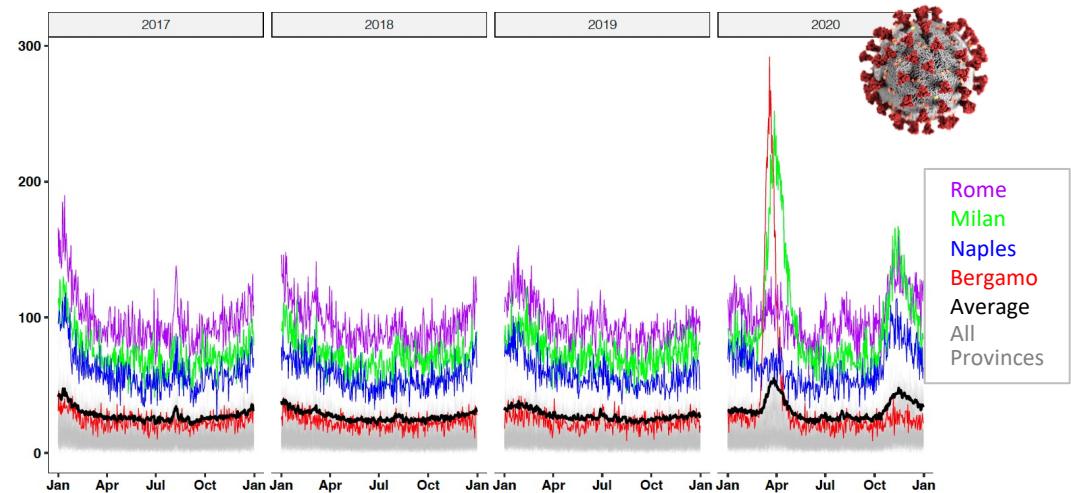
Italian Mortality Data

Example: Analysis of Mortality data in Italian Provinces

- Counts of daily deaths *from all causes* in Italian Provinces



109 Italian Provinces



Daily death counts over 2017-2020, 70+ years

Scimone, Menafoglio, Sangalli, Secchi (SPASTA, 2023)

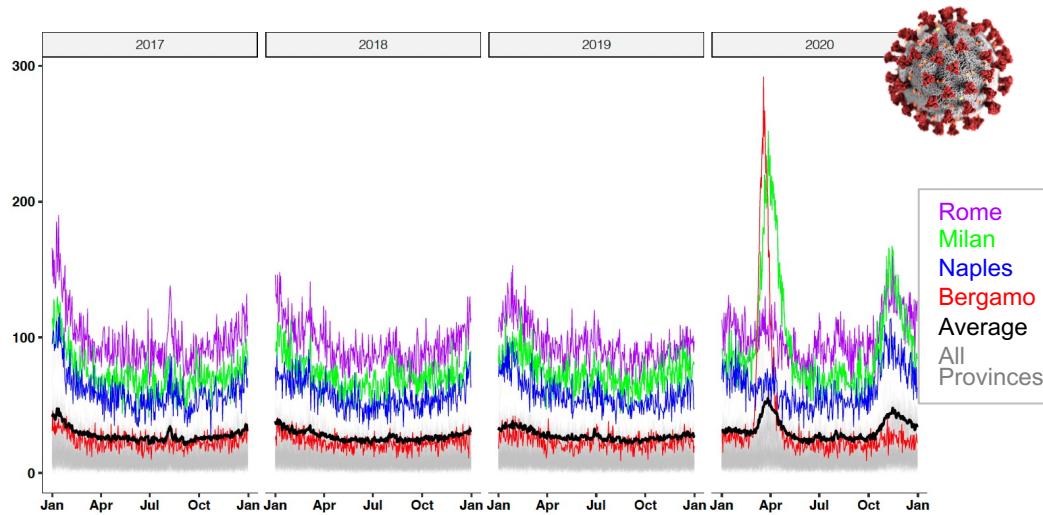
3.2 Examples in B²

Italian Mortality Data

- **A proxy** for measuring the immediate impact of the pandemic on Italian communities
- **High-quality** data (source: ISTAT), recorded at a very fine granular scale over time and space
- **Not affected by varying definitions or testing capabilities**, as for deaths attributed to COVID-19 or number of cases
- **Integrate the direct and indirect effects of the pandemic shock**, registering
 - The consequences of the containment policies
 - The disruption at the local level of the health and welfare system overwhelmed by the struggle against COVID-19.
- From now on, focus on 70+ years population

3.2 Examples in B²

Which data object for the statistical analysis?

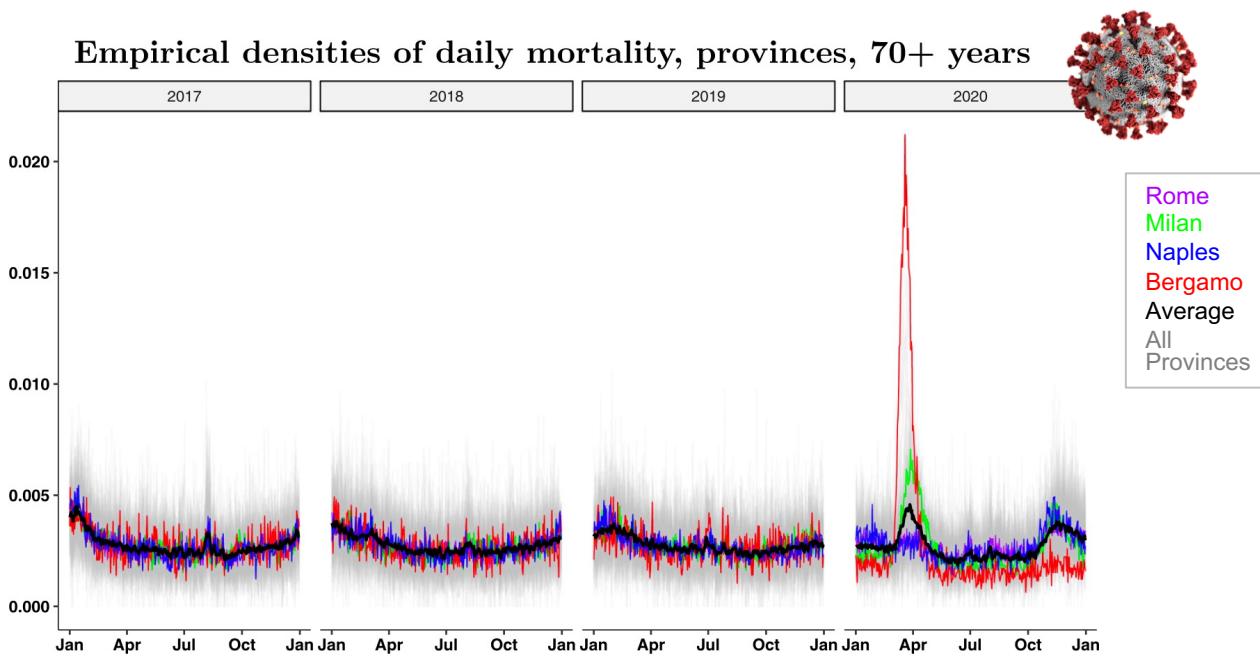


- If data are just considered on an absolute scale, highly populated provinces will mask the entire variability of the data
- If just considering the total over the year, the dynamic along year is lost

3.2 Examples in B^2

Which data object for the statistical analysis?

- **Data object:** the distribution (probability density function - PDF) of the variable *time of death* along the year, considered as an element of B^2
- **Relative scale:** we do not want to look at the absolute counts, but on the relative proportion of deaths within each Province



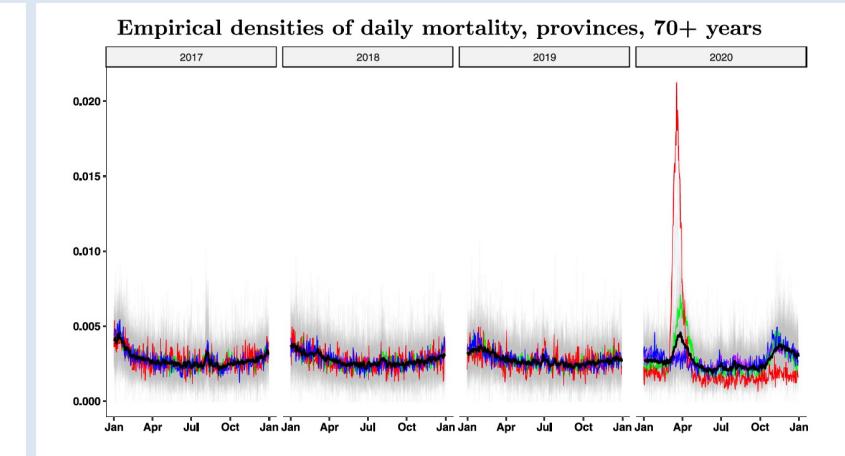
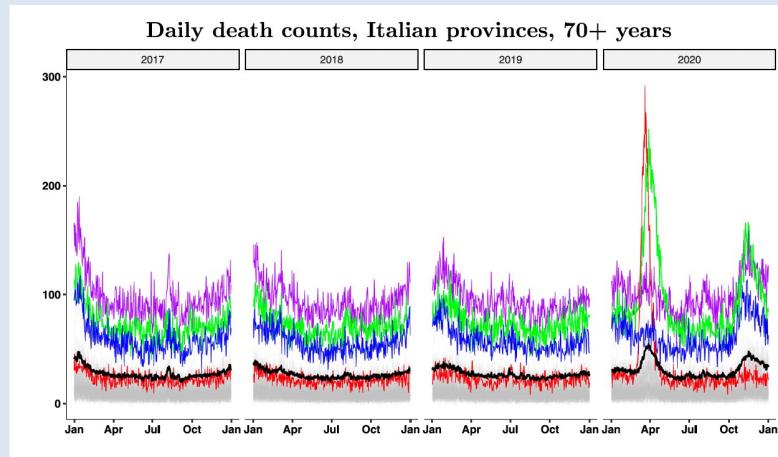
3.2 Examples in B²

Which data object for the statistical analysis?

- **Data object:** the distribution (probability density function - PDF) of the variable *time of death* along the year, considered as an element of B^2
 - **Relative scale:** we do not want to look at the absolute counts, but on the relative proportion of deaths within each Province

Equivalence relation: f and g are equivalent if they are proportional (scale invariance)

Once embedded in B^2 , the scale of the phenomenon is of no relevance!



3.2 Examples in B^2

Detrending the data

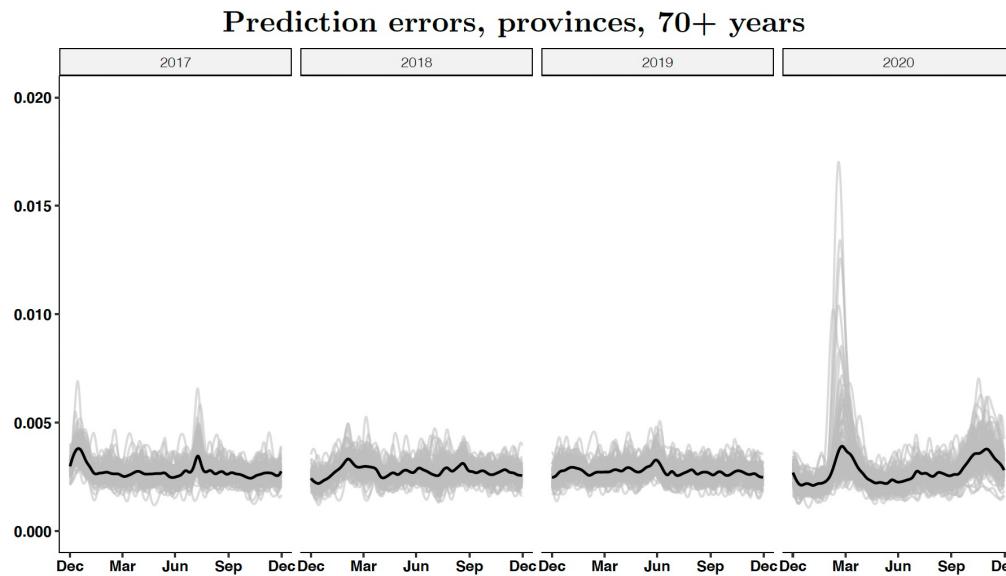
Unpredictable component of mortality

We analyze the prediction error of a model targeting the mortality density at year y given the MD at years $y-1, \dots, y-4$

$$\hat{f}_{iy} = \hat{\beta}_{0y-1} + \langle \hat{\beta}_{y-1}(\cdot, t), \bar{f}_{iy} \rangle_{B^2} \quad \text{Forecast for year } y \text{ calibrated at year } y-1$$

$$\bar{f}_{iy} = \frac{1}{4} \sum_{r=y-4}^{y-1} f_{ir} \quad \text{Predictor: average MD alongy years } y-1 \text{ to } y-4$$

$$\delta_{iy}(t) = f_{iy}(t) - \hat{f}_{iy}(t) \quad \text{Prediction error (not the residuals!)}$$



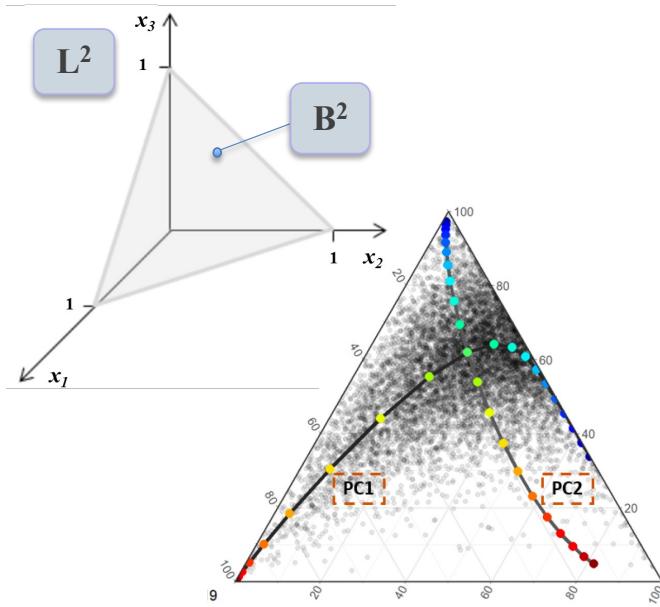
Note. The “0” in B^2 is the uniform distribution

3.2 Examples in B^2

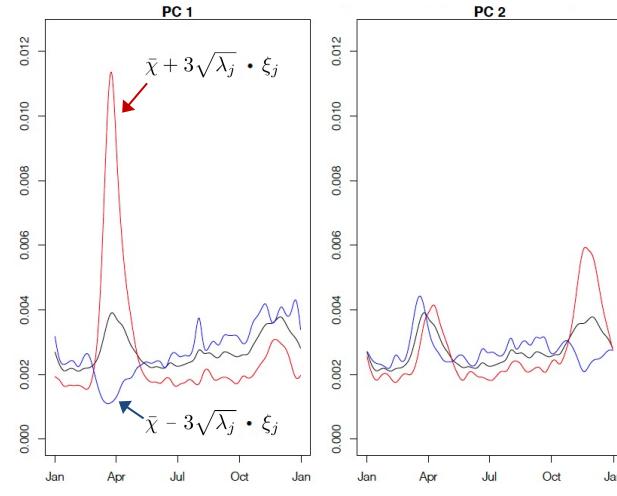
SFPCA: Results

Interpretation of the first two SFPCs

Recall: SFPCs represent direction of max variability in the infinite dimensional simplex B^2



Mean perturbed by the PCs



First wave

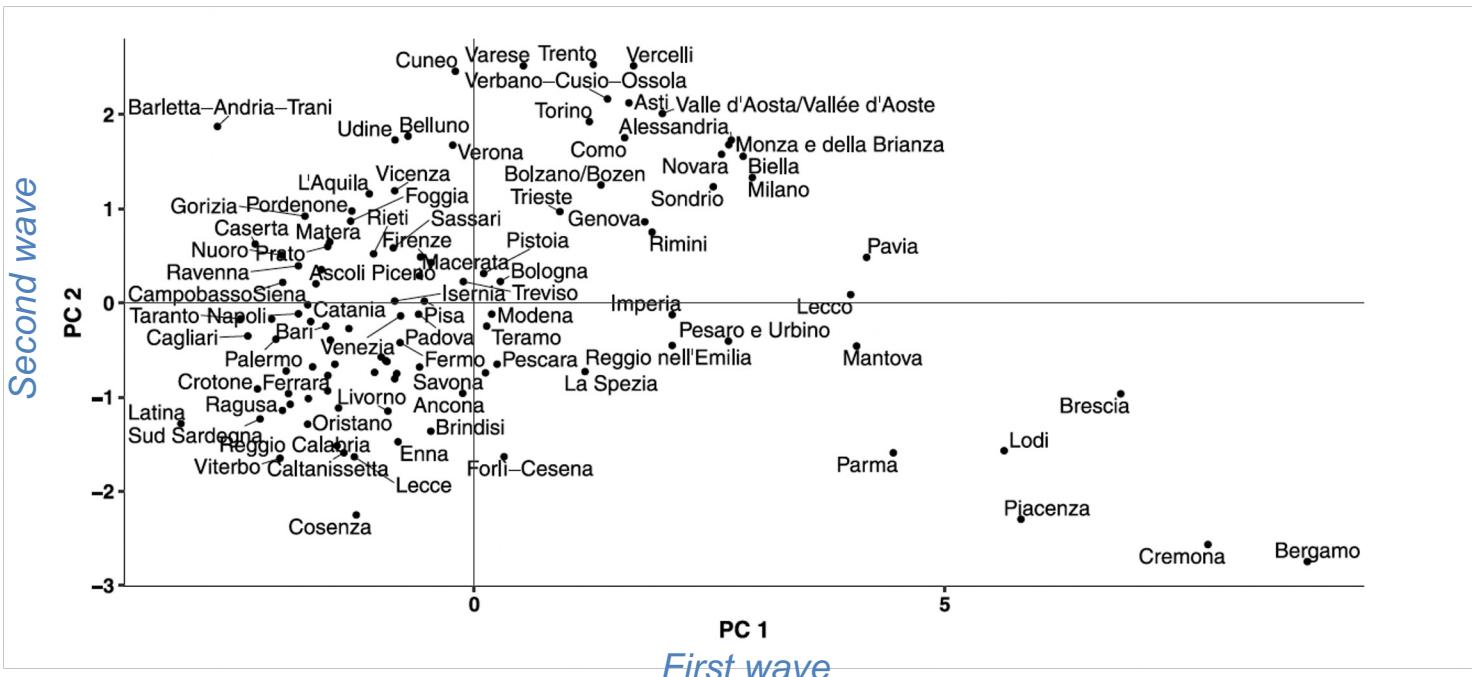
Second wave

SFPCs seems to be associated with the first two waves of Covid pandemic

3.2 Examples in B²

SFPCA: Results

Scores along the first two SFPCs

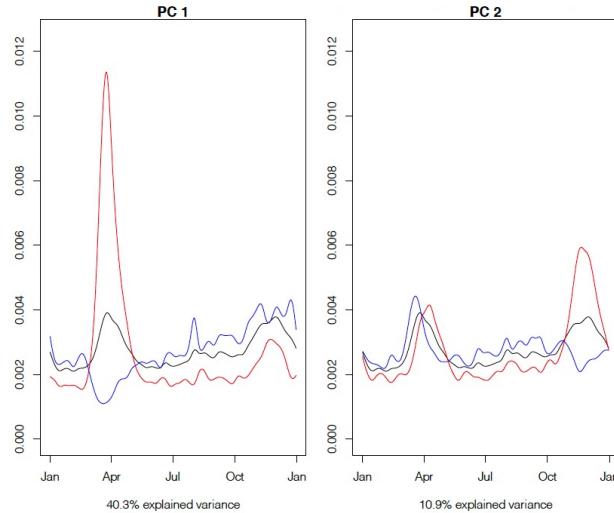


3.2 Examples in B²

SFPCA: Results

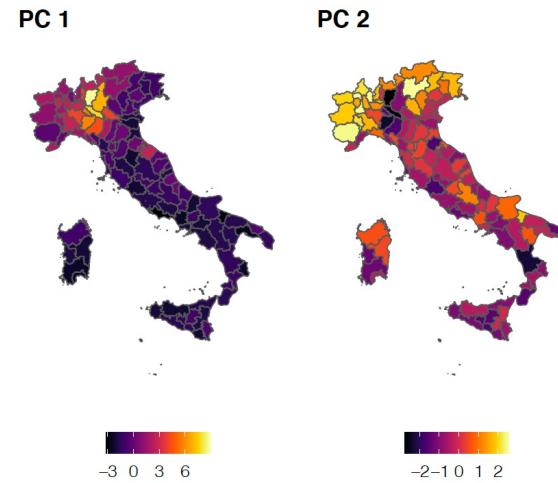
Scores along the first two SFPCs

Mean perturbed by the PCs



First wave

Maps of the PC scores



Second wave

Codes to replicate the analysis are available at:
<https://github.com/RiccardoScimone/Mortality-densities-italy-analysis>