3. FDA & Dimensionality reduction in Hilbert spaces

Course Agenda

1. **Hilbert space model for functional data**
    1.1. Basics notions on Hilbert spaces
    1.2. Hilbert space embedding for functional data
    1.3. Formal definition of functional data
2. **Smoothing and interpolation of functional data**
    2.1. Basis function
    2.2. Least square smoothing
    2.3. Smoothing with a differential penalization
3. **FDA & Dimensionality reduction in Hilbert spaces**
    3.1. Functional Principal Components in Hilbert spaces
    3.2. Examples in L2
4. **Data alignment and clustering**
    4.1 Phase and amplitude variability
    4.2 Landmark and continuous registration
    4.3 Decoupling phase and amplitude variability
    4.4 K-mean alignment

Recall: Principal Component Analysis

Problem: Given a dataset of *N* zero-mean multivariate observations in $\mathbb{R}^p$, $X_1, \ldots, X_N$, find the orthonormal directions $\boldsymbol{a_1}, \ldots, \boldsymbol{a_p}$ of maximum variability (for the dataset).

Equivalently, for *k*= *1,…,p,* find:

$$\begin{aligned} \mathbf{a}_k &= argmax_{\mathbf{a}\in\mathbb{R}^p} Var(\mathbf{a}'\mathbf{X}) \\ \text{subject to:} \quad & \mathbf{a}'\mathbf{a} = 1, \ \mathbf{a}_\mathbf{j}'\mathbf{a} = 0 \text{ for } j < k \end{aligned}$$

- We can re-write the problem as

$$\begin{aligned} \mathbf{a}_k &= argmax_{\mathbf{a}\in\mathbb{R}^p} \frac{1}{N}\sum_{i=1}^{N}(\mathbf{a}'\mathbf{X}_i)^2 \\ \text{subject to:} \quad & \mathbf{a}'\mathbf{a} = 1, \ \mathbf{a}_\mathbf{j}'\mathbf{a} = 0 \text{ for } j < k \end{aligned}$$

or, equivalently

$$\begin{aligned} \mathbf{a}_k &= argmax_{\mathbf{a}\in\mathbb{R}^p} \frac{1}{N}\sum_{i=1}^{N} < \mathbf{a}, \mathbf{X}_i >^2 \\ \text{subject to:} \quad & ||\mathbf{a}|| = 1, \ < \mathbf{a}_\mathbf{j}, \mathbf{a} >= 0 \text{ for } j < k \end{aligned}$$

Note 1. We assume *N>p* and absence of collinearity, i.e. the data matrix is full rank.

Note 2 . If $X_1, \ldots, X_N$ are not zero-mean, they can be centered by subtracting the (sample) mean. For unbisasedness, divide by *N-1* instead of *N*.

Recall: Principal Component Analysis

Problem: Given a dataset of *N* zero-mean multivariate observations in $\mathbb{R}^p$, $X_1, \ldots, X_N$, find the orthonormal directions ***a₁, …, aₙ*** of maximum variability, i.e., those solving for k=1,…p,

$$\begin{aligned} \mathbf{a}_k &= argmax_{\mathbf{a}\in\mathbb{R}^p} \frac{1}{N}\sum_{i=1}^{N} < \mathbf{a}, \mathbf{X}_i >^2 \\ \text{subject to:} &\quad ||\mathbf{a}|| = 1, < \mathbf{a_j}, \mathbf{a} >= 0 \text{ for } j < k \end{aligned}$$

Solution: Call *S* the sample covariance matrix of $X_1, \ldots, X_N$. Then, the **principal components** are found as the eigenvectors of the matrix *S*; for *k=1,…, p*, they solve the eigen-equation

$$Se_k = \lambda_k e_k$$

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.,

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

Recall: Principal Component Analysis

Problem: Given a dataset of *n* zero-mean multivariate observations in $\mathbb{R}$ $X_1, \ldots, X_N$,
find the directions of maximum variability of the dataset, i.e., those maximizing

$$\frac{1}{N}\sum^{N}\langle \boldsymbol{a}, X_i\rangle^2 \quad \text{subject to} \quad \|\boldsymbol{a}\| = 1$$

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

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.,

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

Problem: Given a dataset of *N* zero-mean functional observations in *H*, $X_1, \ldots . X_N$, find the directions of maximum variability (in *H*) of the dataset, i.e., for *k=1,…,N*, find ξ_k maximizing

$$\frac{1}{N}\sum_{i=1}^{N} < \xi, \mathbf{X}_i >_H^2$$

$$\text{subject to:} \qquad ||\xi|| = 1, \; < \xi_j, \xi >_H = 0 \text{ for } j < k$$

- We look for an orthonormal system 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 1. If the data are not zero-mean, they can be centered by subtracting the (sample) mean. *N* should then be replaced by *N-1*.

Note 2. If data are linearly independent and centered on the sample mean, we can find at most *N-1* principal components.

Problem: Given a dataset of *n* zero-mean functional observations in *H*, $X_1, \ldots . X_N$, find the directions of maximum variability (in *H*) of the dataset, i.e., for *k=1,…,N*, find ξ_k maximizing

$$\frac{1}{N}\sum_{i=1}^{N} < \xi, \mathbf{X}_i >_H^2$$

$$\text{subject to:} \qquad ||\xi|| = 1, \; < \xi_j, \xi >_H = 0 \text{ for } j < k$$

- As in 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

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

Note. If data are centered on the sample mean, divide by *N-1* for unbiasedness.

Functional Principal Component Analysis
FPCA and sample covariance operator

Problem: Given a dataset of *n* zero-mean functional observations in *H*, $X_1, \ldots . X_N$, find the directions of maximum variability (in *H*) of the dataset, i.e., for *k=1,…,N*, find ξ_k maximizing

$$\frac{1}{N}\sum_{i=1}^{N} < \xi, \mathbf{X}_i >_H^2$$

$$\text{subject to:} \qquad ||\xi|| = 1, \; < \xi_j, \xi >_H = 0 \text{ for } j < k$$

Solution: Let *S* be the sample covariance operator of $X_1, \ldots, X_N$. Then, the **functional principal components** $\xi_1, \ldots, \xi_N$ 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 eigenvector 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.,

$$x_{ik} =< X_i, \xi_k >$$

Note. If data are centered on the sample mean, we can find at most *N-1* principal components

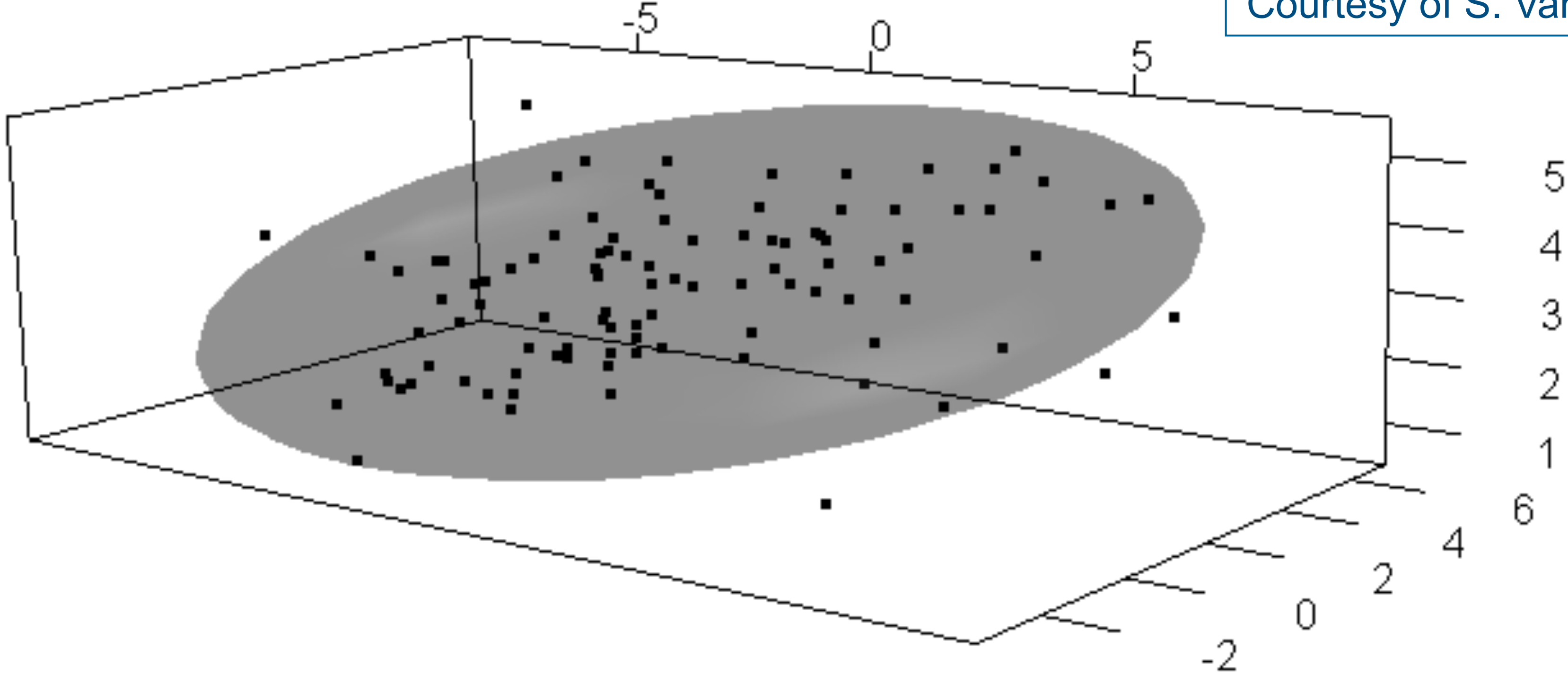
Functional Principal Component Analysis
Dimensionality reduction and Interpretation of the results

- 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:
    $$\overline{X} \pm \sqrt{\lambda_k}\xi_k$$
  - Plotting the projection of the dataset along each component or along the first *p* components
    $$\overline{X} + x_{ik}\xi_k$$
    $$\overline{X} + \sum_{k=1}^{p} x_{ik}\xi_k$$

Courtesy of S. Vantini



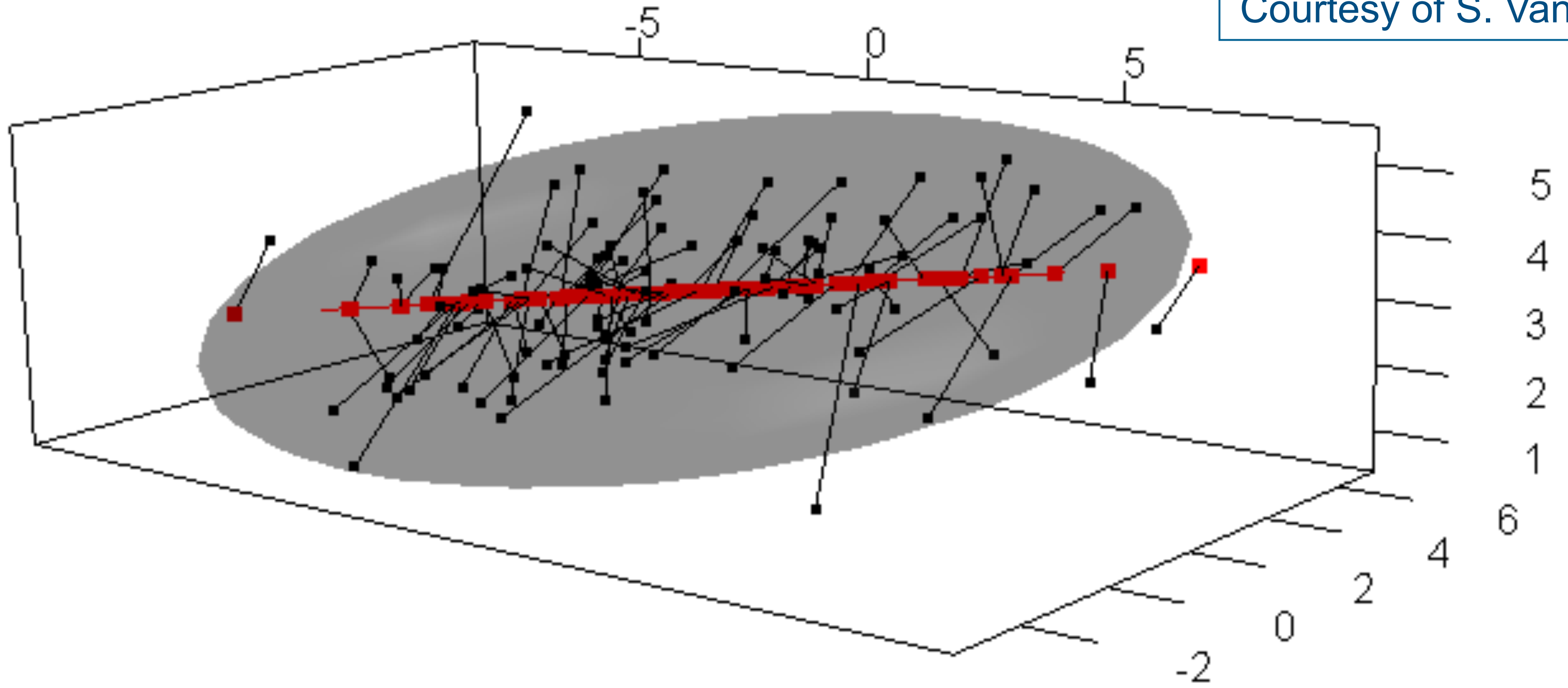
Problem: find the space of dimension *k* that best approximate the data in the mean square sense

If k=0: sample mean

Functional Principal Component Analysis

FPCA as space of best approximation

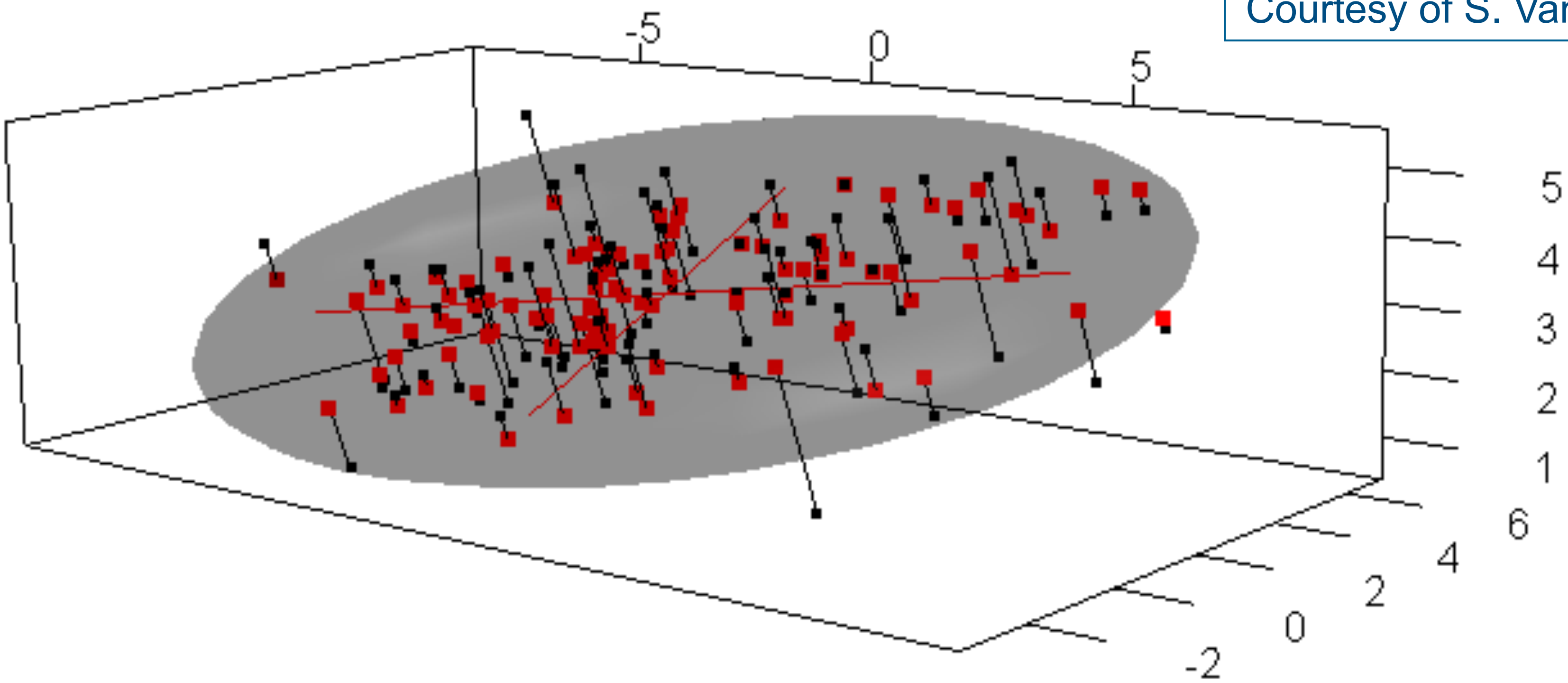
Courtesy of S. Vantini



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

Courtesy of S. Vantini



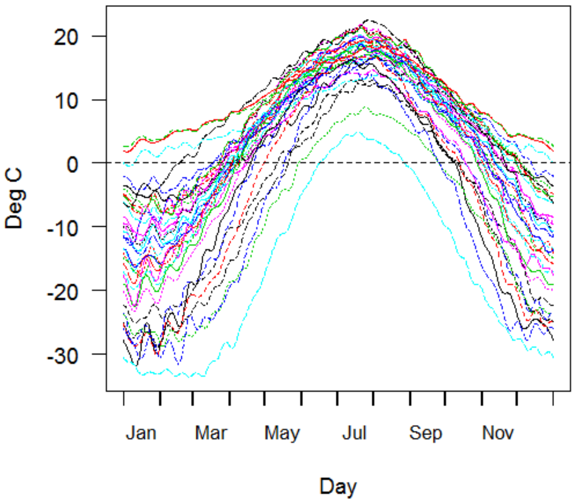
Problem: find the space of dimension *k* that best approximate the data in the mean square sense

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

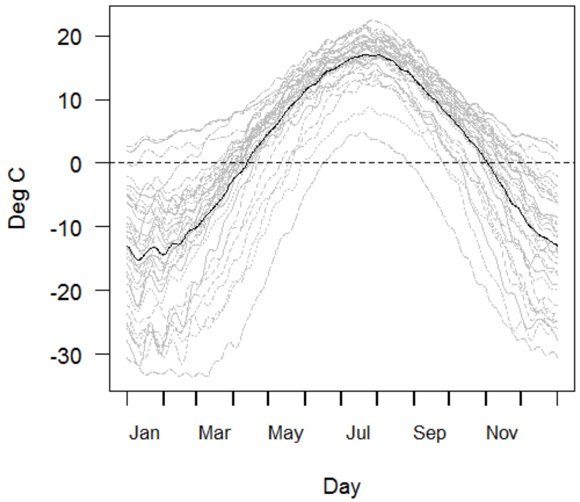
…

Ramsay Silverman 2005 Springer

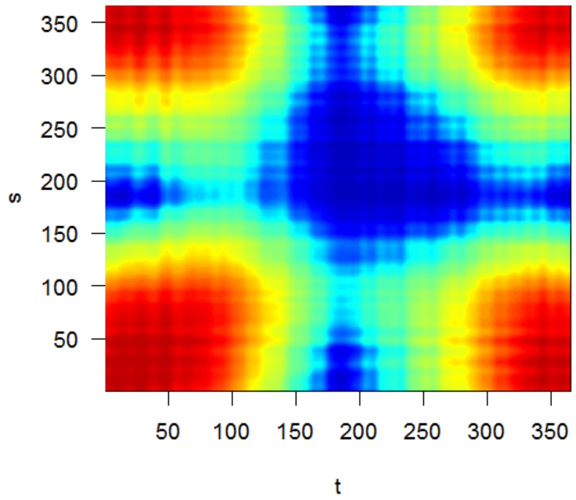
Example. Dataset of Temperatures in Canada (35 observations)



Functional dataset



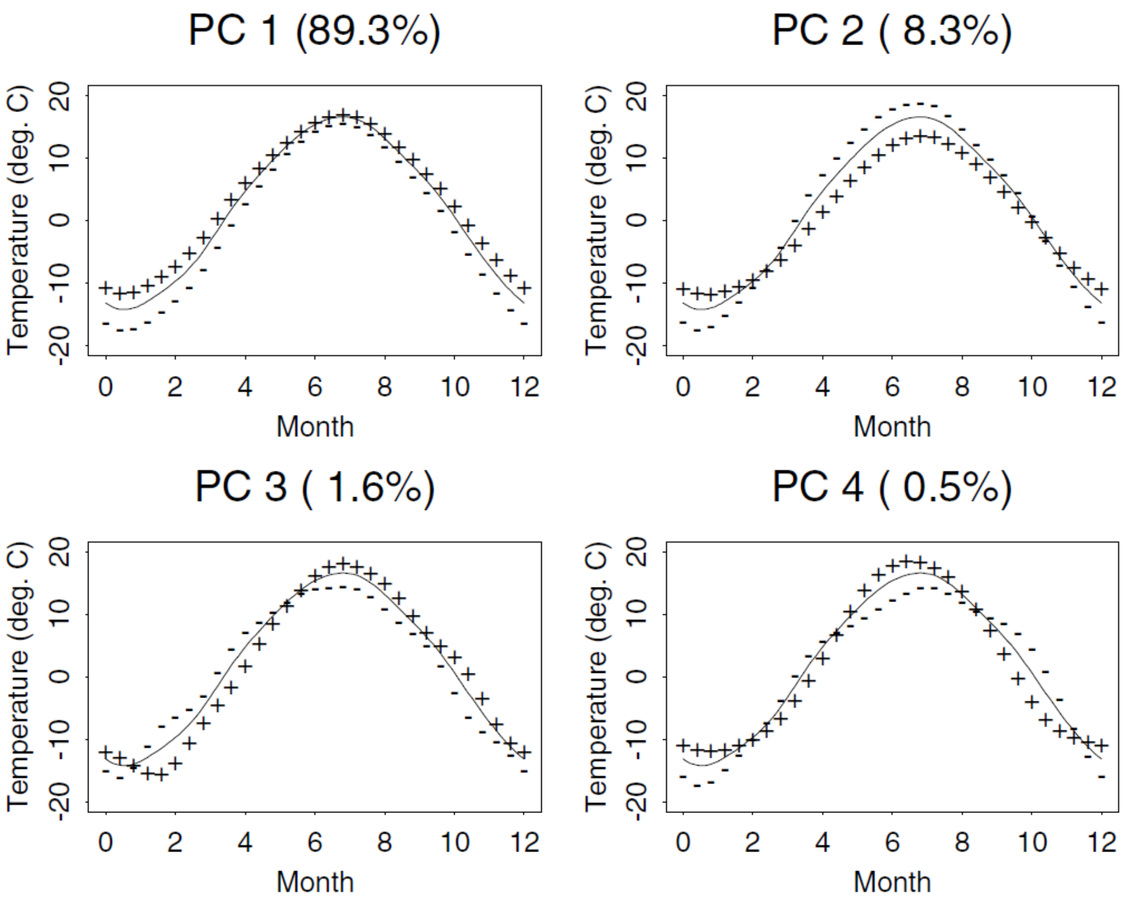
Sample mean



Sample covariance kernel

3.2 Examples in L^2
Dataset of Canadian temperatures

Ramsay Silverman 2005 Springer



Ramsay Silverman 2005 Springer