"Shape Invariant" curves estimation and application to prediction problems

J-M. LOUBES

Institut de Mathématiques de Toulouse

IHP Rencontres INRIA-AMIES-EDF 2012

Outline

Shape invariant model

Curve Registration

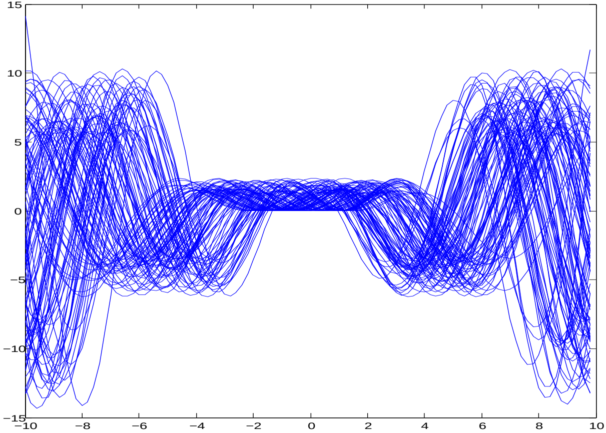
Prediction using shape invariant model

A practical Issue in functional analysis

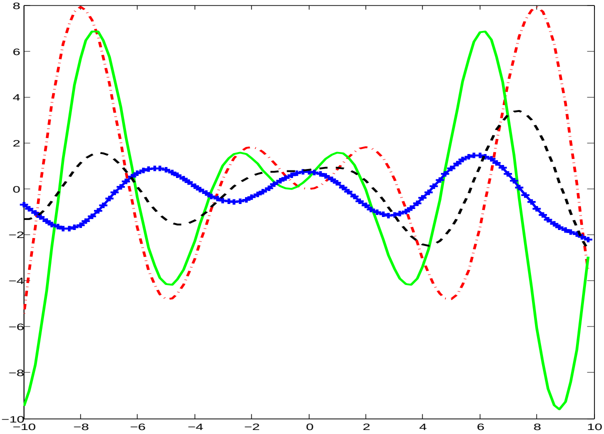
- Functional data usually convey a **general information**, which reflects the inner structure of the observations
- but also different **sources of variation** which blurs the data and prevent the use of the **Euclidean mean**.

Common in biology, medicine (mixed effect model), economics or sociology (panel data) and also in electric charge.
Appears when individuals may differ slightly from a pattern which represents the shape of the observations and break **the Euclidean structure** of the data.

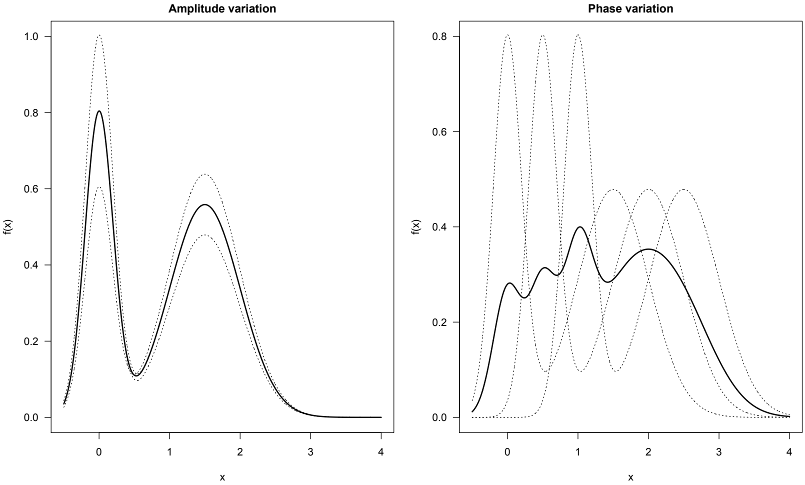
Simulations : Do we trust the mean ?



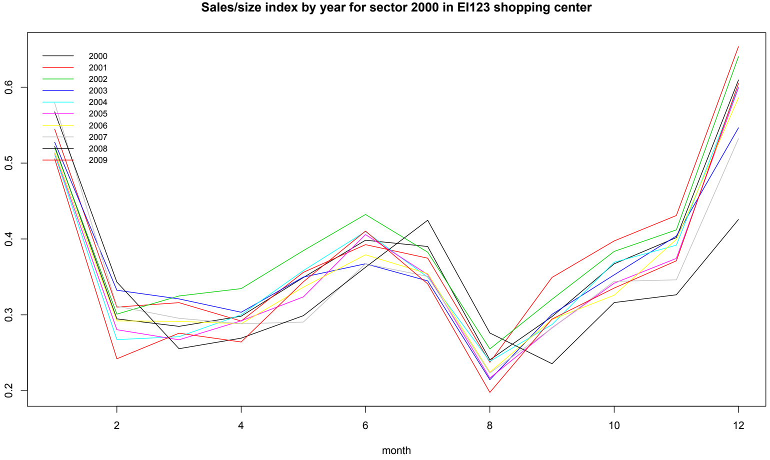
Simulations : Do we trust the mean ?



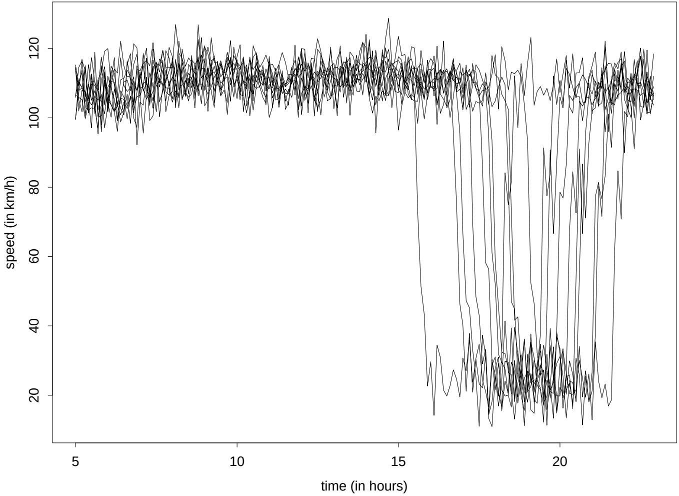
Example : in oligonucleotide array



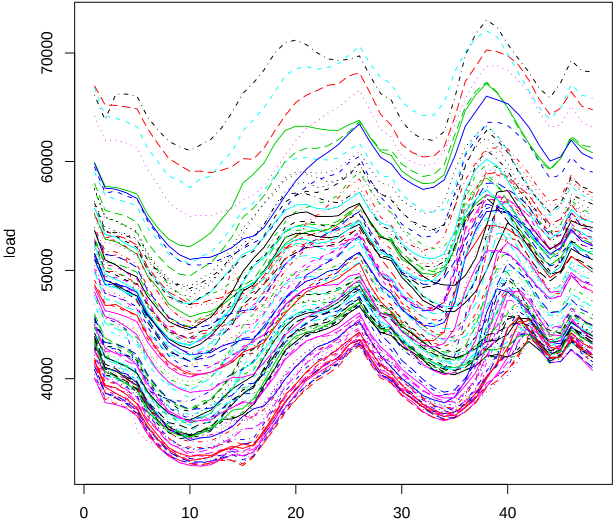
Example : income of stores from a mall



Example : traffic jam on a motorway



Example : EDF daily load



Objectives of the statistical studied

Main issues:

- Recover the **shape or pattern** of the data from the warped observations
- Estimate the individuals deformations
- Understand the deformations : cluster the data according to the distance given by the observations , i.e the **geometry of the data**.
- **Forecast the future deformations** using auxiliary information

$\Longrightarrow$ different framework : shape analysis of curves viewed as a manifold with an inner geometry.

General Model

The **regression model**:

$$Y_{i,j} = f_j^\star(t_{ij}) + \sigma\epsilon_{i,j}, \quad i = 1,\ldots,n,\ j = 1,\ldots,J.$$

where

- $f_j^\star$ models the j^{th} signal (unknown);
- t_{ij} the observation points (known).
- $\epsilon_{i,j}$ is white noise (unknown), and σ variance (unknown)

Assumption: There exists a **common shape** of the signal $f^\star$ and **warping operators** Φ_j,

$$f_j^\star = \Phi_j f^\star, \quad j = 1, \ldots, J.$$

Aim: Estimation of the deformations and the template $f^\star$

Methodology

inverse problem of regression for unknown operator
=> Need for a Model for the warping operator

Parametric Modeling and Semiparametric Statistics

The warping operator can be **parametrized** by $\theta \in \Theta \subset \mathbb{R}^d$.

$$\forall j = 1, \ldots, J, \quad \Phi_j = \Phi_{\theta_j}$$

$$Y_{ij} = \Phi_{\theta_j^\star}[f^\star](t_{ij}) + \sigma\epsilon_{ij}$$

- $f^\star$ is the main feature
- $\Phi_{\theta^\star}[.]$ is a parametric warping operator
- $\theta_j^\star$ local warping parameters

Objective : estimate $f^\star$ by estimating the $\theta_j^\star$'s
Semiparametric framework with unknown distribution $\mathbb{P}_{f^\star,\theta^\star}$

A class of deformations $\theta = (a \quad b \quad \upsilon)'$

$$\Phi_\theta : f(\cdot) \rightarrow af(\cdot - b) + \upsilon$$

Parametric Modeling and Semiparametric Statistics

The warping operator can be **parametrized** by $\theta \in \Theta \subset \mathbb{R}^d$.

$$\forall j = 1, \dots, J, \quad \Phi_j = \Phi_{\theta_j}$$

$$Y_{ij} = \Phi_{\theta_j^\star}[f^\star](t_{ij}) + \sigma \epsilon_{ij}$$

- $f^\star$ is the main feature
- $\Phi_{\theta^\star}[.]$ is a parametric warping operator
- $\theta_j^\star$ local warping parameters

Objective : estimate $f^\star$ by estimating the $\theta_j^\star$'s

Semiparametric framework with unknown distribution $\mathbb{P}_{f^\star, \theta^\star}$

A class of deformations $\theta = (a \quad b \quad \upsilon)'$

$$\Phi_\theta : f(\cdot) \to af(\cdot - b) + \upsilon$$

Parametric Modeling and Semiparametric Statistics

The warping operator can be **parametrized** by $\theta \in \Theta \subset \mathbb{R}^d$.

$$\forall j = 1, \ldots, J, \quad \Phi_j = \Phi_{\theta_j}$$

$$Y_{ij} = \Phi_{\theta_j^\star}[f^\star](t_{ij}) + \sigma \epsilon_{ij}$$

- $f^\star$ is the main feature
- $\Phi_{\theta^\star}[.]$ is a parametric warping operator
- $\theta_j^\star$ local warping parameters

Objective : estimate $f^\star$ by estimating the $\theta_j^\star$'s
Semiparametric framework with unknown distribution $\mathbb{P}_{f^\star,\theta^\star}$

A class of deformations $\theta = (a \quad b \quad \upsilon)'$

$$\Phi_\theta : f(\cdot) \to af(\cdot - b) + \upsilon$$

Curve Alignment

$$Y_{i,j} = a_j^\star f^\star(t_{ij} - b_j^\star) + \upsilon_j^\star + \sigma_j^\star \epsilon_{i,j}, \quad j = 1, \ldots, J, \ i \in \{1, \ldots, n\}^d.$$

Estimation of the parameters

$\mathbf{b_j^\star} \in \mathbb{R}^{\mathbf{d}}, \ \mathbf{a_j^\star} \in \mathbb{R}^*, \ \upsilon_{\mathbf{j}}^\star \in \mathbb{R}, \ \mathbf{j} = \mathbf{1}, \ldots, \mathbf{J}$

Lawton, W.M., Sylvestre, E.A. et Maggio, M.G. (1972): iterative method (**SEMOR**) based on the polynomial approximation of f^*.

Kneip, A. et Gasser, T.(1988): consistency of the SEMOR method.

Gamboa, F., Loubes, J-M. and Maza, E. (EJS 2007): Semi-parametric estimation of shifts.

Bigot J., Loubes, J-M. and Vimond, M. (PTRF 2012): Semiparametric estimation of rigid transformations on compact Lie groups,

Mathematical model

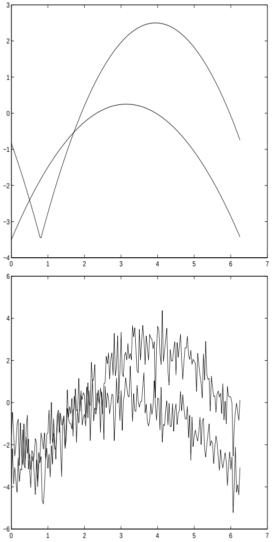
$$Y_{ij} = a_j^* f^*(t_i - b_j^*) + \upsilon_j^* + \sigma_j^* \epsilon_{ij},\ i \in I_n = \{1..n\}^d,$$

Assumptions :

- f^* $2\pi\mathbb{Z}^d$-periodic continuous
- $t_i = 2\pi(i)/n \in [0, 2\pi[^d$,
- $(\epsilon_{i,j})_{ij}$ are i.i.d, $\mathbb{E}\epsilon_{i,j} = 0$, $\mathbb{E}\epsilon_{i,j}^2 = 1$,
- the variance of noise ${\sigma_j^*}^2 = {\sigma^*}^2$,

Method:

1. Estimate the parameters $\theta^\star$
2. Invert the estimated operator to estimate the shape $f^\star$



Identifiability Constraint

Two sets of identifiability constraints are considered:

1. **a natural parametrization**:
   we consider one of the signal as a *reference*,

$$b_1^* = 0, \quad a_1^* = 1 \quad \text{and} \quad \upsilon_1^* = 0$$

2. **an alternative parametrization**:

$$b_1^* = 0, \quad \sum_{j=1}^{J} a_j^{*2} = J \quad \text{and} \quad a_1 > 0, \qquad (1)$$

$$c_0(f^*) = \int_{\mathcal{D}} f(t) \frac{dt}{(2\pi)^d} = 0 \qquad (2)$$

- (2) leads us to an asymptotically independent estimators,
- (1) leads us to an asymptotically efficient estimators (profile likelihood) in the gaussian case,

Construction of a registration method

A new criterion : alignment of an individual warped curve onto the mean of all the warped curves $\Phi = \Phi_\theta$: **parametric model** for deformations => **Semiparametric statistics**

$$\theta = (a, b, \upsilon)', \quad \Phi_\theta : f(\cdot) \rightarrow af(\cdot - b) + \upsilon$$

- For a **candidate** θ, compute the deformations

$$g_j(\theta, x) = \Phi_\theta^{-1} \circ f_j^\star(x) = \Phi_{\theta_j}^{-1} \circ f_j^\star(x)$$

- **Registration Criterion**

$$M_0(\theta) \rightarrow \frac{1}{J}\sum_{j=1}^{J}\left\| g_j(\theta, x) - \frac{1}{J}\sum_{j'=1}^{J} g_{j'}(\theta, x)\right\|_{L^2}^2,$$

M-estimators of the parameters $\hat{\theta}$ using en empirical version $M_n(\theta)$ of $M_0(\theta)$

Semi-parametric framework for translations

Assumption : $\Phi_\theta[f](t) = f(t - \theta)$

$$Y_{ij} = f^\star(t_{ij} - \theta_j^\star) + \sigma\epsilon_{ij}$$

Fourier Transform (DWT) => Equivalent Observation Model

$$d_{jl} = \exp(-il\alpha_j^\star)c_l(f) + w_{jl},$$

$$l = -\frac{n-1}{2}, \ldots, \frac{n-1}{2}, \; j = 1, \ldots, J.$$

$w_{jl} = w_{jl}^x + i w_{jl}^y, \quad w_{jl}^{\{x,y\}} \sim \mathcal{N}(0, \frac{\sigma^2}{n})$, i.i.d: observation noise
$c_l(f)$, $l \in \mathbb{Z}$: Fourier coefficients of f (unknown).
$\alpha_j^\star = \frac{2\pi}{T}\theta_j^\star$ and $\alpha^\star = (\alpha_j^\star)'_{j=1,\ldots,J_n}$ are warping parameters to estimate using d_{jl}.

Construction of Contrast

- Idea : **Aligning one shifted coefficient to mean of the others**

$$\tilde{d}_{jl}(\alpha) = \exp(il\alpha_j)d_{jl} = \exp(il[\alpha_j - \alpha_j^\star])c_l(f) + \exp(il\alpha_j)w_{jl}$$

$$\forall \alpha \in \mathbb{R}^{J_n},\ M_n(\alpha) = \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \frac{1}{J_n} \sum_{j=1}^{J_n} |\tilde{d}_{jl} - \frac{1}{J_n} \sum_{j=1}^{J_n} \tilde{d}_{jl}|^2.$$

- Smoothing Sequence δ_l, such that $\sum_l \delta_l^2 c_l^2 < +\infty$

$$\forall \alpha \in \mathbb{R}^{J_n},\ M_n(\alpha) = \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \frac{1}{J_n} \sum_{j=1}^{J_n} \delta_l^2 |\tilde{d}_{jl} - \frac{1}{J_n} \sum_{j=1}^{J_n} \tilde{d}_{jl}|^2.$$

M-estimation estimator

$$\hat{\alpha}_n = \arg \min_{\alpha \in [0,T]^{J_n} \cap \mathcal{A}_1} M_n(\alpha)$$

Convergence of the estimator

Assumptions

(I) Convergence : $(\delta_l c_l(f))_l \in \ell^2(\mathbb{Z})$ and $\sum_{l\neq 0} \delta_l^2 c_l^2(f) \neq 0$
very weak

(II) Asymptotic Normality : $\sum (\delta_l l)^2 c_l^2(f) < +\infty$,
$\sum (\delta_l l)^4 c_l^2(f) < +\infty$ et $\sum_{|l|<n/2} l^4 \delta_l^4 = o(n^2)$

Theorem (Gamboa-Loubes-Maza (2007))

Under Assumptions (I), $\hat{\alpha}_n$ converges in Probability to $\alpha^\star$.

Proof using **Standard Technics** in M-estimation theory :

- $M_n(\alpha) \xrightarrow{\mathbf{P}_{\alpha^\star}} K(\alpha, \alpha^\star)$ with minimum $\alpha = \alpha^\star$
- $\mathcal{A}$ is compact, $\alpha \to M_n(\alpha)$ and $\alpha \to K(\alpha, \alpha^\star)$ continuous.
- $\lim_n \mathbf{P}_{\alpha^\star}[\sup_{|\alpha-\beta|\leq \eta_k} |M_n(\alpha) - M_n(\beta)| > \epsilon_k] = 0.$

Asymptotic Normality

Theorem

Under Assumptions (II), the estimate $\hat{\alpha}_n$ *is such that*

$$\sqrt{n}(\hat{\alpha}_n - \alpha^\star) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma),$$

with $\Gamma = \frac{(J_n-1)\sum_{l\in\mathbb{Z}} l^2\delta_l^2}{J_n^6(J_n-2)^2(\sum_{l\in\mathbb{Z}}\delta_l l^2|c_l(f)|)^2} V(J).$

For $\delta_l \approx c_l(f)$, if $\sum_l l^2 c_l(f)^2 < +\infty$

$$\sqrt{n\sum_l l^2 c_l(f)^2}(\hat{\alpha}_n - \alpha^\star) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V(J))$$

$\rightarrow$ optimality of Warping Procedure with $\hat{\alpha}_n$.

Complete Model $\Phi_\theta : f(\cdot) \longrightarrow af(\cdot - b) + \upsilon$

$$d_{j0} = a_j^\star c_0(f^\star) + \upsilon_j^\star + w_{j0}$$

$$d_{jl} = a_j^\star \exp(-il\frac{2\pi}{T} b_j^\star) c_l(f^\star) + w_{jl}, \quad l \neq 0.$$

Inverting the operator for a candidate $\theta = (\alpha \quad b \quad \upsilon)'$

$$\tilde{d}_{jl}(\theta) = \frac{\exp(il\frac{2\pi}{T} b_j)}{a_j} d_{jl}, \; l \neq 0, \quad \tilde{d}_{j0}(\theta) = \frac{1}{a_j}(d_{j0} - \upsilon_j).$$

$$\widehat{\theta}_n = \arg\min_\theta M_n(\theta) \longrightarrow \theta^\star$$

Shape invariant curve estimator of $f^\star$

$$\widehat{f}(t_i) = \frac{1}{J} \sum_{j=1}^{J} \Phi_{\widehat{\theta}_j}^{-1}[Y_{ij}]$$

Other kind of deformations

Let v a vector field, v_j, $j = 1, \ldots, J$
ϕ_j is a diffeomorphism defined as a solution of PDE

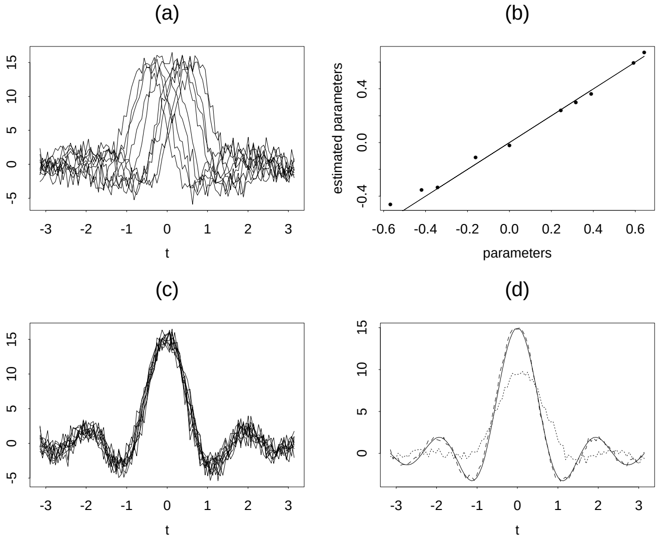
$$\frac{\partial}{\partial t}\phi_j = v_j \circ \phi$$

Model proposed by Trouvé and Youness for image transformations and used for curve warping

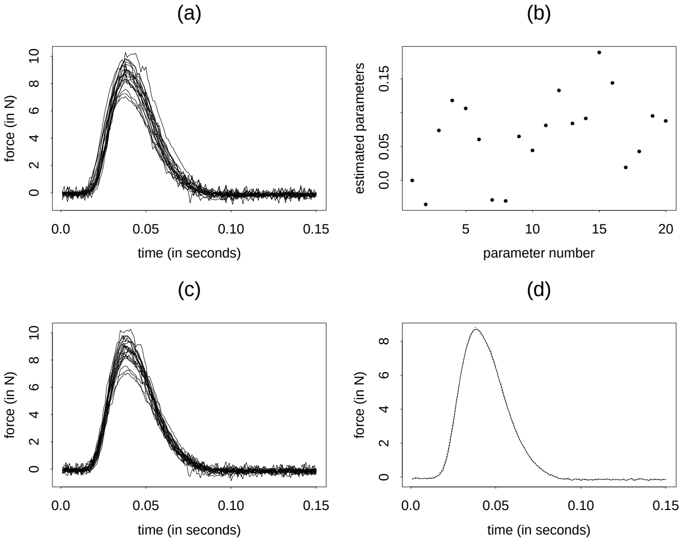
- Drawback : Fourier transform has not the same *nice* form but direct minimization of the quadratic criterion is still feasible ;
- Advantage : allows non linear warping families

Bigot, Gadat, Loubes (2011)

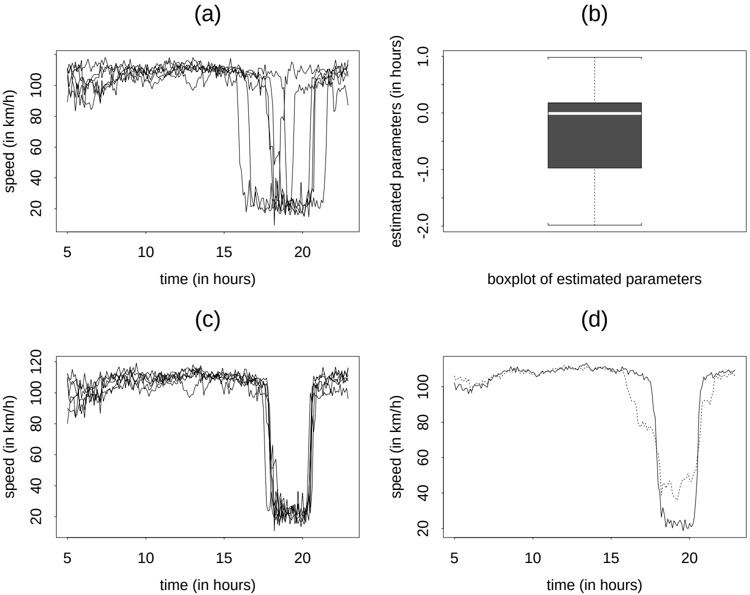
Simulations with sinusoidal functions



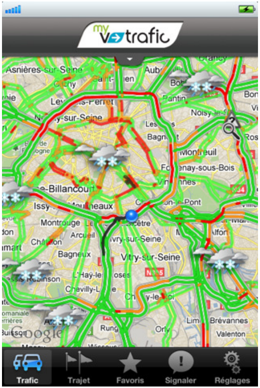
Pinch Force Data



Velocities of Cars

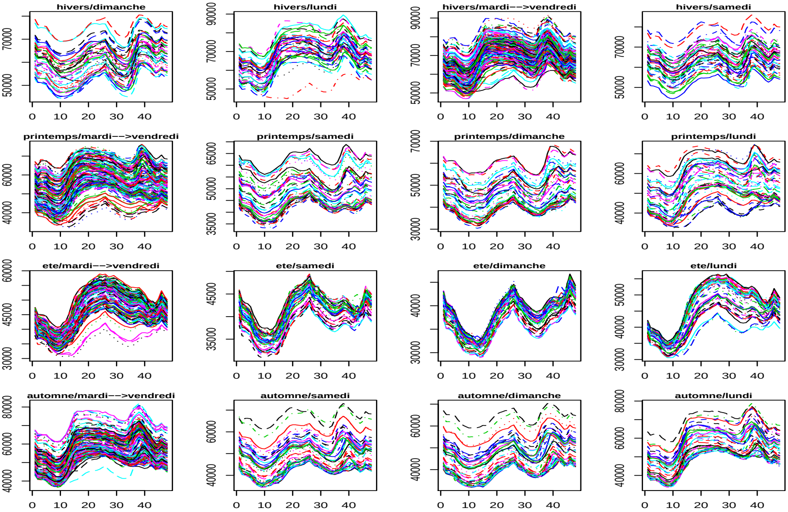


Velocities of Cars

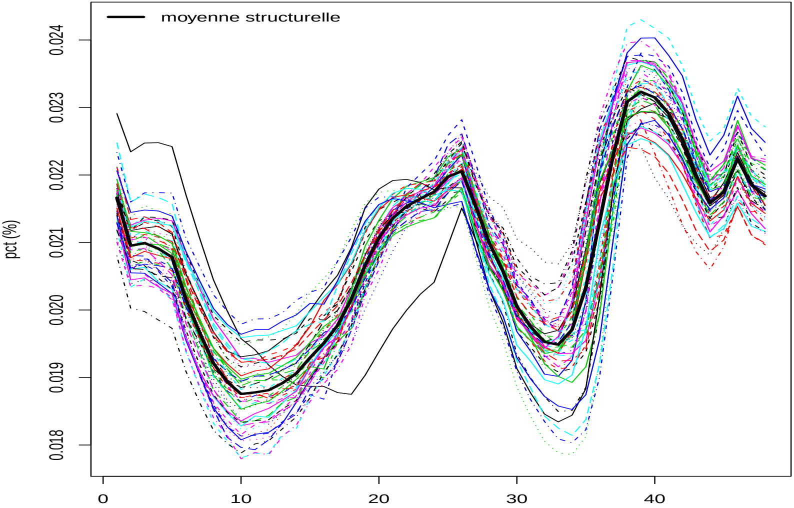


It works : good models of road trafficking behaviours joint work with **V-Trafic** i-phone application with Meteo as auxiliary variable for prediction

Clustering Load curves with warping distance



Registered Load curves



Conclusions : (partial)

- The **shape** represents a **mean** behaviour of a large number of individuals **sharing the same behaviour** with some fluctuations.
- Provides a **different distance** based on deformations
  - To aggregate the individual curves and provide a **structural mean curve**
  - To cluster the individuals with respect to the pseudo-distance to the pattern
  - To analyze the deformations with geometric PCA
- Based on sharp estimation of the deformations which relies on **semi-parametric** technics
- If other variables are available ... use them ! And obtain better results and enable to **predict** using auxiliary variables

Incorporate additional information $X \in \mathcal{X}$

Framework : the outcome depends on **exogenous information** which explain the individual deviations from the pattern.

$$Y_{ij} = \Phi_{\theta_j^\star}[f^\star](t_i) + \sigma\varepsilon_{ij} = F(X_j, t_i).$$

X_j are observed parameters which characterize the behaviour of the individual j

- **Objective** : model the relationship and estimate it

$$\theta \quad : \quad X_j \longrightarrow \theta(X_j) = \theta_j^\star$$

Forecast with shape invariant model

$$\widehat{f_j} = \Phi_{\hat{\theta}(X_j)}[\hat{f}]$$

Example : confidence index for income analysis, characteristics of the observation day (holidays, week end, standard) for vehicle speed, type of days or temperature evolution for electric load.

Building a surface response

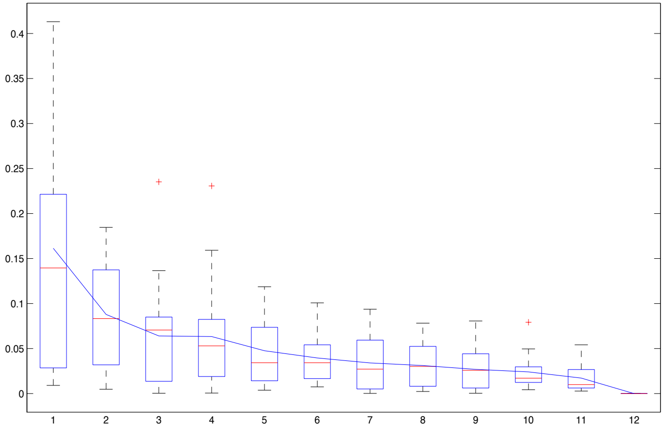
1. Data : X_j, $j = 1, \ldots, J$ and the outcome Y_{ij}
2. Estimation of the parameters $\widehat{\theta}_j \in \mathbb{R}^3$ and $\widehat{f}$ of the shape $f^\star$
3. $\widehat{\theta}(X_j)$ is viewed as a sample of a random process $\widehat{\theta}(.)$ observed at random locations. $X_j, \quad j = 1, \ldots, J$
   $\Rightarrow$ Prediction using standard **Kriging method** choosing a proper covariance structure K.

$$x_o \notin \{X_1, \ldots, X_j\} \quad x_o \mapsto \tilde{\theta}(x_0)$$

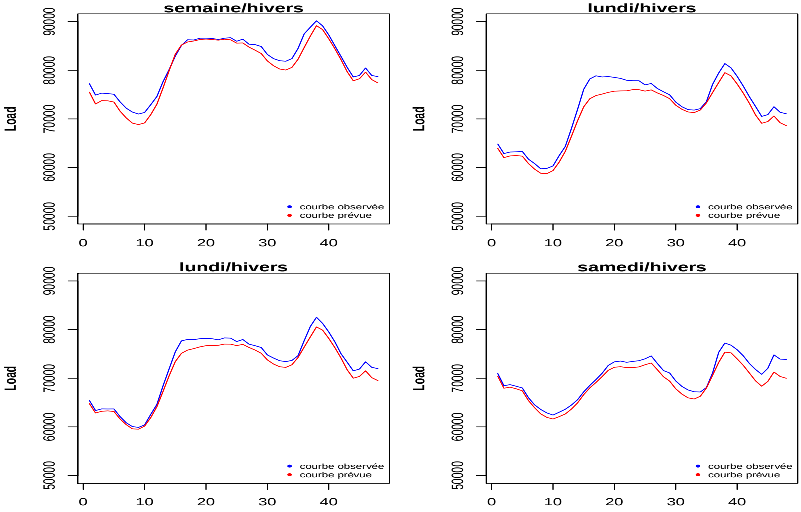
4. Forecast : the predicted shape built by the deformations induced by the parameters x_0.

$$\widehat{f}_0 = \Phi_{\tilde{\theta}(x_0)}[\widehat{f}]$$

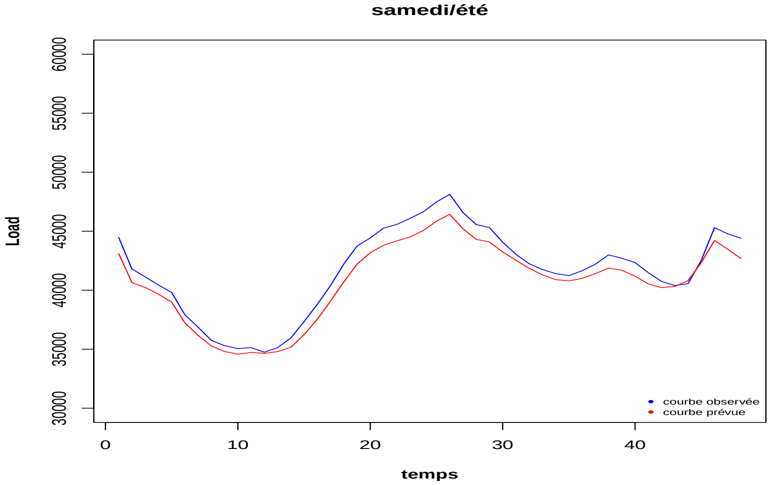
Forecasting Sales



Forecasting Load curves $J+1$



Forecasting Load curves $J + 1$



Remarks on the forecast

- Semiparametric framework enables efficient and sharp estimation of the underlying $\widehat{\theta}(.)$.
  Necessary since the prediction are based on a good approximation result.
- Choice of the Kernel (covariance) induces a *distance* on the covariates X
  For curves (temperature), the kernel must handle functional distance and may depend on a tuning parameter (see work by Sapatinas 2012).
- Sensitivity Analysis to analyze the influence of the parameters X.
- The deformations should be local ... for the moment work on parts of curves but work still in progress
- If scale issues : transform the problem into forecast of the daily distribution of consumption (deformation of densities) and the daily consumption (regression)