

Méthodes pour l'analyse des données fonctionnelles

Cristian PREDA

Polytech' Lille

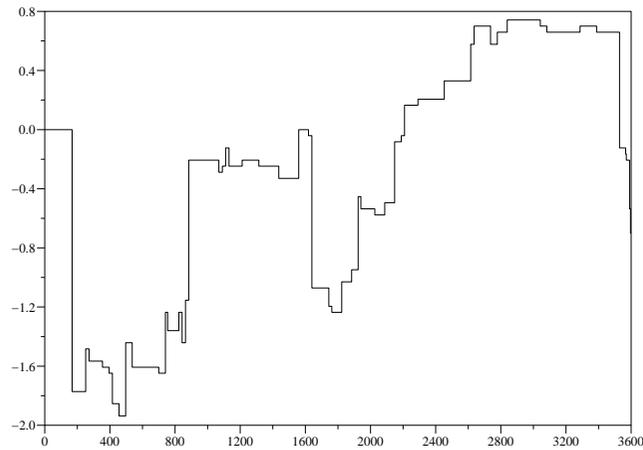
Université des Sciences et Technologies de Lille, France

Google :

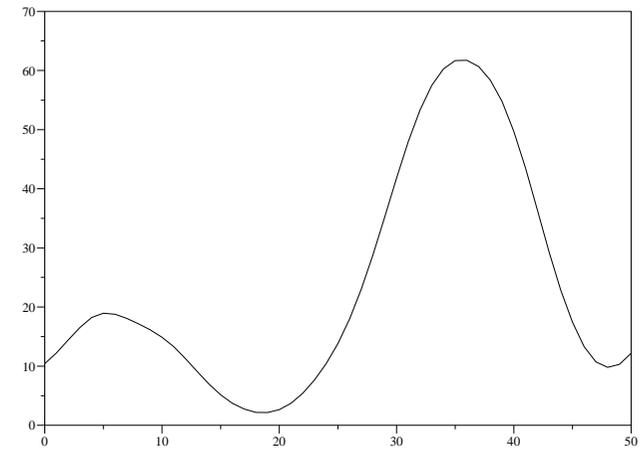
- chemometrics : 1050000
- functional+ data : 70900000
- functional + data + chemometrics : 17300

Données fonctionnelles

Exemples :

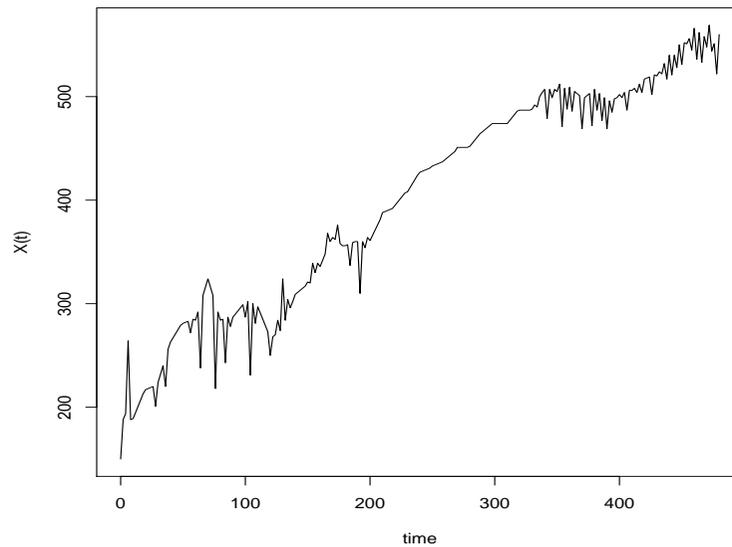


Indice d'action boursière

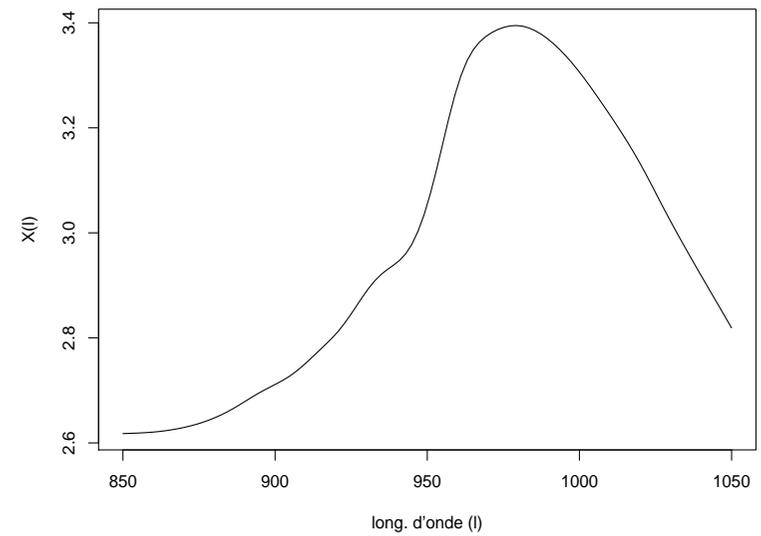


Marche : angle du genou

Données spectrométriques



Pétrissage: résistance de la pâte



Données spectrométriques

Variable fonctionnelle et donnée fonctionnelle

Définition [Ferraty et Vieu (2006)] : *Une variable aléatoire est dite fonctionnelle si ses valeurs sont dans un espace de dimension infinie. Une observation d'une variable fonctionnelle est appelée donnée fonctionnelle*

$$\mathbf{X} = \{X_t : t \in \mathcal{T}\},$$
$$X_t : \Omega \rightarrow \mathbb{R}$$

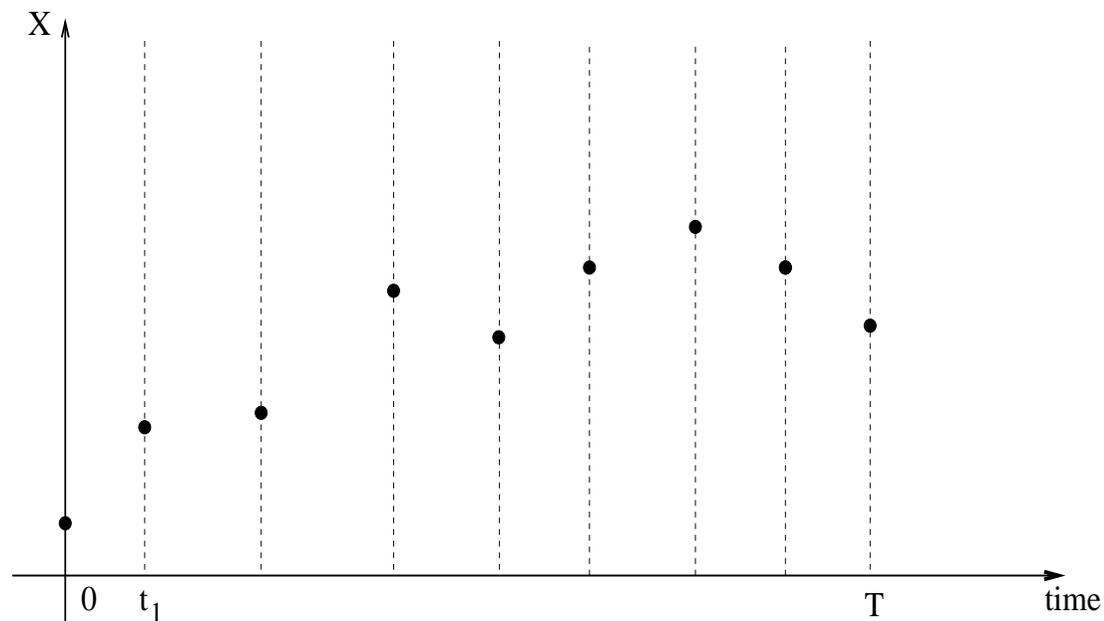
- $\mathcal{T} \subseteq \mathbb{R}$: courbe
- $\mathcal{T} \subseteq \mathbb{R}^2$: image

“Tableau” des données :

En pratique, une donnée fonctionnelle est observée dans un nombre fini d’instants :

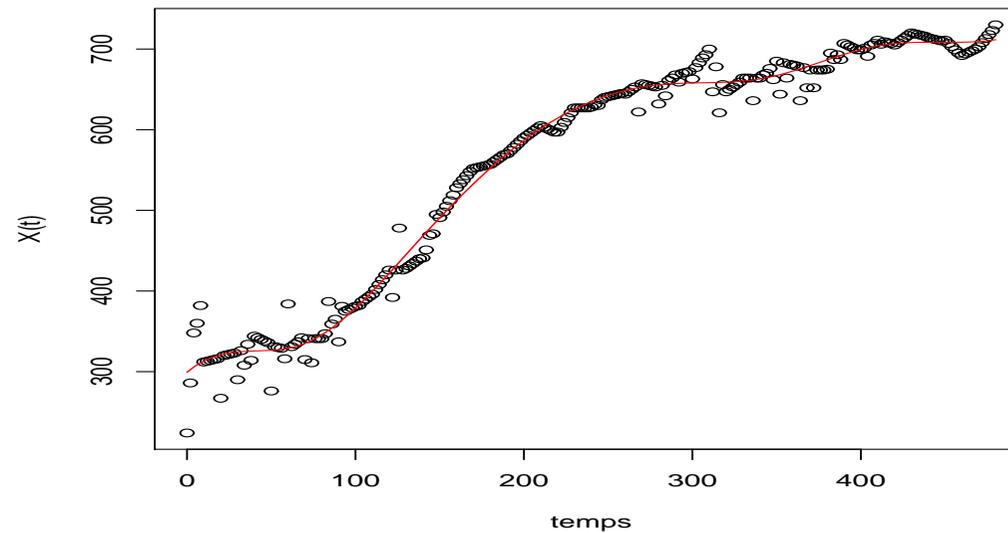
$$0 = t_0 < t_1 < \dots < t_k = T :$$

$$\{(t_0, X_{t_0}), (t_1, X_{t_1}), \dots, (t_k, X_{t_k})\}.$$



Caractère “fonctionnel” de la courbe :

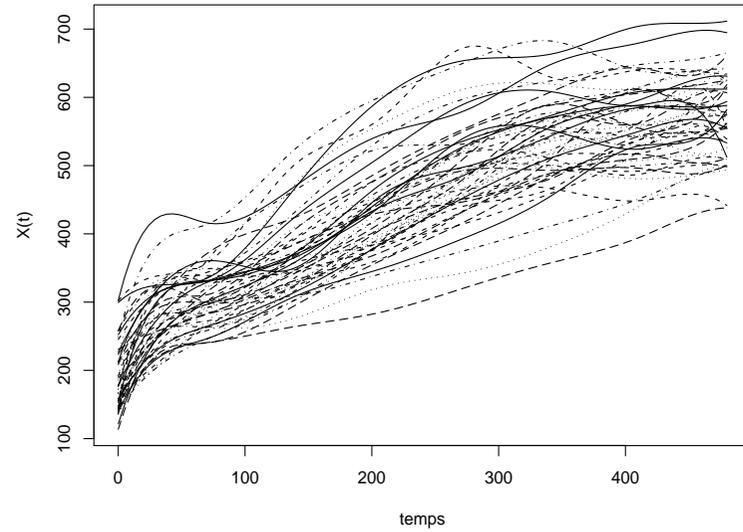
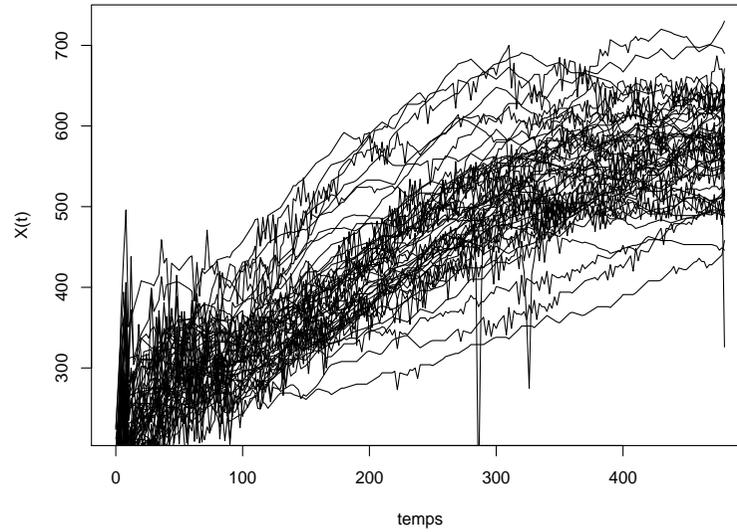
- interpolation : linéaire, spline, ...
- lissage si les données sont observées avec erreurs



Lissage utilisant une base de fonctions (B-splines cubiques)

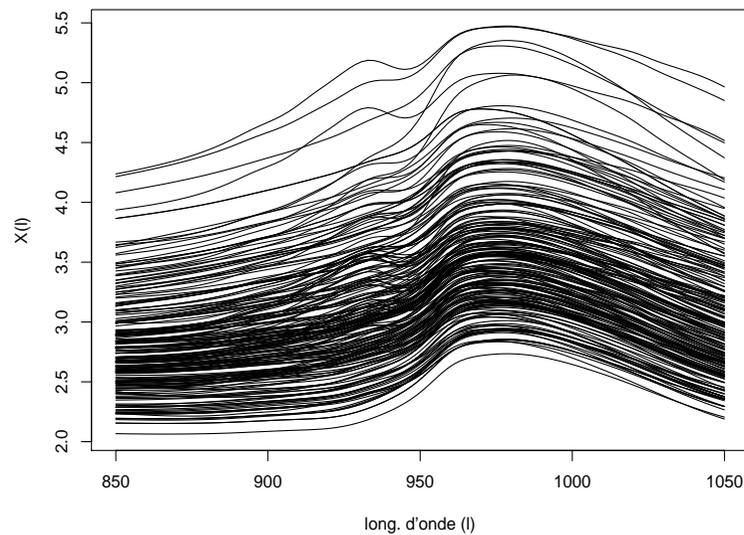
Exemples :

- lissage :



- interpolation :

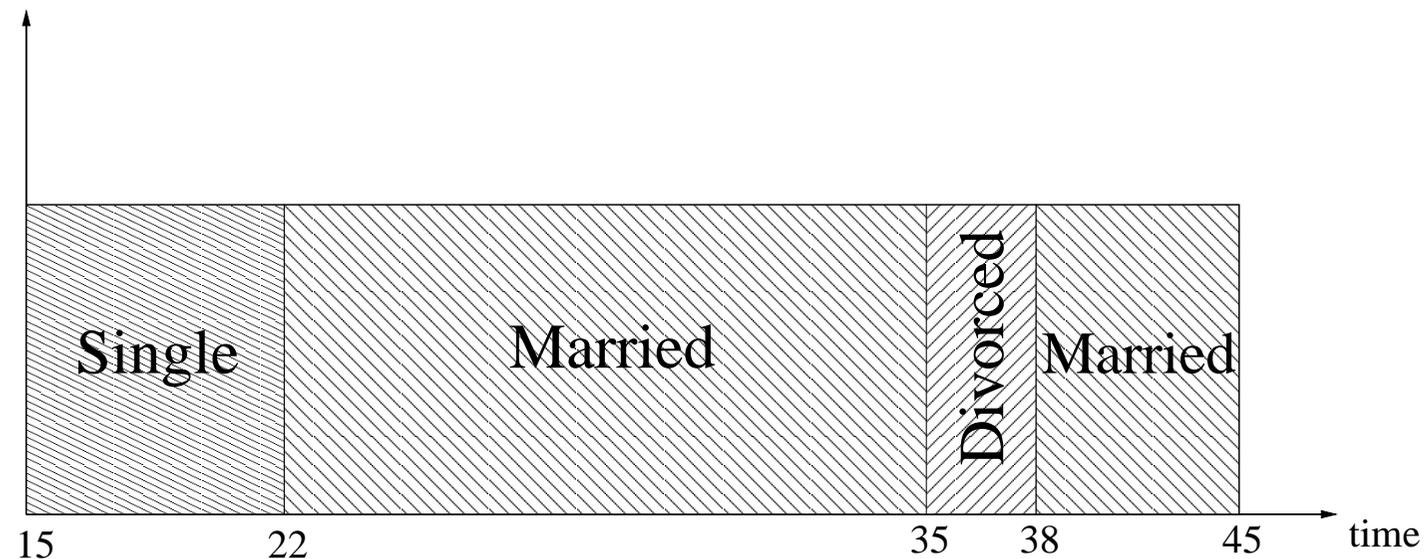
Données spectrométriques



Données fonctionnelles qualitatives

Deville (1982)

$$X_t \in \{\text{"Single"}, \text{"Married"}, \text{"Divorced"}, \text{"Widowed"}\}$$



Evolution de l'état civil des femmes de 15 à 45 ans.

- $\mathbf{E} = \{1, 2, \dots, m\}$ = espace d'états,
- $(\Omega, \mathcal{A}, P), \mathbf{X} = \{X_t : t \in [0, T]\}, X_t : \Omega \rightarrow \mathbf{E}$ (processus de sauts).

Méthodes pour les données fonctionnelles

Deville (1974)

Saporta (1981)

Ramsay et Silverman (1997, 2002, 2005)

Ferraty et Vieu (2006)

- Analyses factorielles
- Modèles de régression
- Classification

Exemples :

Données	Vecteur de dimension finie $X = (X_1, X_2, \dots, X_p)$	Données fonctionnelles $X = \{X_t\}_{t \in [0, T]}$
ACP	Matrice de cov. $V : \in M_{p \times p}(\mathbb{R})$ f.p. : $u = (u_1, \dots, u_p) \in \mathbb{R}^p$: $Vu = \lambda u$ c.p. : $\xi = u_1 X_1 + \dots + u_p X_p$	Opérateur de cov. C de noyau $C(t, s)$ f.p. : $u : \mathcal{T} \rightarrow \mathbb{R}$: $\int_{\mathcal{T}} C(t, s) u(s) ds = u(t)$ c.p. : $\xi = \int_{\mathcal{T}} u(t) X_t dt$
Régression linéaire	$\mathbb{E}(Y X = (x_1, \dots, x_p)) :$ $\beta_0 + \sum_{i=1}^p \beta_i x_i$	$\mathbb{E}(Y X = \{x(t), t \in \mathcal{T}\}) :$ $\beta_0 + \int_{\mathcal{T}} \beta(t) x(t) dt$

ACP des données fonctionnelles

$\{X_t : t \in [0, T]\}$:

- $E(X_t^2) < \infty$,
- L_2 -continu,
- $\forall \omega \in \Omega : (X_t(\omega))_{t \in [0, T]} \in L_2([0, T])$,
- $E(X_t) = 0, \forall t \in [0, T]$.

$\mathbf{C} : L_2([0, T]) \rightarrow L_2([0, T])$,

$$f \xrightarrow{\mathbf{C}} g, \quad g(t) = \int_0^T C(t, s) f(s) ds, \quad \forall t \in [0, T],$$

$C(t, s) = \text{Cov}(X_t X_s), \forall s, t \in [0, T]$.

\mathbf{C} : autoadjoint, positif, compact et de trace finie.

$\{\lambda_1 \geq \lambda_2 \geq \dots\}, \{f_i\}_{i \geq 1} :$

$$\mathbf{C}f_i = \lambda_i f_i,$$

- $\lambda_i \geq 0, \quad \lim_{i \rightarrow \infty} \lambda_i = 0, \quad \sum_{i=1}^{\infty} \lambda_i < \infty$
- $(f_i | f_j) = \delta_{i,j}.$

$$\xi_i = \int_0^T f_i(t) X_t dt \quad \forall i \geq 1.$$

$\{\xi_i\}_{i \geq 1} : \text{Var}(\xi_i) = \lambda_i, \text{Cov}(\xi_i, \xi_j) = 0, i \neq j.$

$\mathbf{W} : L_2(\Omega) \rightarrow L_2(\Omega),$

$$\mathbf{W}z = \int_0^T X_t E(X_t z) dt, \quad \forall z \in L_2(\Omega).$$

$$\mathbf{W}\xi_i = \lambda_i \xi_i, \quad \forall i \geq 1.$$

Reconstitution des données en ACP ou décomposition de Karhunen-Loève :

$$X_t = \sum_{i \geq 1} f_i(t) \xi_i, \quad \forall t \in [0, T].$$

Approximation d'ordre K ($K \geq 1$) :

$$X_t^{(K)} = \sum_{i=1}^K f_i(t) \xi_i, \quad \forall t \in [0, T].$$

Loève (1945) :

$X^{(K)}$ est la meilleure approximation de X en norme L_2 .

$$V_{(X_t)} = \int_0^T E(X_t^2) dt = \text{Tr}(\mathbf{C}) = \sum_{i \geq 1} \lambda_i.$$

$$I_K = \frac{V_{(X_t^{(K)})}}{V_{(X_t)}} = \frac{\sum_{i=1}^K \lambda_i}{\sum_{i \geq 1} \lambda_i} \leq 1.$$

Exemple :

$X = \{X_t, t \in [0, T]\}$: processus à accroissements stationnaires non corrélés (Poisson, le mouvement brownien)

$$C(t, s) = c \times \min(t, s), \quad c > 0.$$

$$- \lambda_k = \frac{4cT^2}{(2k-1)^2\pi^2}, \quad k = 1, \dots, \infty.$$

$$- f_k(t) = \sqrt{\frac{2}{T}} \sin\left(\left(k - \frac{1}{2}\right)\pi \frac{t}{T}\right)$$

$$- I_1 = 81.1\%$$

$$- I_2 = 90.1\%$$

$$- I_3 = 93.3\%$$

$$- I_4 = 95.0\%$$

$$X_t \approx \xi_1 f_1(t) = \xi_1 \sqrt{\frac{2}{T}} \sin\left(\frac{\pi t}{2T}\right)$$

Estimation et approximation de l'ACP

- Estimation

Deville '74, Dauxois et al. '82

$\{X_1, X_2, \dots, X_N\}$

$$\hat{C}(t, s) = \frac{1}{N} \sum_{i=1}^N X_i(t)X_i(s), \quad \forall (s, t) \in [0, T] \times [0, T],$$

$$(\hat{C}f)(t) = \int_0^T \hat{C}(t, s)f(s)ds, \quad \forall f \in L_2([0, T]).$$

- $\lim_{N \rightarrow \infty} E(\|\hat{C} - C\|^2) = 0,$
- $\lim_{N \rightarrow \infty} E(\hat{\lambda}_i - \lambda_i)^2 = 0 \text{ (} o(N^{-\frac{1}{2}} \text{))},$
- $\lim_{N \rightarrow \infty} E(\|\hat{f}_i - f_i\|^2) = 0 \text{ (} o(N^{-\frac{1}{2}} \text{))}.$

- Interpolation

$\{0 = t_1 < t_2 < \dots < t_p = T\}, \{X_{t_0}, X_{t_2}, \dots, X_{t_p}\}.$

(Deville '74 et Besse '79) : $\text{ACP} \{X_{t_0}, X_{t_2}, \dots, X_{t_p}\} \not\approx \text{ACP} (X_t)_{t \in [0, T]}.$

Interpolation linéaire :

$$IX_t = \sum_{i=0}^p c_i \phi_i(t), \quad \forall t \in [0, T],$$

$\{\phi_i\}_{i=0, \dots, p}$, est une base dans un espace de fonctions et $\{c_i\}_{i=0..p}$ v.a.r. :

$$IX_{t_i} = X_{t_i} \text{ p.s.}$$

$\text{ACP} (IX_t)_{t \in [0, T]} \equiv$ l'ACP $\{X_{t_0}, X_{t_1}, \dots, X_{t_p}\}$ avec pour métrique $M = \Phi\Phi'$, où $\Phi = A^{-1}$, $A = (a_{i,j})_{0 \leq i, j \leq p}$, $a_{i,j} = \langle \phi_i, \phi_j \rangle.$

Le choix des fonctions d'interpolation est équivalent à celui d'une métrique.

Deville '74 : constantes par morceaux,

Pardoux '82 : splines linéaires,

Aguilera '93 : B-splines,

Besse '89 : splines,

- Lissage

$\{\phi_1, \dots, \phi_p\}$ dans $L_2([0, T])$, lin. indep.

$$X_t \approx \sum_{i=1}^p \alpha_i \phi(t).$$

ACP de X est équivalente à l'ACP de $\{\alpha_1, \dots, \alpha_p\}$ avec la métrique $\Phi = \{\langle \phi_i, \phi_j \rangle\}$.

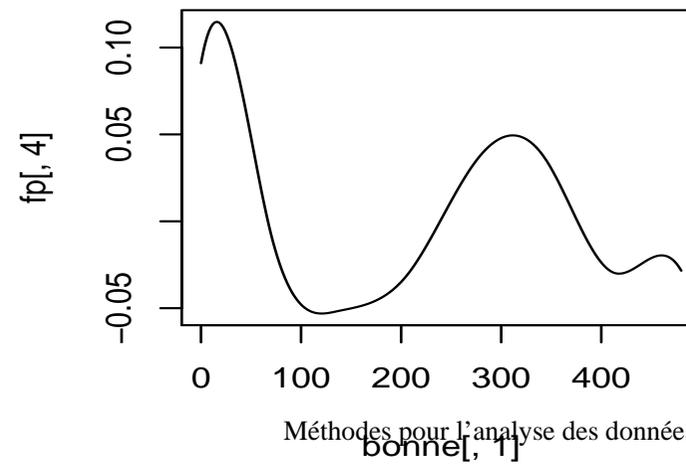
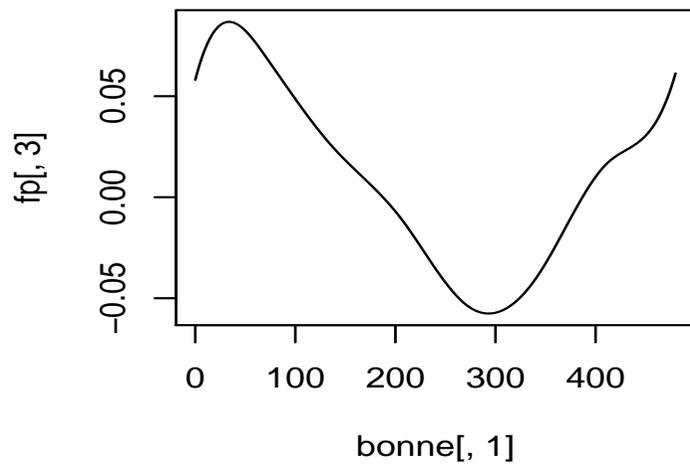
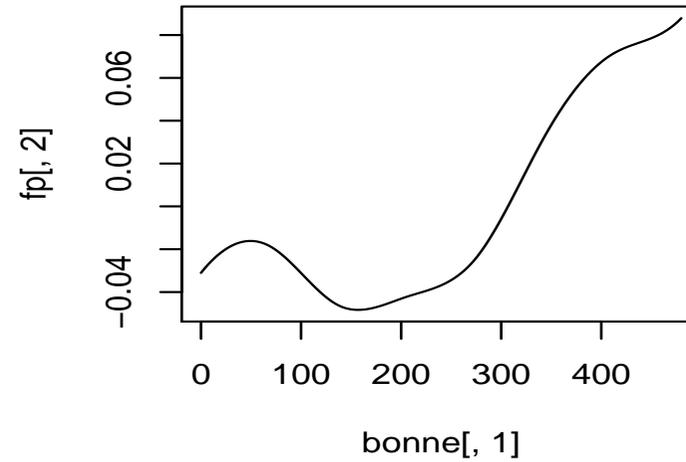
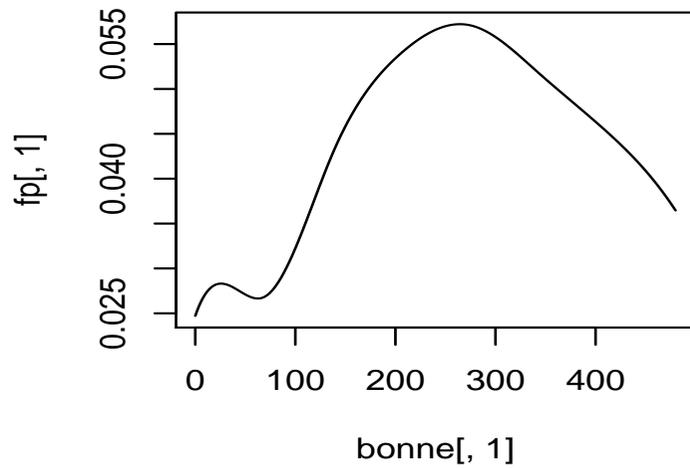
Implémentation R

Le package **fda**

Quelques fonctions utiles :

- création d'une base de fonctions :
`create.bspline.basis`
- création d'un objet "donnée fonctionnelle" : `data2fd`
- réalisation de l'ACP fonctionnelle : `pca.fd`
- régression linéaire avec prédicteur fonctionnel sur les composantes principales : `pcr`, `pls`

Exemple. Les facteurs principaux : farines de Danone.



Régression sur données fonctionnelles

$$Y = f(X) + \varepsilon$$

- Y :
- scalaire (\mathbb{R})
 - vectorielle scalaire (\mathbb{R}^p)
 - fonctionnelle ($L_2([0, T])$)
 - qualitative

Modélisation paramétrique et non-paramétrique.

Le modèle linéaire

f - linéaire : théorème de représentation de Riesz :

$$f(X) = \langle X, \beta \rangle_{L_2([0,T])} = \int_0^T X_t \beta(t) dt, \quad \beta \in L_2([0,T]).$$

Moindres carrés :

$$\mathbb{E}(Y X_t) = \int_0^T \mathbb{E}(X_t X_s) \beta(s) ds \quad (W - H)$$

En général, c'est un problème mal posé car

Théorème (Picard)

$$\beta \text{ unique} \iff \sum_{i \geq 1} \frac{c_i^2}{\lambda_i^2} < \infty,$$

$$c_i = \int_0^T \mathbb{E}(X_t Y) f_i(t) dt, \quad \{\lambda_i\}_{i \geq 1}, \{f_i\}_{i \geq 1} \text{ (ACP)} \quad \beta(t) = \sum_{i \geq 1} \frac{c_i}{\lambda_i} f_i(t), \quad t \in [0, T].$$

Exemple : Le processus X est un processus de Poisson sur $[0, T]$ et $Y = X_{T+h}$, $h > 0$.

Alors,

$$\sum_{i \geq 1} \frac{c_i^2}{\lambda_i^2} = \sum_{i \geq 1} \frac{2}{T} = \infty,$$

$\hat{Y} = X_T$ et β est la distribution de Dirac en T , δ_T .

Régularisation des moindres carrés

- *réduction de la dimension* : $X \approx X_S, S = \{\phi_1, \dots, \phi_p\}$

Régression sur les composantes principales (PCR) :

$S = \{f_1, \dots, f_p\}$ - facteurs principaux de l'ACP de X

(Deville (1978), Aguilera (1998), Cardot et al. (1999))

- *pénalisation des moindres carrés* :

$$\hat{\beta} = \arg \min_{\beta \in \overline{S}} \frac{1}{n} \sum_{i=1}^n (Y_i - \langle X_i, \beta \rangle)^2 + \lambda \|\beta^{(m)}\|^2$$

(Ramsay et Silverman (1997), Cardot et al. (2003))

- *les moindres carrés partiels (PLS)*.

$$\hat{\beta}_{PLS} = \arg \min_{\substack{\beta \in L_2([0, T]) \\ \|\beta\| = 1}} \underbrace{\frac{1}{n} \sum_{i=1}^n (Y_i - \langle X_i, \beta \rangle)^2}_{\text{}} \times \underbrace{\frac{1}{n} \langle X_i, \beta \rangle^2}_{\text{}}$$

L'approche PLS pour données fonctionnelles

$$X = \{X_t\}_{t \in [0, T]}, Y = (Y_1, \dots, Y_p), p \geq 1.$$

$\{t_i\}_{i \geq 1}$ – variables aléatoires non-corrélées dans $L_2(X)$

Théorème [Critère de Tucker]

$$\max_{w, c} \text{Cov}^2 \left(\int_0^T X_t w(t) dt, \sum_{i=1}^p c_i Y_i \right)$$

$$w \in L_2([0, T]), \|w\| = 1$$

$$c \in \mathbf{R}^p, \|c\| = 1$$

est réalisé pour w et c , les vecteurs propres associés à la plus grande valeur propre des opérateurs \mathbf{U}_X , respectivement \mathbf{U}_Y , où $\mathbf{U}_X = \mathbf{C}_{XY} \circ \mathbf{C}_{YX}$ et $\mathbf{U}_Y = \mathbf{C}_{YX} \circ \mathbf{C}_{XY}$.

Les composantes PLS

Soit $w_1 \in L_2([0, T])$: $\mathbf{U}_X w_1 = \lambda_{max} w_1$,

$$t_1 = \int_0^T X_t w_1(t) dt$$

Proposition.

$$\mathbf{W}^X \mathbf{W}^Y t_1 = \lambda_{max} t_1$$

Itération PLS :

Let $X_{0,t} = X_t, \forall t \in [0, T]$ and $Y_{0,i} = Y_i, \forall i = 1, \dots, p$.

Au pas $h, h \geq 1$: $\mathbf{W}_{h-1}^X \mathbf{W}_{h-1}^Y t_h = \lambda_{max}^{(h)} t_h$,

Régression linéaire simple sur t_h :

$$X_{h,t} = X_{h-1,t} - p_h(t) t_h, \quad t \in [0, T],$$

$$Y_{h,i} = Y_{h-1,i} - c_{h,i} t_h, \quad i = 1, \dots, p.$$

$\{t_h\}_{h \geq 1}$ - variables aléatoires

$\{p_h\}_{h \geq 1}$ - fonctions de $L_2([0, T])$

$\{c_h\}_{h \geq 1}$ - vecteurs de \mathbb{R}^p .

Théorème [Décomposition PLS] Pour tout $h \geq 1$:

a) $\{t_i\}_{1 \leq i \leq h}$ – système orthogonal dans $L_2(X)$,

$$b) \mathbf{Y} = \underbrace{c_1 t_1 + c_2 t_2 + \dots + c_h t_h}_{\hat{\mathbf{Y}}_{PLS(h)}} + \mathbf{Y}_h = \underbrace{\int_0^T X_t \beta_{PLS(h)}(t) dt}_{\hat{\mathbf{Y}}_{PLS(h)}} + \mathbf{Y}_h,$$

c) $X_t = p_1(t)t_1 + p_2(t)t_2 + \dots + p_h(t)t_h + X_{h,t}$,

d) $\mathbb{E}(Y_{h,j}t_i) = 0, \forall i = 1, \dots, h, \forall j = 1, \dots, p$,

e) $\mathbb{E}(X_{h,t}t_i) = 0, \forall t \in [0, T], \forall i = 1, \dots, h$.

Théorème [Convergence] $\lim_{h \rightarrow \infty} \mathbb{E} \left(\|\hat{\mathbf{Y}}_{PLS(h)} - \hat{\mathbf{Y}}\|^2 \right) = 0.$

Théorème [PLS vs PCR] L'ajustement à l'aide des premières h composantes PLS est meilleur que celui avec les premières h composantes principales,

$$R^2(Y, \hat{Y}_{PCR(h)}) \leq R^2(Y, \hat{Y}_{PLS(h)}).$$

Remarque :

- analogie avec l'analyse canonique
- prise en compte des relations existant entre les variables $\{Y_i\}_{i=1,\dots,p}$.
- réponse fonctionnelle : $\mathbf{Y} = \{Y_t\}_{t \in [0, T']}$.
- choix de h en pratique : validation croisée.

PLS et l'analyse discriminante linéaire fonctionnelle

$$X = \{X_t\}_{t \in [0, T]}, \quad X_t : \Omega \rightarrow \mathbb{R},$$

$$Y : \Omega \rightarrow \{1, \dots, K\}, \quad K \geq 1,$$

$$- \mathbb{P}(Y = i) \neq 0, \quad \forall i \in \{1, \dots, K\}$$

Problème :

Déterminer des scores discriminantes linéaires,

$$\Phi(X) = \int_0^T X_t \beta(t) dt, \quad \beta \in L_2[0, T],$$

tels que

$$\max_{\beta \in L_2[0, T]} \frac{\mathbb{V}(\mathbb{E}(\Phi(X)|Y))}{\mathbb{V}(\Phi(X))}$$

(James et Hastie (2001), Ratcliffe et al. (2002), Ferraty et Vieu (2003), Biau et al. (2005))

Réponse binaire : un problème mal posé

$Y \in \{0, 1\}$:

$$\begin{cases} p_0 = \mathbb{P}(Y = 0), & p_1 = 1 - p_0 = \mathbb{P}(Y = 1), \\ \mu_0(t) = \mathbb{E}(X_t|Y = 0), & \mu_1(t) = \mathbb{E}(X_t|Y = 1), t \in [0, 1]. \end{cases}$$

Soient

- \mathbf{C} : l'opérateur de covariance de X
- \mathbf{B} : l'opérateur de covariance interclasses définit par

$$f \xrightarrow{\mathbf{B}} g, \quad g(t) = \int_0^T B(t, s) f(s) ds,$$

où $B(t, s) = p_0 p_1 (\mu_0(t) - \mu_1(t)) (\mu_0(s) - \mu_1(s)) = \phi \otimes \phi$,
avec $\phi = \sqrt{p_0 p_1} (\mu_0 - \mu_1)$.

Alors,

La fonction discriminante, β , vérifie :

$$\mathbf{B}\beta = \lambda \mathbf{C}\beta, \quad \text{avec } \langle \beta, \mathbf{C}\beta \rangle_{L_2[0,T]} = 1.$$

$$\text{Recoder } Y \text{ par : } \begin{cases} 0 \rightsquigarrow \sqrt{\frac{p_1}{p_0}}, \\ 1 \rightsquigarrow -\sqrt{\frac{p_0}{p_1}} \end{cases}$$

Alors, β est solution de l'équation Wiener-Hopf

$$\mathbb{E}(Y X_t) = \int_0^T \mathbb{E}(X_t X_s) \beta(s) ds \quad (\text{W-H})$$

En pratique

$$Y \in \{0, 1\} \underset{\text{recodage}}{\Longleftrightarrow} Y \in \left\{ \sqrt{\frac{p_1}{p_0}}, -\sqrt{\frac{p_0}{p_1}} \right\}.$$

Régression PLS de Y sur $X = \{X_t\}_{t \in [0, T]}$:

- fonction discriminante : $\beta_{PLS(h)}$.

- score discriminant : $\Phi(X) = \int_0^T X_t \beta_{PLS(q)}(t) dt$.

Prédiction :

si $\Phi(X) \leq 0$ alors $Y = 1$ sinon $Y = 0$.

Réponse multiple : $K > 2$

$$Y \in \{1, \dots, K\}.$$

Soit $\mathbf{Y} = \{Y_1, \dots, Y_{K-1}\}$ le vecteur aléatoire dont les composantes sont les variables indicatrices des modalités $\{1, \dots, K - 1\}$.

Méthodologie basée sur l'analogie entre l'analyse canonique et analyse discriminante :

- régression PLS de \mathbf{Y} sur $X = \{X_t\}_{t \in [0, T]}$:
- déterminer les composantes PLS $\{t_1, \dots, t_h\}$,
- réaliser l'analyse linéaire discriminante classique de Y sur les composantes PLS.

Modélisation non-paramétrique

$$Y = f(X) + \varepsilon.$$

$$\hat{f}(x) = \frac{\sum_{i=1}^n Y_i K(d(x, X_i)/h)}{\sum_{i=1}^n K(d(x, X_i)/h)},$$

(Ferraty et al. (2006) ! Convergence : $\phi_x(h) = \mathbb{P}(d(x, X) < h)$).

Estimation non-paramétrique de la fonction de répartition conditionnelle et du mode conditionnel.

Précision des vitesses de convergences.

f : minimiser le risque empirique :

$$R_{emp}[f] = \frac{1}{n} \sum_{i=1}^n \mathcal{C}(X_i, Y_i, f(X_i))$$

Fonction de perte :

$$\mathcal{C}(x, y, f(x)) = \begin{cases} (y - f(x))^2 & \text{- quadratique} \\ \max\{0, |f(x) - y| - \varepsilon\} & \text{- } \varepsilon \text{- insensible} \\ -yf(x) + \log(1 + e^{f(x)}) & \text{- logistique (} Y \in \{0, 1\}\text{)} \end{cases}$$

Risque régularisé

$$R_{reg}[f] = \frac{1}{n} \sum_{i=1}^n \mathcal{C}(x_i, y_i, f(x_i)) + \lambda \Omega[f],$$

où $\lambda > 0$ est un paramètre de régularisation et $\Omega[f]$ un terme de stabilisation.

$(\mathcal{F}, \|\cdot\|)$,

$$R_{reg}[f] = \frac{1}{n} \sum_{i=1}^n \mathcal{C}(x_i, y_i, f(x_i)) + \underbrace{\lambda \|f\|_{\mathcal{F}}^2}_{\text{pénalité}}.$$

$Im(X) \in H$ ($H = L_2([0, T])$)

$\mathcal{F} = \mathcal{H}_{\mathbf{K}}$:

$\mathbf{K} : H \times H \rightarrow \mathbb{R}$, symétrique et définie positive : $\mathcal{H}_{\mathbf{K}} = \overline{\left\{ \sum_{i=1}^n a_i \mathbf{K}(t_i, \cdot), t_i \in H, a_i \in \mathbb{R}, n \geq 1 \right\}}$

$$\mathbf{K}(x, y) = \begin{cases} \langle x, y \rangle_H^d, & d \in \mathbb{N} & \text{polynomial} \\ (c + \langle x, y \rangle_H)^d, & d \in \mathbb{N}, c > 0 & \text{polynomial non-homogène} \\ e^{-\frac{\|x - y\|_H^2}{2\sigma^2}}, & \sigma > 0, & \text{gaussien} \end{cases}$$

Régression dans RKHS

Théorème [Représentation] Soit $\{(x_i, y_i)\}_{i=1}^n$, $x_i \in H$, $y_i \in \mathbb{R}$, $n > 0$, $n \in \mathbb{N}$, $\lambda > 0$ et $\mathcal{C} : H \times \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ une fonction de perte convexe. Alors, la solution du problème : trouver $\hat{f} \in \mathcal{H}_{\mathbf{K}}$ qui minimise

$$\frac{1}{n} \sum_{i=1}^n \mathcal{C} \left(x_i, y_i, \hat{f}(x_i) \right) + \lambda \|\hat{f}\|_{\mathcal{H}_{\mathbf{K}}}^2$$

existe, est unique et admet la représentation

$$\hat{f}(x) = \sum_{i=1}^n \alpha_i \mathbf{K}(x, x_i), \quad \forall x \in H,$$

où $\alpha_i \in \mathbb{R}$, $i = 1, \dots, n$.

Exemple.

$$\mathcal{C}(x, y, f(x)) = (y - f(x))^2,$$

$$\{(x_i, y_i)\}_{i=1}^n,$$

$$(\lambda n \mathbf{I} + [\mathbf{K}])\alpha = y,$$

où $\alpha = (\alpha_1, \dots, \alpha_n)'$, $y = (y_1, \dots, y_n)'$, $[\mathbf{K}] \in \mathcal{M}_n(\mathbb{R})$,
 $[\mathbf{K}]_{i,j} = \mathbf{K}(x_i, x_j)$, $1 \leq i, j \leq n$.

Perte logistique, ε - insensitive : méthode du gradient.

Consistance dans RKHS

Théorème [Consistance] Soit \mathcal{C} la fonction de perte quadratique ou logistique et supposons qu'il existe $M_Y > 0$ tel que $|Y| \leq M_Y$ p.s. Supposons aussi qu'il existe κ tel que $\forall x \in H, \sqrt{\mathbf{K}(x, x)} \leq \kappa$. Alors, pour la fonction de perte quadratique on a

$$P \left(\left| R_{emp}[\hat{f}^n] - R[\hat{f}^n] \right| > \beta + \varepsilon \right) \leq 2 \exp \left(- \frac{n\varepsilon^2}{2 \left(n\beta + M_Y \frac{\kappa}{\sqrt{\lambda}} \right)^2} \right),$$

$$\text{où } \beta = \frac{8\kappa^2 M_Y^2 \left(1 + \frac{1}{\sqrt{\lambda}} \right)^2}{n\lambda},$$

et pour la perte logistique,

$$P \left(\left| R_{emp}[\hat{f}^n] - R[\hat{f}^n] \right| > \varepsilon + \frac{\kappa^2}{n\lambda} \right) \leq 2 \exp \left(- \frac{2n\varepsilon^2 \lambda^2}{9\kappa^4} \right).$$

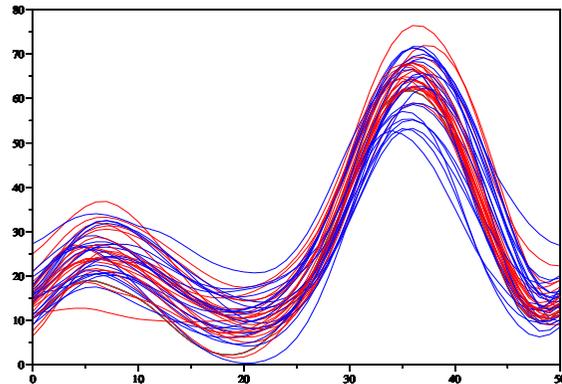
On a aussi :

Corollaire. *Avec les conditions du théorème de représentation on a :*

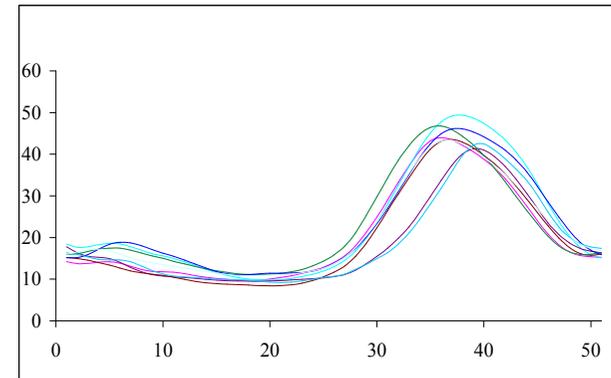
$$\begin{aligned} R_{emp}[f^{opt}] - R_{emp}[\hat{f}^n] &\xrightarrow{P} 0, \\ R[\hat{f}^n] - R[f^{opt}] &\xrightarrow{P} 0. \end{aligned}$$

Applications. Données fonctionnelles.

Parkinsoniens et l'angle de flexion du genou. Courbes de marche.



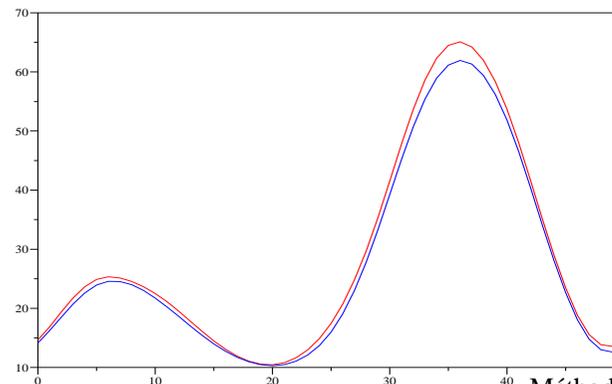
Sujets sains – (jeune–rouge, âgé–bleu)



Sujets parkinsoniens.

Question : Pour les sujets sains, l'âge a-t-il un effet sur la marche ?

Courbes moyennes :



30 sujets jeunes

30 sujets âgés.

Echantillon apprentissage : 40. Echantillon test : 20. Nombre échantillons : 100

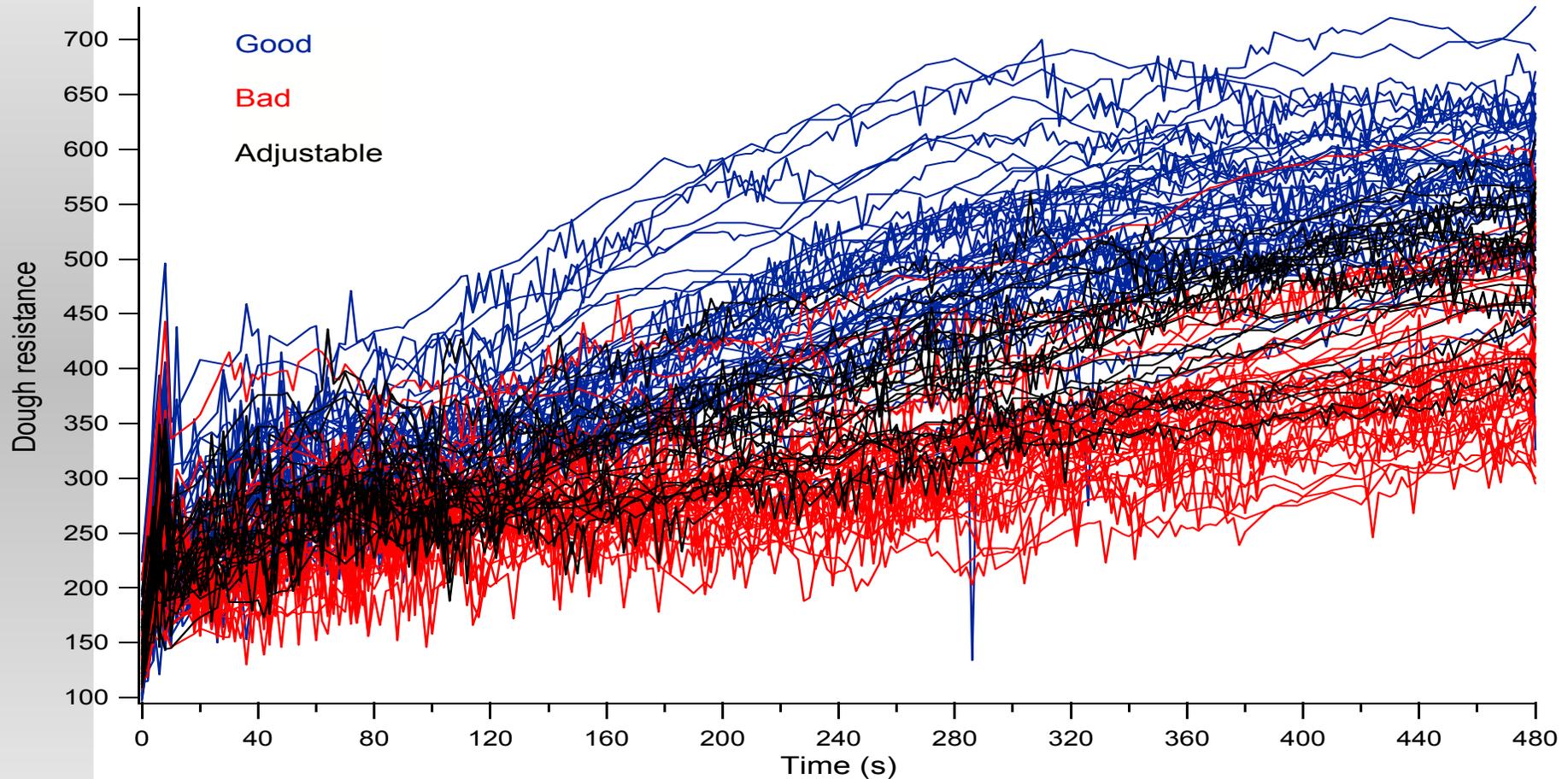
Modèle	PLS_FLDA	PCR_FLog(4))	Gaussien (1) $\lambda = 0.12$	IP(1,2) $\lambda = 0.18$	IP(1,3) $\lambda = 0.22$
Aire ROC	0.780	0.790	0.760	0.780	0.750

Moyenne de l'aire de la surface sous la courbe ROC.

Taux moyen de mal classés pour le meilleur modèle $\approx 30\%$

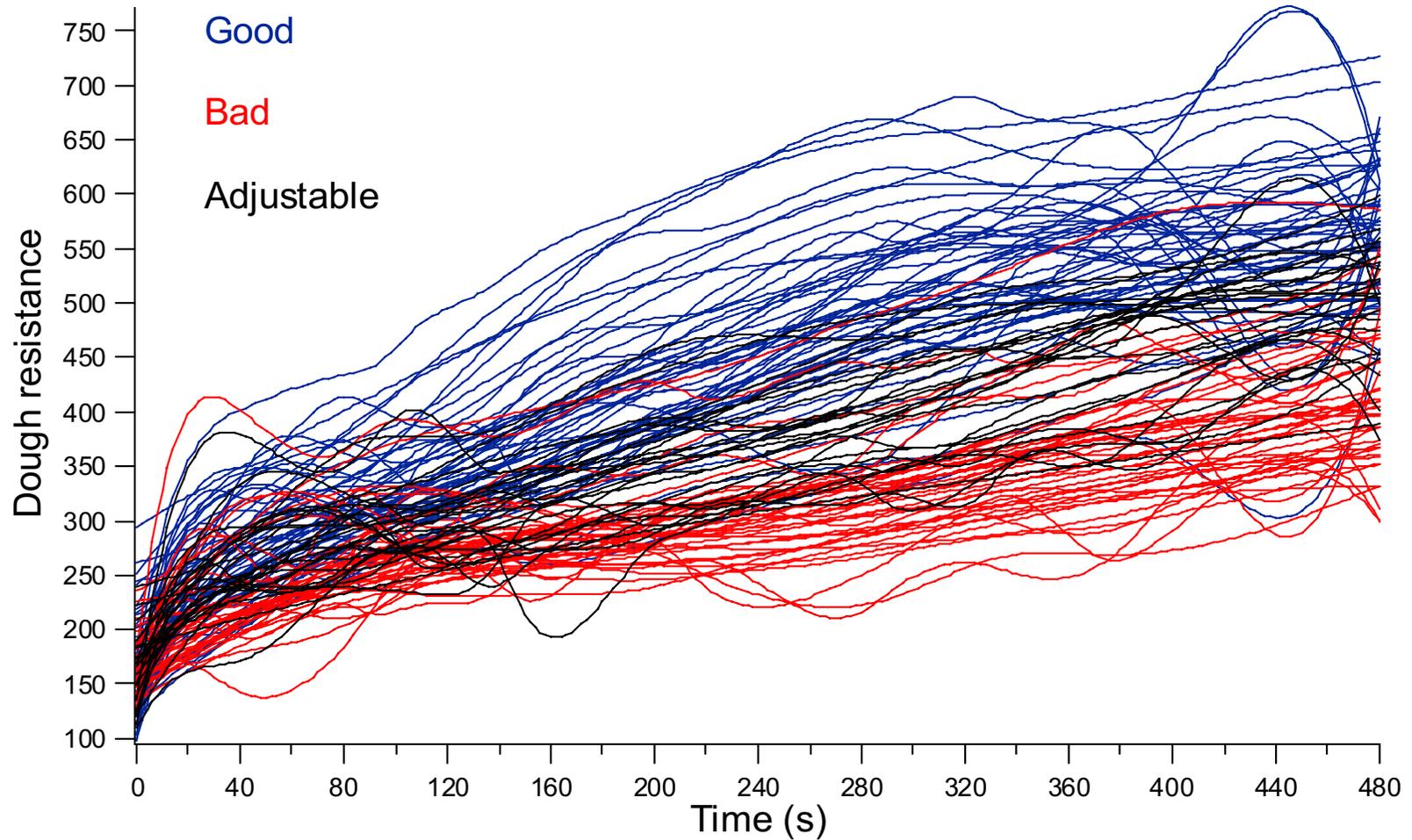
Les farines de Danone

$X = \{X_t\}_{t \in [0, 480]}$: résistance de la pâte durant 480s de pétrissage



115 farines ;

Qualité des biscuits : 50 (bonne), 40 (mauvaise), 25 (ajustable)



Lissage avec des fonction B-splines cubique avec 16 noeuds

Résultats :

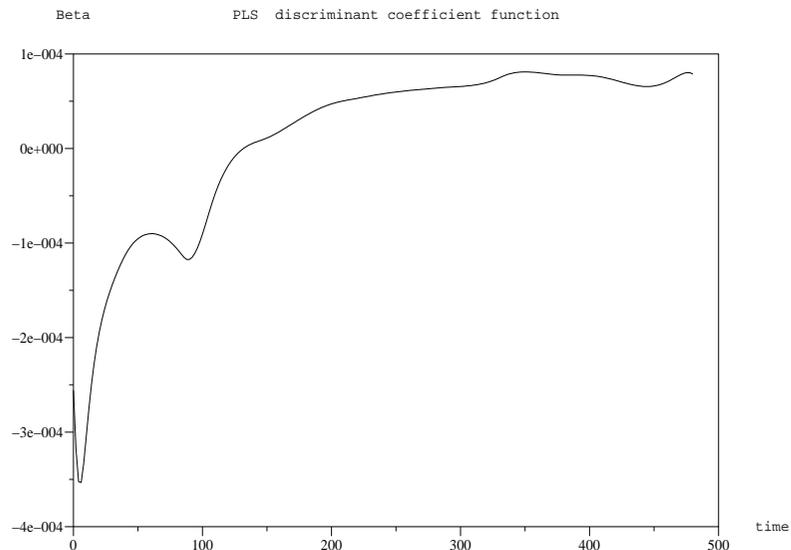
$Y \in \{\text{bonne, mauvaise}\}$

100 échantillons ; (apprentissage : 60 ; test : 30)

Modèle	PLS_FLDA	K-NN(13)	PC_FLDA	Gaussien(6)	LDA
Taux de mal classés	0.112	0.103	0.142	0.108	0.154

Moyenne du taux des mal classés sur un ensemble de 100 échantillons test.

La fonction discriminante $\Phi(X) = -1.46 + \int_0^{480} X_t \hat{\beta}_{PLS}(t) dt$.



$Y \in \{\text{bonne, ajustable, mauvaise}\}$

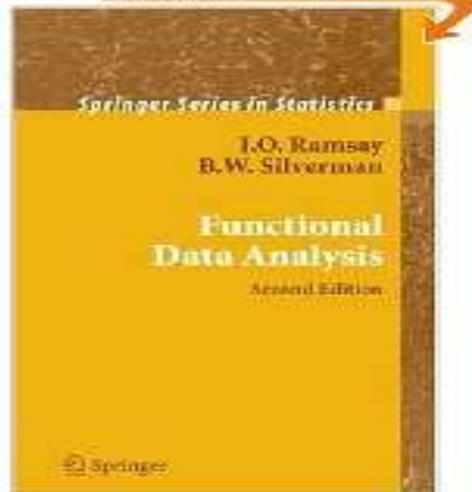
100 échantillons ; (apprentissage : 70 ; test : 35)

Modèle	PLS_FLDA	K-NN(13)	PC_FLDA	Gaussien(6)	LDA
Taux de mal classés	0.258	0.245	0.262	0.247	0.282

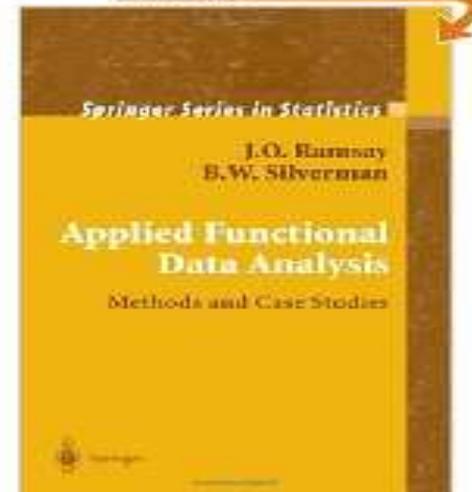
Moyenne du taux des mal classés sur un ensemble de 100 échantillons test.

Bibliographie

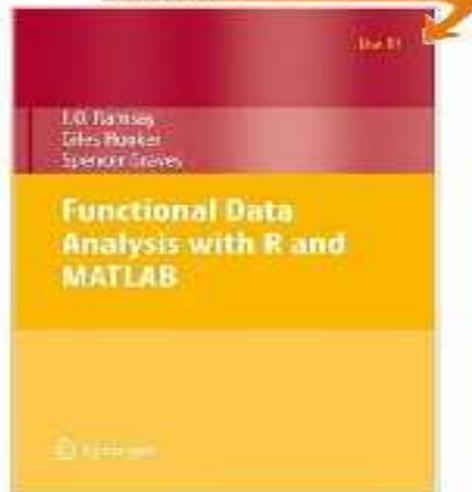
Cliquez pour **Feuilleter!**



Cliquez pour **Feuilleter!**



Cliquez pour **Feuilleter!**



Cliquez pour **Feuilleter!**

