

Modèle probabiliste à base de processus latent pour la description et la classification de signaux.

Faïcel Chamroukhi



Institut National de Recherche sur les Transports et leur Sécurité

Directeur : Patrice Aknin, Encadrant : Allou Samé

Université de Technologie de Compiègne

Directeur : Gérard Govaert

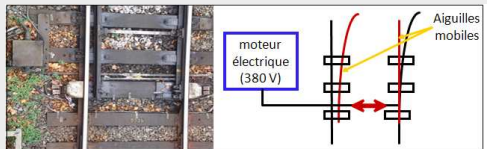
08 Avril 2010

- 1 Introduction
- 2 Aspects bibliographiques
 - Modèle de régression par morceaux
 - Modèle de régression markovien
- 3 Modèle de régression à processus latent proposé
- 4 Apprentissage supervisé à partir d'un ensemble de signaux
- 5 Apprentissage non supervisé à partir de signaux (clustering)
- 6 Conclusion et perspectives

Problématique de diagnostic

- ▶ Collaboration avec la SNCF
- ▶ Diagnostic d'un composant de l'infrastructure ferroviaire

Mécanisme d'aiguillage



Objectifs visés

Aide à la décision dans le cadre d'une maintenance préventive :

- ▶ Détecter les défauts afin de déployer l'équipe de maintenance
- ▶ Suivi de l'état de fonctionnement au cours du temps

Apprentissage statistique

Apprentissage statistique pour le diagnostic

Apprendre les modes de fonctionnement du système étudié à partir de mesures issues de capteurs

$$f : x \rightarrow y$$

Contexte

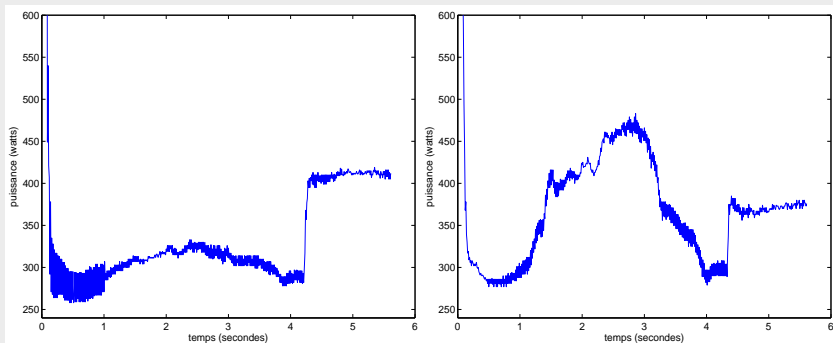
- 1 Apprentissage supervisé : données (x,y)
- 2 Apprentissage non supervisé : données $x, y?$

Génératif/Discriminatif? (Jebara, 2001)

- 1 Approche discriminative : apprendre directement $p(y|x)$
- 2 Approche générative : apprendre $p(x,y) \Rightarrow p(y|x) \propto p(x|y)p(y)$
 - Comment les données ont été générées?
 - S'adapte "facilement" au contexte non supervisé

Données exploitées

- ▶ Signaux de puissance consommée par le moteur d'aiguille durant chaque manœuvre



- ▶ Particularité des signaux : plusieurs régimes se succédant de manière brusque ou lente
- ▶ Solution proposée : utiliser un modèle de régression adapté, dont les paramètres vont servir de caractéristiques des signaux.

Description de signaux et classification dans l'espace des descripteurs

Modèle de régression par morceaux (McGee and Carleton, 1970)

Données

- ▶ échantillon $((x_1, t_1), \dots, (x_n, t_n))$ où x_i est l'observation à l'instant t_i

Modèle

$$x_i = \beta_k^T \mathbf{r}_i + \sigma_k \epsilon_i \quad (i = 1, \dots, n)$$

- ▶ k vérifiant $i \in I_k = (\gamma_k, \gamma_{k+1}]$: indices des éléments du segment k
- ▶ $\mathbf{r}_i = (1, t_i, \dots, t_i^p)^T$: vecteur-covariable aux coeff. de régression $\beta_k \in \mathbb{R}^{p+1}$

Paramètres du modèle et estimation

- ▶ (ψ, γ) avec $\psi = (\beta_1, \dots, \beta_K, \sigma_1^2, \dots, \sigma_K^2)$ et $\gamma = (\gamma_1, \dots, \gamma_{K+1})$.
- ▶ minimiser $J(\psi, \gamma) = \sum_{k=1}^K \sum_{i \in I_k} \left[\log \sigma_k^2 + \frac{(x_i - \beta_k^T \mathbf{r}_i)^2}{\sigma_k^2} \right]$ par rapport à (ψ, γ)
- ▶ J est additif sur k
- ▶ Optimization globale par la programmation dynamique (Bellman, 1961; Stone, 1961; Lechevalier, 1990;)

Approximation et segmentation d'un signal

- ▶ $\hat{x}_i = \sum_{k=1}^K \hat{z}_{ik} \hat{\beta}_k^T r_i \quad (i = 1, \dots, n)$
- ▶ $\hat{z}_{ik} = 1$ si $i \in (\hat{\gamma}_k, \hat{\gamma}_{k+1}]$ (x_i appartient au k ème segment) et $\hat{z}_{ik} = 0$ sinon

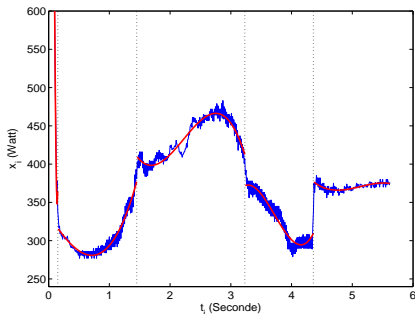


FIGURE – Exemple de résultat graphique obtenu avec le modèle de régression par morceaux

- ▶ La programmation dynamique peut s'avérer coûteuse en temps de calcul
- ▶ Fournit une segmentation dure \Rightarrow adapté pour les changements brusques

Modèle de régression régi par un HMM (Fridman, 1993)

Modèle

$$x_i = \beta_{z_i}^T \mathbf{r}_i + \sigma_{z_i} \epsilon_i \quad (i = 1, \dots, n)$$

- ▶ z_i : variable latente discrète $\in \{1, \dots, K\} \rightarrow$ le label du modèle de régression de x_i
- ▶ $\mathbf{z} = (z_1, \dots, z_n)$ est une chaîne de Markov homogène d'ordre 1

Paramètres du modèle

$$\Psi = (\pi, A, \beta_1, \dots, \beta_K, \sigma_1^2, \dots, \sigma_K^2)$$

Modèle de régression régi par un HMM (Fridman, 1993)

Modèle

$$x_i = \beta_{z_i}^T \mathbf{r}_i + \sigma_{z_i} \epsilon_i \quad (i = 1, \dots, n)$$

- ▶ z_i : variable latente discrète $\in \{1, \dots, K\} \rightarrow$ le label du modèle de régression de x_i
- ▶ $\mathbf{z} = (z_1, \dots, z_n)$ est une chaîne de Markov homogène d'ordre 1

Paramètres du modèle

$$\Psi = (\pi, A, \beta_1, \dots, \beta_K, \sigma_1^2, \dots, \sigma_K^2)$$

Contraintes sur la matrice de transition

- ▶ Signaux d'aiguillages : mouvements successifs liés à la mécanique du système
- ▶ \Rightarrow Contrainte : $p(z_i = k | z_{i-1} = \ell) = 0$ si $k > \ell + 1$ ou $k < \ell$
 \Rightarrow Pas de saut d'état et pas de retour à un état précédent
- ▶ Cas particulier du modèle gauche-droit (Rabiner, 1989)

Estimation des paramètres par Maximum de Vraisemblance

- ▶ log-vraisemblance :

$$\mathcal{L}(\Psi; \mathbf{x}) = \log \sum_{z_1, \dots, z_n} p(z_1; \pi) \prod_{i=2}^n p(z_i | z_{i-1}; A) \prod_{i=1}^n \mathcal{N}(x_i; \beta_{z_i}^T \mathbf{r}_i, \sigma_{z_i}^2)$$

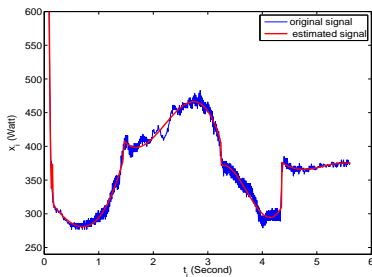
- ▶ Maximiser $\mathcal{L}(\Psi; \mathbf{x})$ par l'algorithme Espérance-Maximisation (EM) (Baum et al., 1970; Dempster et al., 1977)

Estimation des paramètres par Maximum de Vraisemblance

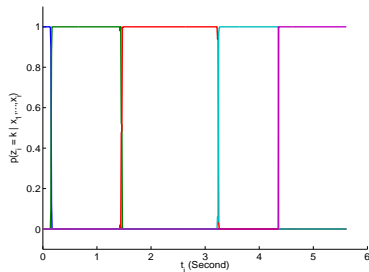
- log-vraisemblance :

$$\mathcal{L}(\Psi; \mathbf{x}) = \log \sum_{z_1, \dots, z_n} p(z_1; \pi) \prod_{i=2}^n p(z_i | z_{i-1}; A) \prod_{i=1}^n \mathcal{N}(x_i; \beta_{z_i}^T \mathbf{r}_i, \sigma_{z_i}^2)$$

- Maximiser $\mathcal{L}(\Psi; \mathbf{x})$ par l'algorithme Espérance-Maximisation (EM) (Baum et al., 1970; Dempster et al., 1977)



Signal et modèle estimé



Probabilités de filtrage

FIGURE – Exemple de résultat graphique obtenu avec le modèle markovien

Modèle de régression à processus latent proposé

(Chamroukhi et al., 2009)

Modèle

$$x_i = \beta_{z_i}^T \mathbf{r}_i + \sigma_{z_i} \epsilon_i \quad (i = 1, \dots, n)$$

- ▶ $\epsilon_i \sim \mathcal{N}(0, 1)$
- ▶ $\mathbf{z} = (z_1, \dots, z_n)$ est un processus logistique caché

Modèle de régression à processus latent proposé

(Chamroukhi et al., 2009)

Modèle

$$x_i = \beta_{z_i}^T \mathbf{r}_i + \sigma_{z_i} \epsilon_i \quad (i = 1, \dots, n)$$

- ▶ $\epsilon_i \sim \mathcal{N}(0, 1)$
- ▶ $\mathbf{z} = (z_1, \dots, z_n)$ est un processus logistique caché

$z_i \sim \mathcal{M}(1, \pi_{i1}(\mathbf{w}), \dots, \pi_{iK}(\mathbf{w}))$; où

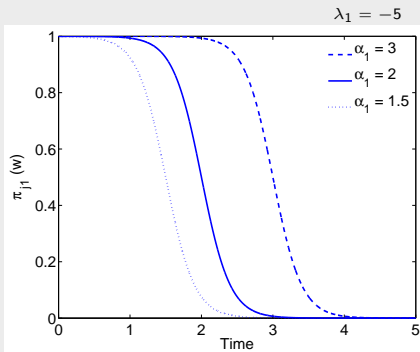
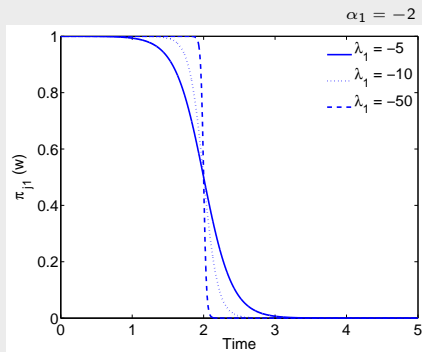
$$\pi_{ik}(\mathbf{w}) = p(z_i = k; \mathbf{w}) = \frac{\exp(\mathbf{w}_k^T \mathbf{v}_i)}{\sum_{\ell=1}^K \exp(\mathbf{w}_\ell^T \mathbf{v}_i)}$$

- ▶ $\mathbf{v}_i = (1, t_i)^T \in \mathbb{R}^2$
- ▶ $\mathbf{w}_k = (\mathbf{w}_{k0}, \mathbf{w}_{k1})^T$ vecteur paramètre de la *kième* fonction logistique $\in \mathbb{R}^2$
- ▶ $\mathbf{w} = (\mathbf{w}_1, \dots, \mathbf{w}_K)$ vecteur paramètre pour les K fonctions logistiques $\in \mathbb{R}^{2K}$

Flexibilité de la transformation logistique :

Variation des proportions logistiques en fonction de w :

- ▶ Paramétrisation de w_k par $w_k = \lambda_k(\alpha_k, 1)^T$
- ▶ Exemple pour $K = 2$:



- ⇒ Le paramètre λ_k contrôle la qualité des transitions entre les régimes
- ⇒ Le paramètre α_k contrôle l'instant de transition

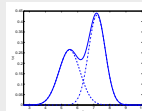
Estimation des paramètres par maximum de vraisemblance

- ▶ Vecteur paramètre du modèle :

$$\boldsymbol{\theta} = (\mathbf{w}, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_K, \sigma_1^2, \dots, \sigma_K^2)$$

- ▶ Densité mélange

$$p(x_i; \boldsymbol{\theta}) = \sum_{k=1}^K \pi_{ik}(\mathbf{w}) \mathcal{N}(x_i; \boldsymbol{\beta}_k^T \mathbf{r}_i, \sigma_k^2)$$



- ▶ Log-vraisemblance de $\boldsymbol{\theta}$:

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{x}) = \sum_{i=1}^n \log \sum_{k=1}^K \pi_{ik}(\mathbf{w}) \mathcal{N}(x_i; \boldsymbol{\beta}_k^T \mathbf{r}_i, \sigma_k^2).$$

- ▶ Maximisation de $\mathcal{L}(\boldsymbol{\theta}; \mathbf{x})$ par un algorithme Expectation-Maximization (EM) dédié (Chamroukhi et al., 2009)

Algorithme EM

- **Initialisation** : $\theta^{(0)}$, $q \leftarrow 0$

Algorithme EM

- **Initialisation** : $\theta^{(0)}$, $q \leftarrow 0$

- 1 **Étape E : Espérance** Calcul de l'espérance conditionnelle de la log-vraisemblance pour les données complétées

$$\begin{aligned}
 Q(\theta, \theta^{(q)}) &= \mathbb{E} \left[\mathcal{L}(\theta; \mathbf{x}, \mathbf{z}) | \mathbf{x}, \theta^{(q)} \right] \\
 &= \underbrace{\sum_{i=1}^n \sum_{k=1}^K \tau_{ik}^{(q)} \log \pi_{ik}(\mathbf{w})}_{Q_1(\mathbf{w})} + \underbrace{\sum_{k=1}^K \sum_{i=1}^n \tau_{ik}^{(q)} \log \mathcal{N}(x_i; \beta_k^T \mathbf{r}_i, \sigma_k^2)}_{Q_2(\beta_k, \sigma_k^2; k=1, \dots, K)},
 \end{aligned}$$

$$\text{avec } \tau_{ik}^{(q)} = p(z_i = k | x_i; \theta^{(q)}) = \frac{\pi_{ik}(\mathbf{w}^{(q)}) \mathcal{N}(x_i; \beta_k^T \mathbf{r}_i, \sigma_k^2(q))}{\sum_{\ell=1}^K \pi_{i\ell}(\mathbf{w}^{(q)}) \mathcal{N}(x_i; \beta_\ell^T \mathbf{r}_i, \sigma_\ell^2(q))}$$

Algorithme EM

- **Initialisation** : $\theta^{(0)}$, $q \leftarrow 0$

- 1 **Étape E : Espérance** Calcul de l'espérance conditionnelle de la log-vraisemblance pour les données complétées

$$\begin{aligned}
 Q(\theta, \theta^{(q)}) &= \mathbb{E} \left[\mathcal{L}(\theta; \mathbf{x}, \mathbf{z}) | \mathbf{x}, \theta^{(q)} \right] \\
 &= \underbrace{\sum_{i=1}^n \sum_{k=1}^K \tau_{ik}^{(q)} \log \pi_{ik}(\mathbf{w})}_{Q_1(\mathbf{w})} + \underbrace{\sum_{k=1}^K \sum_{i=1}^n \tau_{ik}^{(q)} \log \mathcal{N}(x_i; \beta_k^T \mathbf{r}_i, \sigma_k^2)}_{Q_2(\beta_k, \sigma_k^2; k=1, \dots, K)},
 \end{aligned}$$

$$\text{avec } \tau_{ik}^{(q)} = p(z_i = k | x_i; \theta^{(q)}) = \frac{\pi_{ik}(\mathbf{w}^{(q)}) \mathcal{N}(x_i; \beta_k^T \mathbf{r}_i, \sigma_k^2(q))}{\sum_{\ell=1}^K \pi_{i\ell}(\mathbf{w}^{(q)}) \mathcal{N}(x_i; \beta_\ell^T \mathbf{r}_i, \sigma_\ell^2(q))}$$

- 2 **Étape M : Maximisation** Calcul de $\theta^{(q+1)} = \arg \max_{\theta} Q(\theta, \theta^{(q)})$

- $q \leftarrow q + 1$
- Convergence (Xu and Jordan, 1996)

Détails de l'étape M

- 1 Maximisation de Q_2 par rapport aux $\{\beta_k, k = 1, \dots, K\}$: résolution analytique de K problèmes de régression polynomiale pondérés

- $\beta_k^{(q+1)} = \arg \min_{\beta_k} \sum_{i=1}^n \tau_{ik}^{(q)} (x_i - \beta_k^T r_i)^2 = (\mathbf{M}^T \Gamma_k^{(q)} \mathbf{M})^{-1} \mathbf{M}^T \Gamma_k^{(q)} \mathbf{x}$,

avec \mathbf{M} la matrice de régression et $\Gamma_k^{(q)} = \text{diag}(\tau_{1k}^{(q)}, \dots, \tau_{nk}^{(q)})$

Maximisation de Q_2 par rapport aux σ_k^2 :

- $\sigma_k^{2(q+1)} = \frac{1}{\sum_{i=1}^n \tau_{ik}^{(q)}} \sum_{i=1}^n \tau_{ik}^{(q)} (x_i - \beta_k^{T(q+1)} r_i)^2$

Détails de l'étape M

- 1 Maximisation de Q_2 par rapport aux $\{\beta_k, k = 1, \dots, K\}$: résolution analytique de K problèmes de régression polynomiale pondérés

- $\beta_k^{(q+1)} = \arg \min_{\beta_k} \sum_{i=1}^n \tau_{ik}^{(q)} (x_i - \beta_k^T r_i)^2 = (\mathbf{M}^T \Gamma_k^{(q)} \mathbf{M})^{-1} \mathbf{M}^T \Gamma_k^{(q)} \mathbf{x}$,

avec \mathbf{M} la matrice de régression et $\Gamma_k^{(q)} = \text{diag}(\tau_{1k}^{(q)}, \dots, \tau_{nk}^{(q)})$

Maximisation de Q_2 par rapport aux σ_k^2 :

- $\sigma_k^{2(q+1)} = \frac{1}{\sum_{i=1}^n \tau_{ik}^{(q)}} \sum_{i=1}^n \tau_{ik}^{(q)} (x_i - \beta_k^{T(q+1)} r_i)^2$

- 2 Maximisation de Q_1 par rapport \mathbf{w} : problème d'optimisation convexe de régression logistique multinomiale pondéré \Rightarrow Algorithme IRLS (Green, 1984)

$$\mathbf{w}^{(m+1)} = \mathbf{w}^{(m)} - \left[\frac{\partial^2 Q_1(\mathbf{w})}{\partial \mathbf{w} \partial \mathbf{w}^T} \right]_{\mathbf{w}=\mathbf{w}^{(m)}}^{-1} \frac{\partial Q_1(\mathbf{w})}{\partial \mathbf{w}} \Big|_{\mathbf{w}=\mathbf{w}^{(m)}}$$

\Rightarrow L'algorithme IRLS fournit le paramètre $\mathbf{w}^{(q+1)}$

Approximation et segmentation d'un signal

Approximation d'un signal

Comme dans la régression classique, $x_i (i = 1, \dots, n)$ est approximé par son espérance :

$$\begin{aligned}\hat{x}_i = \mathbb{E}(x_i; \hat{\theta}) &= \int_{\mathbb{R}} x_i p(x_i; \hat{\theta}) dx_i \\ &= \sum_{k=1}^K \pi_{ik}(\hat{\mathbf{w}}) \hat{\beta}_k^T \mathbf{r}_i\end{aligned}$$

Une somme de polynômes pondérés par des fonctions logistiques $\pi_{ik}(\hat{\mathbf{w}})$

⇒ Adapté pour les changement lents ou rapides

⇒ Assure la continuité du signal estimé

Segmentation d'un signal

Estimer le label \hat{z}_i du composant générant x_i par la règle suivante :

$$\hat{z}_i = \arg \max_{1 \leq k \leq K} \pi_{ik}(\hat{\mathbf{w}})$$

Expérimentations sur des données simulées

Critères d'évaluation

- ▶ écart quadratique moyen $\frac{1}{n} \sum_{i=1}^n [\mathbb{E}(x_i; \theta) - \hat{x}_i]^2$ θ étant le vrai paramètre
- ▶ Erreur de segmentation
- ▶ Comparaison de l'algorithme proposé avec les deux alternatives

Protocole de simulations

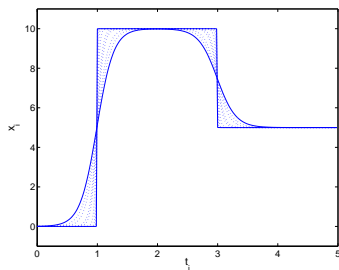
évaluation de la qualité des estimations :

- ▶ Expérience 1 : effet du niveau de souplesse des transitions
- ▶ Expérience 2 : effet de la taille d'échantillon n
- ▶ Expérience 3 : effet du niveau de bruit σ

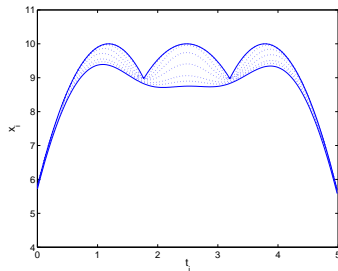
20 jeux de données différents ont été simulés pour chaque situation et les résultats ont été moyennés

Situation 1	$\beta_1 = 0$ $\beta_2 = 10$ $\beta_3 = 5$	$w_1 = [3341.33, -1706.96]$ $w_2 = [2436.97, -810.07]$ $w_3 = [0, 0]$
Situation 2	$\beta_1 = [-0.64, 14.4, -6]$ $\beta_2 = [-21.25, 25, -5]$ $\beta_3 = [-78.64, 45.6, -6]$	$w_1 = [3767.58, -1510.19]$ $w_2 = [2468.99, -742.55]$ $w_3 = [0, 0]$

Niveau de souplesse des transitions	1	2	3	4	5	6	7	8	9	10
(a) $\lambda_k = w_{k1}/$	1	2	5	10	20	40	50	80	100	125
(b) $\lambda_k = w_{k1}/$	1	10	50	100	150	200	250	275	300	400

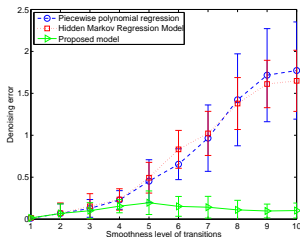


(a)

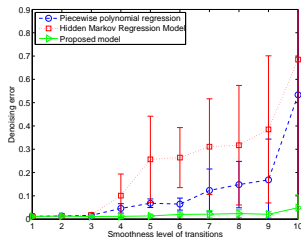


(b)

Résultats : EQM en fonction de la qualité des transitions

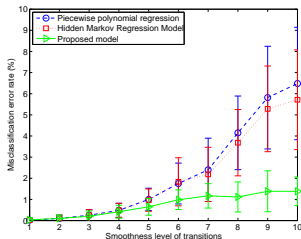


(a)

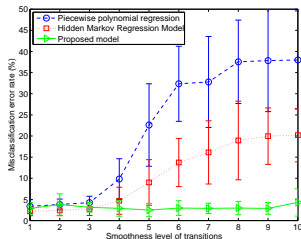


(b)

Erreur de segmentation en fonction de la qualité des transitions

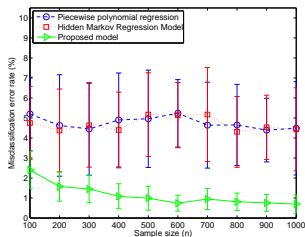
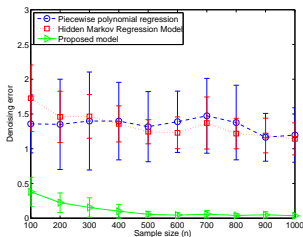


(a)

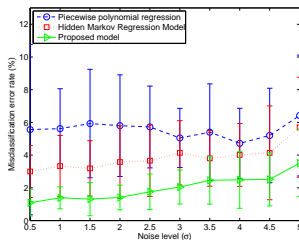
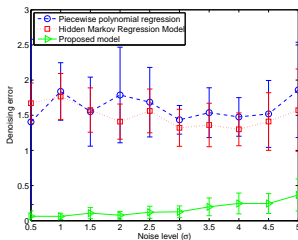


(b)

Résultats : EQM et erreur de segmentation en fonction de la taille d'échantillon n



EQM et erreur de segmentation en fonction du niveau de bruit



Exemple de signaux de manœuvres d'aiguillage :

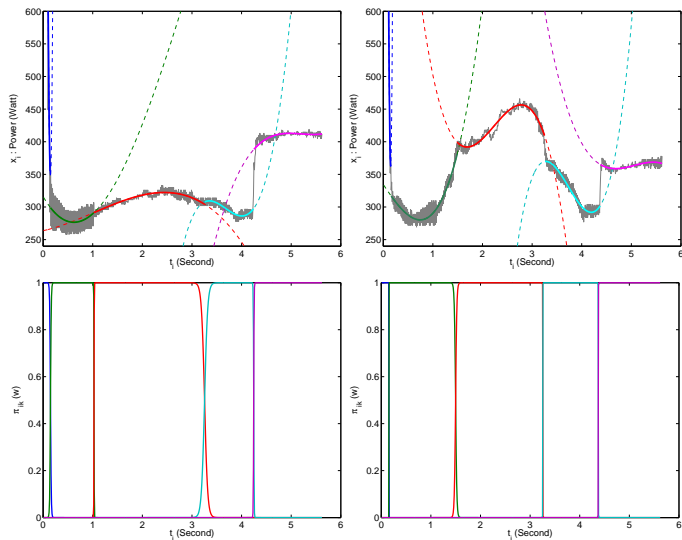


FIGURE – Régimes polynomiaux estimés et fonctions logistiques associées

Exemple de signaux de manœuvres d'aiguillage :

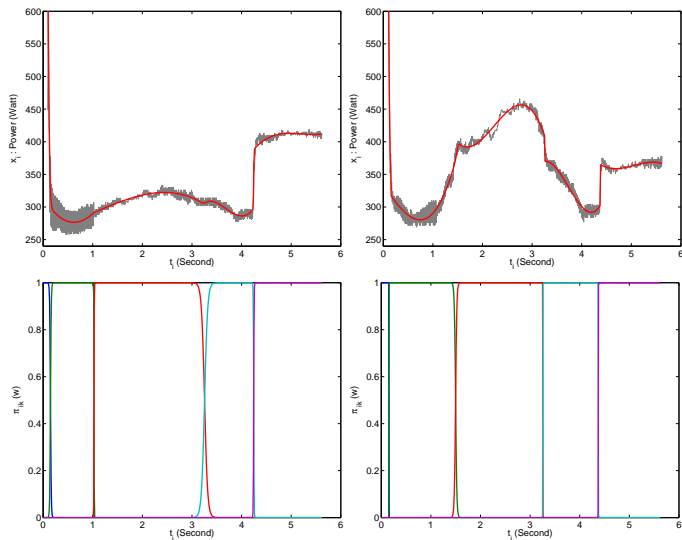


FIGURE – Modèles estimés et fonctions logistiques associées

Classement des signaux de manœuvres d'aiguillage

Apprentissage des paramètres des classes

- ▶ $N = 120$ signaux couvrant 3 classes : sans défaut, défaut mineur, défaut critique
- ▶ Extraction des paramètres par les 3 approches
- ▶ Apprentissage des paramètres des classes par MDA (Mixture Discriminant Analysis) (Hastie and Tibshirani, 1996)
- ▶ Classement des signaux par Maximum A Posteriori (MAP)

Classement des signaux de manœuvres d'aiguillage

Apprentissage des paramètres des classes

- ▶ $N = 120$ signaux couvrant 3 classes : sans défaut, défaut mineur, défaut critique
- ▶ Extraction des paramètres par les 3 approches
- ▶ Apprentissage des paramètres des classes par MDA (Mixture Discriminant Analysis) (Hastie and Tibshirani, 1996)
- ▶ Classement des signaux par Maximum A Posteriori (MAP)

Résultats de discrimination

Modèle	Taux d'erreur de test (%)
Régression par morceaux	17 \pm (4.5)
Régression par HMM	11 \pm (4.25)
Modèle proposé	9 \pm (3.72)

Apprentissage supervisé à partir d'un ensemble de signaux

Apprentissage supervisé à partir de plusieurs signaux

Travaux existants

- ▶ Cadre de l'analyse de données fonctionnelles (FDA) (Ramsay and Silverman, 2005)
- ▶ Approche Neuronale fonctionnelle (Rossi and Conan-Guez, 2005)
- ▶ SVMs pour des données fonctionnelles (Rossi and Villa, 2006)
- ▶ Functional Linear Discriminant Analysis basée sur les splines (James and Hastie, 2001)
- ▶ Régression par morceaux pour des données fonctionnelles ()

Approche proposée : (Chamroukhi et al., 2010)

- ▶ Modèle : Modèle de régression à processus latent par classe de signaux
- ▶ Optimization : Algorithme EM dédié

Apprentissage supervisé des paramètres des classes

Données

- ▶ échantillon indépendant : $((\mathbf{x}_1, c_1), \dots, (\mathbf{x}_n, c_n))$ n signaux étiquetés
- ▶ $\mathbf{x}_i = ((x_{i1}, t_1), \dots, (x_{im}, t_m))$: i ème signal
- ▶ $c_i \in \{1, \dots, G\}$: classe du signal \mathbf{x}_i ($i = 1, \dots, n$)

Modèle par classe de signaux :

$$x_{ij} = \beta_{gz_j}^T \mathbf{r}_j + \sigma_{gz_j} \epsilon_{ij} \quad (i \text{ vérifie } c_i = g, j = 1, \dots, m)$$

- ▶ $\epsilon_{ij} \sim \mathcal{N}(0, 1)$
- ▶ $z_j \in \{1, \dots, K\}$ variable latente : label du régime polynomial de x_{ij}
- ▶ $\mathbf{z} = (z_1, \dots, z_m)$ est un processus logistique

Estimation des paramètres par Maximum de Vraisemblance

▶ $\theta_g = (\mathbf{w}^g, \beta_{g1}, \dots, \beta_{gK}, \sigma_{g1}^2, \dots, \sigma_{gK}^2)$ Vecteur paramètre de la classe g

▶ Vraisemblance pour un signal

$$p(\mathbf{x}_i; \theta_g) = \prod_{j=1}^m \sum_{k=1}^K \pi_{jk}(\mathbf{w}^g) \mathcal{N}(x_{ij}; \beta_{gk}^T \mathbf{r}_j, \sigma_{gk}^2)$$

▶ log-vraisemblance :

$$L(\theta_g; \{\mathbf{x}_i | c_i = g\}) = \sum_i \sum_{j=1}^m \log \sum_{k=1}^K \pi_{jk}(\mathbf{w}^g) \mathcal{N}(x_{ij}; \beta_{gk}^T \mathbf{r}_j, \sigma_{gk}^2)$$

▶ Maximum de vraisemblance mis en œuvre par EM (Chamroukhi et al., 2010)

Algorithme EM

- **Initialisation** : $\theta_g^{(0)}$, $q \leftarrow 0$

1 Étape E

$$Q(\theta_g, \theta_g^{(q)}) = E \left[CL(\theta_g; \mathbf{X}, \mathbf{z}) | \mathbf{X}; \theta_g^{(q)} \right]$$

$$= \sum_{i|c_i=g} \sum_{j=1}^m \sum_{k=1}^K \tau_{ijk}^{(q)} \log \pi_{jk}(\mathbf{w}^g) + \sum_{i|c_i=g} \sum_{j=1}^m \sum_{k=1}^K \tau_{ijk}^{(q)} \log \mathcal{N}(x_{ij}; \beta_{gk}^T \mathbf{r}_j, \sigma_{gk}^2)$$

$$\tau_{ijk}^{(q)} = p(z_j = k | x_{ij}; \theta_g^{(q)}) = \frac{\pi_{jk}(\mathbf{w}^{g(q)}) \mathcal{N}(x_{ij}; \beta_{gk}^T \mathbf{r}_j, \sigma_{gk}^2(q))}{\sum_{\ell=1}^K \pi_{j\ell}(\mathbf{w}^{g(q)}) \mathcal{N}(x_{ij}; \beta_{g\ell}^T \mathbf{r}_j, \sigma_{g\ell}^2(q))} \text{ probabilités a posteriori}$$

2 Étape M : Formules de mise à jour

▶ $\beta_{gk}^{(q+1)} = (\Lambda^T \mathbf{W}_k^{(q)} \Lambda)^{-1} \Lambda^T \mathbf{W}_k^{(q)} \mathbf{X}$ avec Λ mat. de rég. et $\mathbf{W}_k^{(q)} = (\tau_{11k}^{(q)}, \dots, \tau_{n_g m k}^{(q)})$

▶ $\sigma_{gk}^{2(q+1)} = \frac{1}{n_g \sum_{j=1}^m \tau_{ijk}^{(q)}} \sum_{i=1}^n \sum_{j=1}^m \tau_{ijk}^{(q)} (x_{ij} - \beta_{gk}^T \mathbf{r}_j)^2$

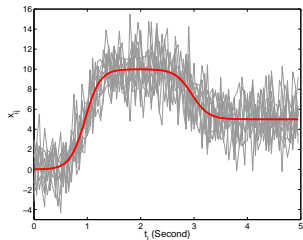
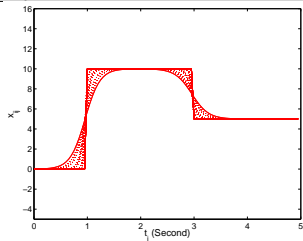
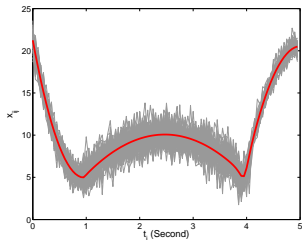
▶ $\mathbf{w}^{g(q+1)} = \arg \max_{\mathbf{w}} \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^K \tau_{ijk}^{(q)} \log \pi_{jk}(\mathbf{w}^g) \rightarrow \text{IRLS}$

- $q \leftarrow q + 1$

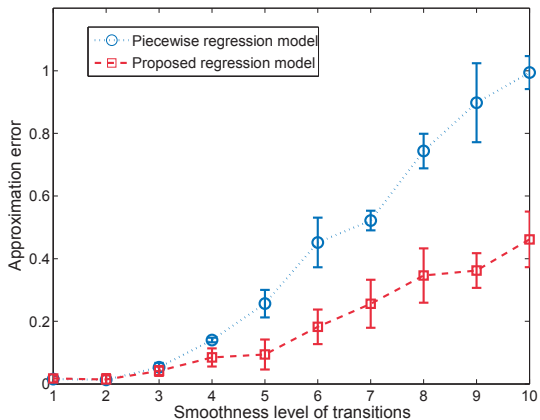
Données simulées

$\beta_1 = 0$	$w_1 = [3341.33, -1706.96]$
$\beta_2 = 10$	$w_2 = [2436.97, -810.07]$
$\beta_3 = 5$	$w_3 = [0, 0]$

$\beta_1 = [23, -36, 18]$	$w_1 = [92.72, -46.72]$	$\sigma_1 = 1$
$\beta_2 = [-3.9, 11.08, -2.2]$	$w_2 = [61.16, -15.28]$	$\sigma_2 = 1.25$
$\beta_3 = [-337, 141.5, -14]$	$w_3 = [0, 0]$	$\sigma_3 = 0.75$

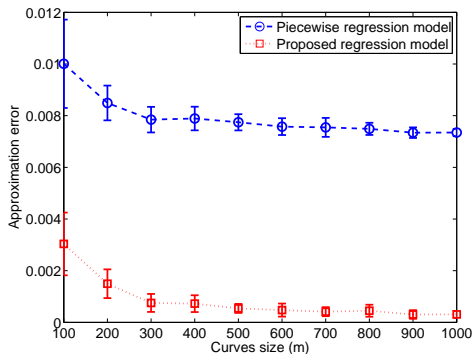
(a) ($n=10, m=100, \sigma = 2$)(b) (exemple : $n=50, m=100$)

Résultats : Erreur d'approximation des courbes en fonction de la qualité des transitions

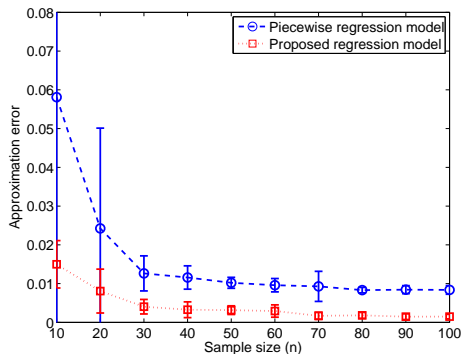


$n = 10$ courbes de $m = 100$ points

Résultats : Erreur d'approximation des courbes en fonction du nombre de courbes n et de la taille des courbes m

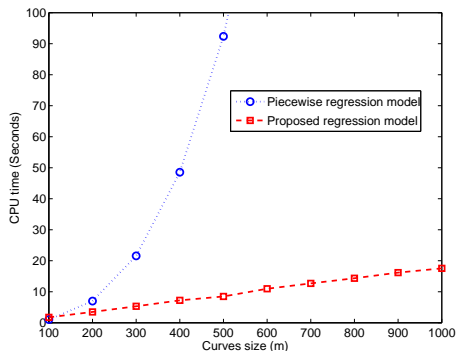


$n = 50$ courbes

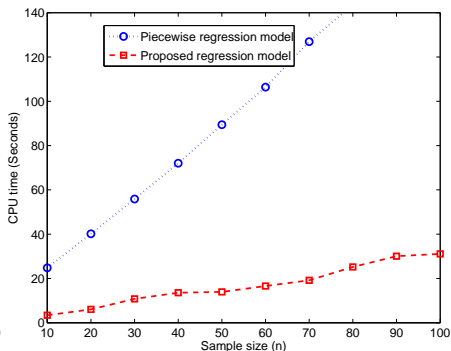


$m = 500$ points

Résultats : Temps de calcul moyen en fonction du nombre de courbes n et de la taille des courbes m :



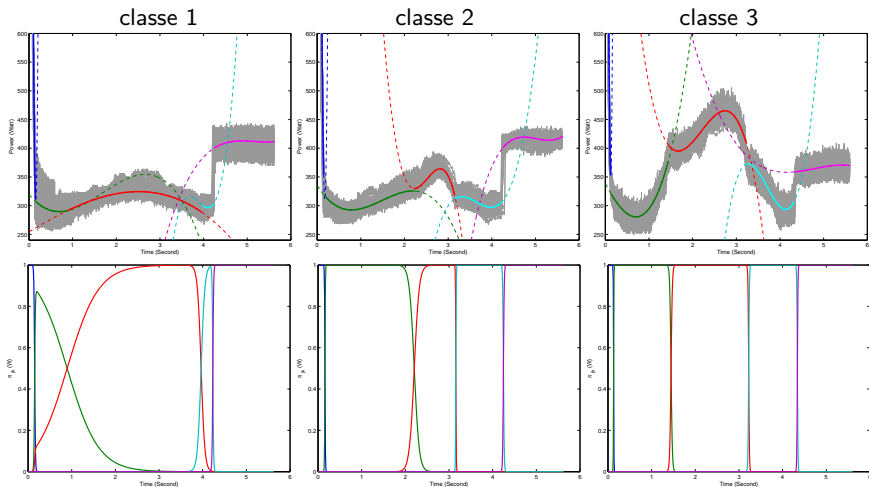
$n = 50$ courbes



$m = 500$ points

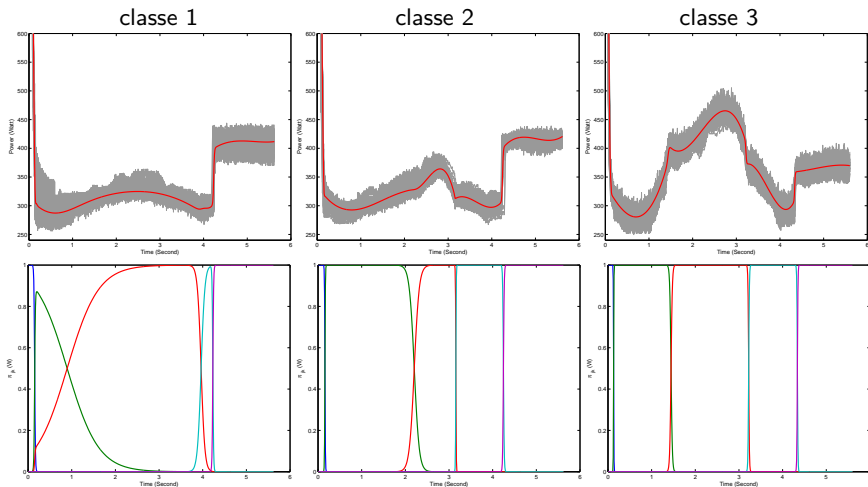
Signaux de manoeuvres d'aiguillage

FIGURE — Signaux, modèles estimés et les fonctions logistiques associées



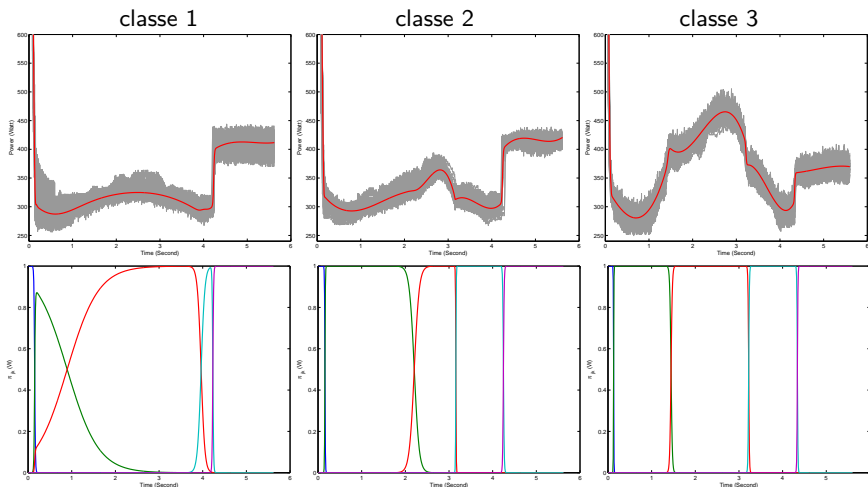
Signaux de manoeuvres d'aiguillage

FIGURE — Signaux, modèles estimés et les fonctions logistiques associées



Signaux de manoeuvres d'aiguillage

FIGURE — Signaux, modèles estimés et les fonctions logistiques associées



Taux d'erreur de classement (%)

Régression par morceaux
Modèle proposé

$1.82 \pm (5.74)$
 $1.67 \pm (2.28)$

Apprentissage non supervisé pour la classification automatique de signaux (clustering)

Classification automatique de signaux (clustering)

Travaux existants

- ▶ Mélange de régressions (Gaffney and Smyth, 2005)
- ▶ Régression splines (Liu and Yang, 2009)
- ▶ Régression par morceaux (Hébrail et al., 2010)

Approche proposée

- ▶ Modèle : Mélange de régressions à processus latents
- ▶ Optimization : Version classifiante de l'algorithme EM (CEM) (Celeux and Govaert, 1992)

Modèle en non supervisé

Modèle et paramètre

- ▶ Modèle défini précédemment \Rightarrow passage au contexte non supervisé
- ▶ $\Psi = (\alpha_1, \dots, \alpha_G, \theta_1, \dots, \theta_G)$: vecteur paramètre du modèle
- ▶ $\alpha_g = p(c_i = g)$: proportion de la classe g
- ▶ θ_g : paramètres de la classe g

Critère

- ▶ log-vraisemblance classifiante

$$\mathcal{L}_c(\Psi; \mathbf{X}, \mathbf{c}) = \sum_{i=1}^n \sum_{g=1}^G c_{ig} \log p(\mathbf{x}_i | c_i = g; \theta_g)$$

- ▶ $c_{ig}^{(q)} = 1$ si $c_i = g$ et 0 sinon
- ▶ $p(\mathbf{x}_i | c_i = g; \theta_g) = \prod_{j=1}^m \sum_{k=1}^K \pi_{kj}(\mathbf{w}_g^{(q)}) \mathcal{N}(x_{ij}; \beta_{gk}^{T(q)} \mathbf{r}_j, \sigma_{gk}^{2(q)})$

- **Initialisation** : $\Psi^{(0)}$, $q \leftarrow 0$

1 Étape E : Espérance

$$\begin{aligned}
 Q^*(\Psi, \Psi^{(q)}) &= \mathbb{E} \left[\mathcal{L}(\Psi; \mathbf{X}, \mathbf{c}, \mathbf{z}_1, \dots, \mathbf{z}_G) \mid \mathbf{X}; \Psi^{(q)} \right] \\
 &= \sum_{i=1}^n \sum_{g=1}^G h_{ig}^{(q)} \log \alpha_g + \sum_{i=1}^n \sum_{g=1}^G \sum_{j=1}^m \sum_{k=1}^K h_{ig}^{(q)} \tau_{ijgk}^{(q)} \log \pi_{jk}(\mathbf{w}_g) \\
 &\quad + \sum_{i=1}^n \sum_{g=1}^G \sum_{j=1}^m \sum_{k=1}^K h_{ig}^{(q)} \tau_{ijgk}^{(q)} \log \mathcal{N}(x_{ij}; \beta_{gk}^T \mathbf{r}_j, \sigma_{gk}^2)
 \end{aligned}$$

$h_{ig}^{(q)}$ = $p(c_i = g \mid \mathbf{x}_i; \Psi^{(q)})$: probabilités a posteriori d'appartenance à la classe g

$\tau_{ijgk}^{(q)}$ = $p(z_{jgk} = 1 \mid c_i = g, x_{ij}; \Psi^{(q)})$: probabilités a posteriori d'appartenance aux phases au sein de la classe g

2 Étape C : Classification $c_i^{(q+1)} = \arg \max_{1 \leq g \leq G} h_{ig}^{(q)}$

3 Étape M : Maximisation : $\Psi^{(q+1)} = \arg \max_{\Psi} \mathbb{E}_{\mathbf{z}} \left[\mathcal{L}_c(\Psi; \mathbf{X}, \mathbf{c}^{(q+1)}, \mathbf{z}_1, \dots, \mathbf{z}_G) \right]$

\Rightarrow pondérations : $c_{ig}^{(q+1)} \tau_{ijgk}^{(q)}$

- $q \leftarrow q + 1$

Illustration

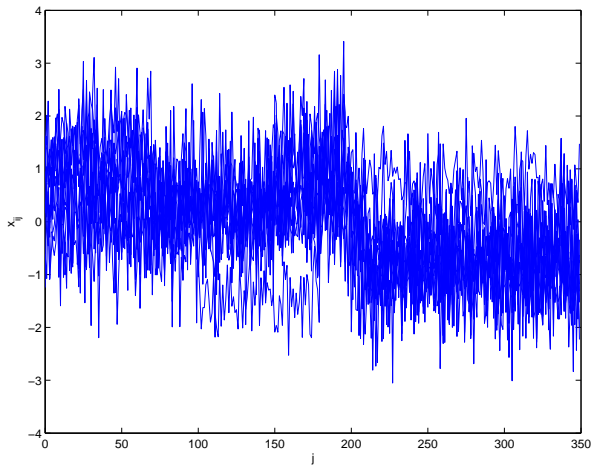


Illustration : Résultats graphiques

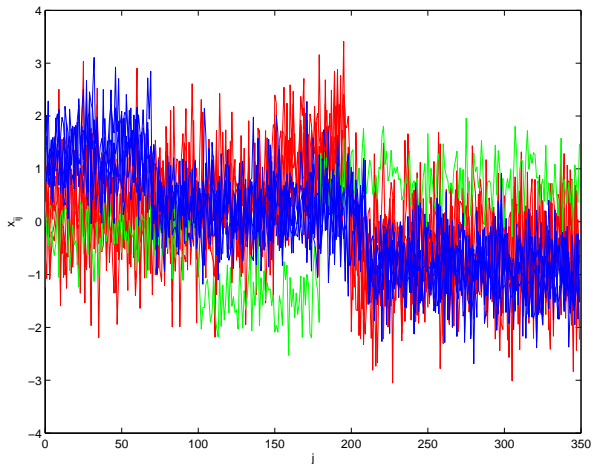
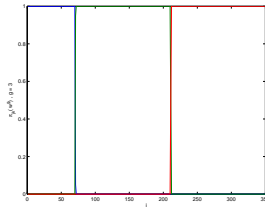
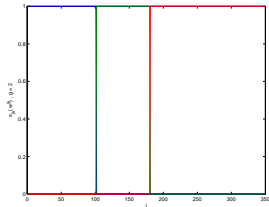
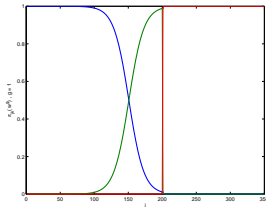
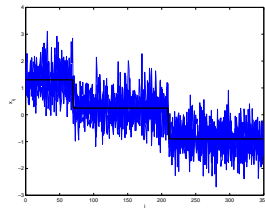
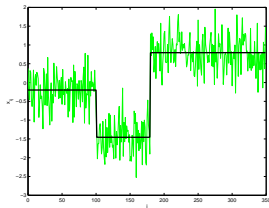
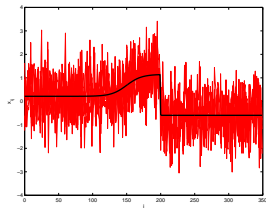
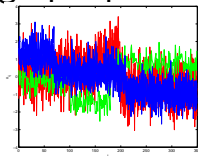


Illustration : Résultats graphiques



Vagues de Breiman (Breiman et al., 1984) : problème à 3 classes

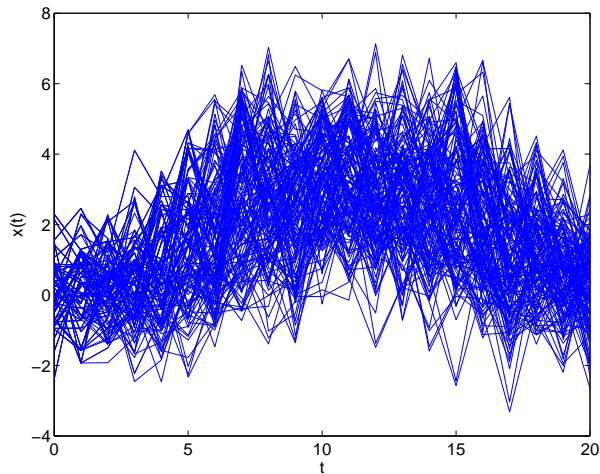


FIGURE – 150 courbes à 21 points

Vagues de Breiman : Résultats graphiques

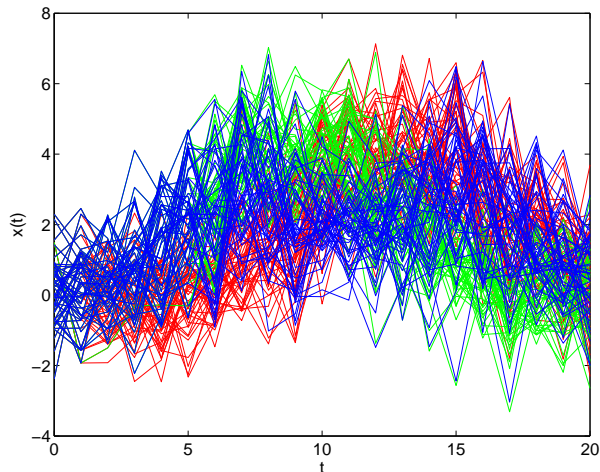
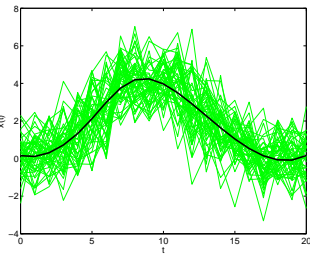
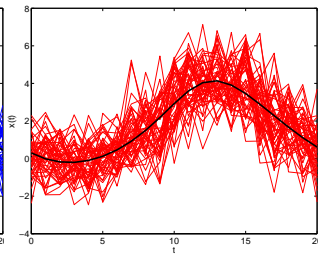
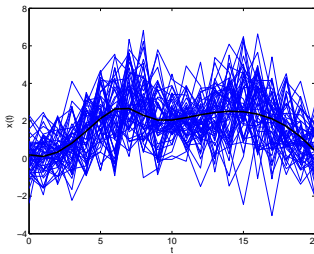
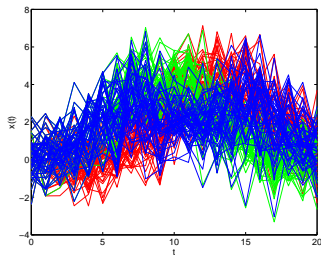
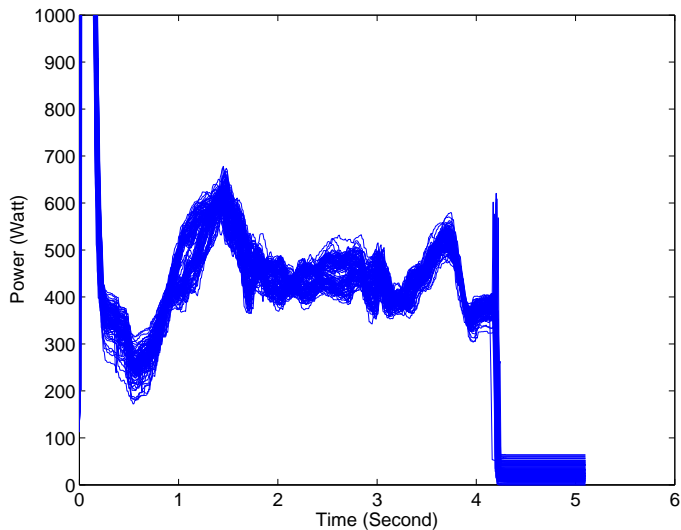


FIGURE – 150 courbes à 21 points

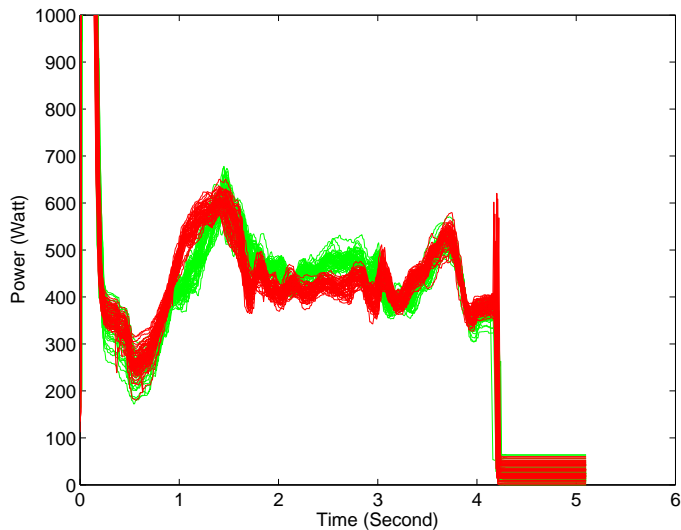
Vagues de Breiman : Résultats graphiques



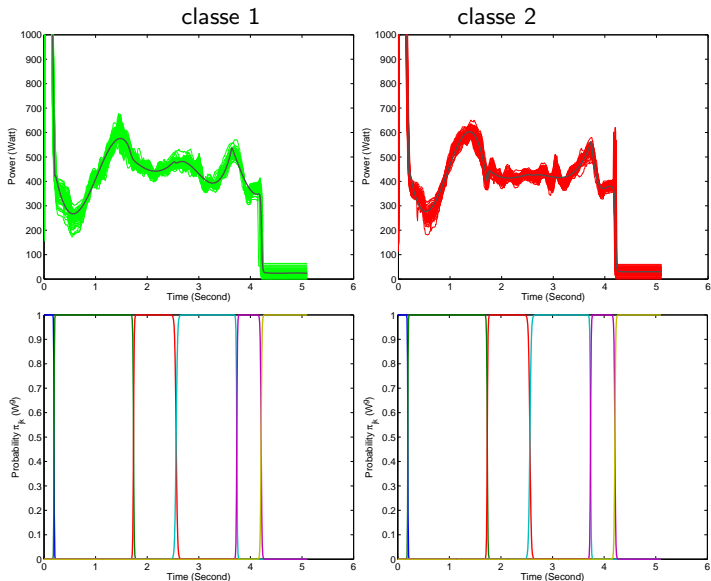
Données réelles : 115 signaux de 511 points



Données réelles : Résultats graphiques



Données réelles : Résultats graphiques



Conclusion

- ▶ Modèle de régression à processus latent proposé pour modéliser les signaux de manœuvres d'aiguillage.
- ▶ Avantages :
 - Adaptation aux changements de régimes souples ou rapides
 - L'EM constitue un cadre élégant pour l'optimisation des paramètres
- ▶ Apprentissage des modes de fonctionnement directement à partir d'un ensemble de signaux
- ▶ On obtient des résultats précis en terme de discrimination de signaux
- ▶ Adaptation au cadre non supervisé

Conclusion

- ▶ Modèle de régression à processus latent proposé pour modéliser les signaux de manœuvres d'aiguillage.
- ▶ Avantages :
 - Adaptation aux changements de régimes souples ou rapides
 - L'EM constitue un cadre élégant pour l'optimisation des paramètres
- ▶ Apprentissage des modes de fonctionnement directement à partir d'un ensemble de signaux
- ▶ On obtient des résultats précis en terme de discrimination de signaux
- ▶ Adaptation au cadre non supervisé

Perspectives

- ▶ En cours : Simulation poussée pour le cas non supervisée
Apprentissage en ligne d'un Modèle AR régi par un processus latent (Markovien, logistique) pour le suivi de l'état de fonctionnement du système
- ▶ À court terme : Functional Mixture Discriminant Analysis

Merci de votre attention !

- Baum, L., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. Annals of Mathematical Statistics, 41 :164–171.
- Bellman, R. (1961). On the approximation of curves by line segments using dynamic programming. Communications of the Association for Computing Machinery (CACM), 4(6) :284.
- Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). Classification And Regression Trees. Wadsworth, New York.
- Celeux, G. and Govaert, G. (1992). A classification EM algorithm for clustering and two stochastic versions. Computational Statistics and Data Analysis, 14 :315–332.
- Chamroukhi, F., Samé, A., Govaert, G., and Aknin, P. (2009). Time series modeling by a regression approach based on a latent process. Neural Networks, 22(5-6) :593–602.
- Chamroukhi, F., Samé, A., Govaert, G., and Aknin, P. (2010). A hidden process regression model for functional data description. application to curve discrimination. Neurocomputing, 73(7-9) :1210–1221.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. Journal of The Royal Statistical Society, B, 39(1) :1–38.
- Fridman, M. (1993). Hidden markov model regression. Technical report, Institute of mathematics, University of Minnesota.
- Gaffney, J. and Smyth, P. (2005). Joint probabilistic curve clustering and alignment. In In Advances in Neural Information Processing Systems (NIPS).
- Green, P. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. Journal of The Royal Statistical Society, B, 46(2) :149–192.
- Hastie, T. and Tibshirani, R. (1996). Discriminant analysis by gaussian mixtures. Journal of the Royal Statistical Society, B, 58 :155–176.
- Hébrail, G., Hugueney, B., Lechevallier, Y., and Rossi, F. (2010). Exploratory analysis of functional data via clustering and optimal segmentation. Neurocomputing, 73(7-9) :1125–1141.
- James, G. M. and Hastie, T. J. (2001). Functional linear discriminant analysis for irregularly sampled curves. Journal of the Royal Statistical Society Series B, 63 :533–550.
- Jebara, T. (2001). Discriminative, Generative and Imitative learning. Phd thesis, Media Laboratory, MIT.
- Lechevalier, Y. (1990). Optimal clustering on ordered set. Technical report, The French National Institute for Research in Computer Science and Control (INRIA).
- Liu, X. and Yang, M. (2009). Simultaneous curve registration and clustering for functional data. Computational Statistics and Data Analysis.
- McGee, V. E. and Carleton, W. T. (1970). Piecewise regression. Journal of the American Statistical Association, 65 :1109–1124.
- Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2) :257–286.
- Ramsay, J. O. and Silverman, B. W. (2005). Functional Data Analysis. Springer Series in Statistics. Springer.
- Rossi, F. and Conan-Guez, B. (2005). Functional multi-layer perceptron : a nonlinear tool for functional data analysis. Neural Networks, 18(1) :45–60.