

Modélisation probabiliste à base de processus latent de signaux monodimensionnels.

Faïcel Chamroukhi



Institut National de Recherche sur les Transports et leur Sécurité (INRETS)
Laboratoire des Technologies Nouvelles (LTN)

Université de Technologie de Compiègne (UTC)
Laboratoire HEUDIASYC, UMR CNRS 6599

Directeurs de thèse : Gérard Govaert (UTC)
Patrice Aknin (INRETS)

Encadrant : Allou Samé (INRETS)

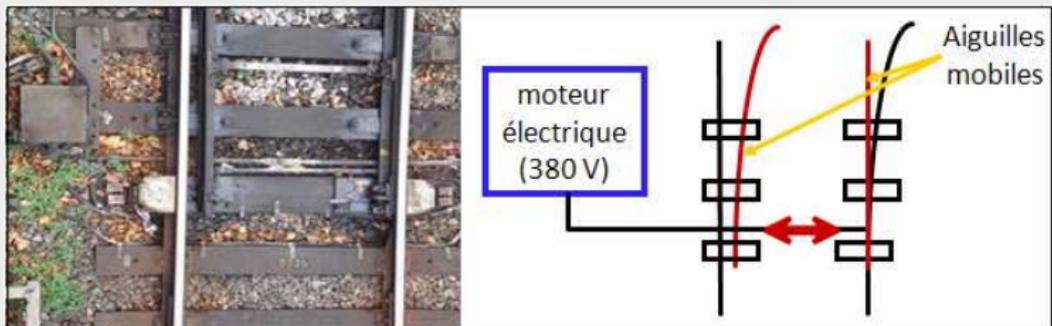
8 Octobre 2009

- 1 Introduction
- 2 Aspects bibliographiques
 - Modèle de régression par morceaux
 - Modèle de régression régi par une chaîne de Markov caché (HMM)
- 3 Modèle de régression à processus latent proposé
- 4 Expérimentations
- 5 Apprentissage à partir de plusieurs signaux
 - Expérimentations
- 6 Conclusion et perspectives

Contexte applicatif

- ▶ Collaboration avec la SNCF
- ▶ Diagnostic et suivi de l'état de fonctionnement du mécanisme d'aiguillage

Mécanisme d'aiguillage



Objectifs visés

- ▶ Détecter précisément le type défaut (défaut mécanique, défaut électrique, défaut de graissage,...) afin de déployer l'équipe concernée
- ▶ Suivi de l'état de fonctionnement :
Estimer la dynamique d'évolution de l'état de fonctionnement du système

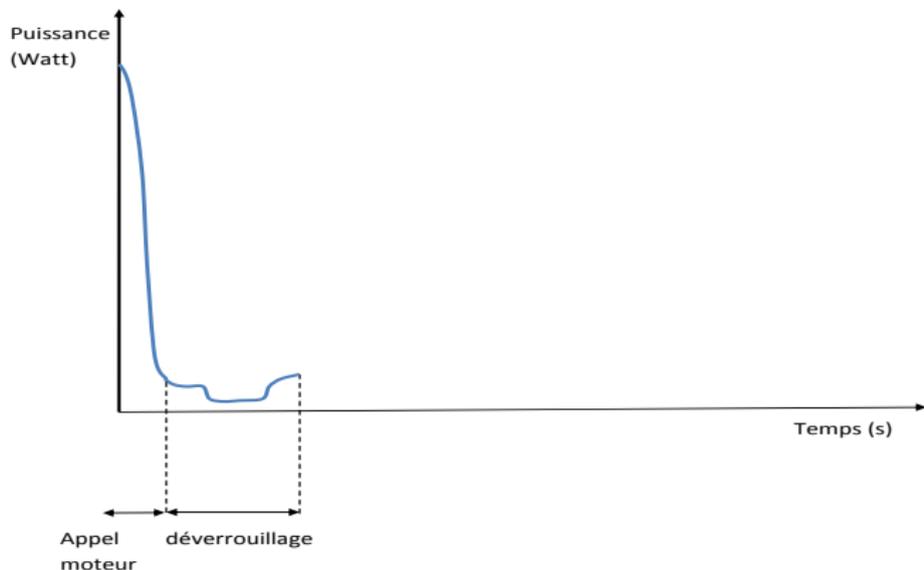
Données exploitées : Puissance consommée par le moteur durant la manoeuvre



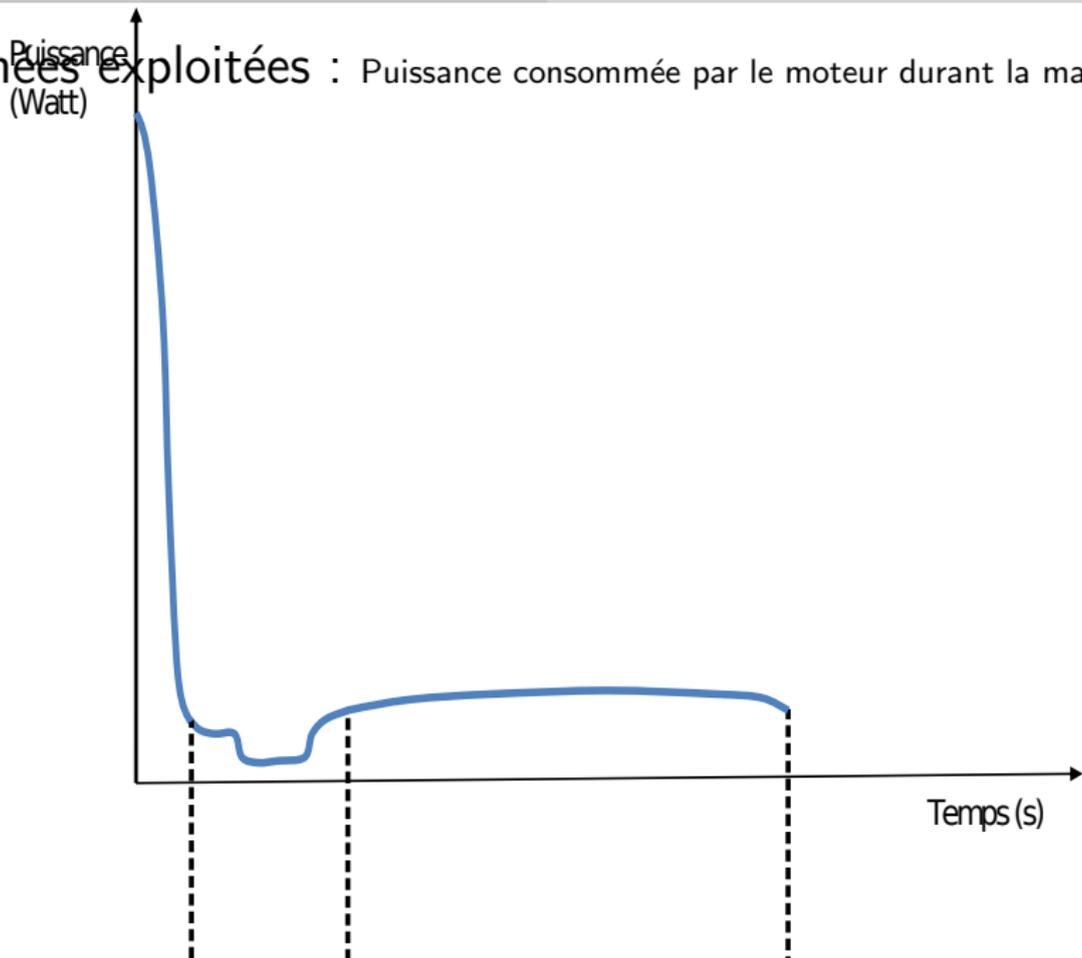
Données exploitées : Puissance consommée par le moteur durant la manoeuvre



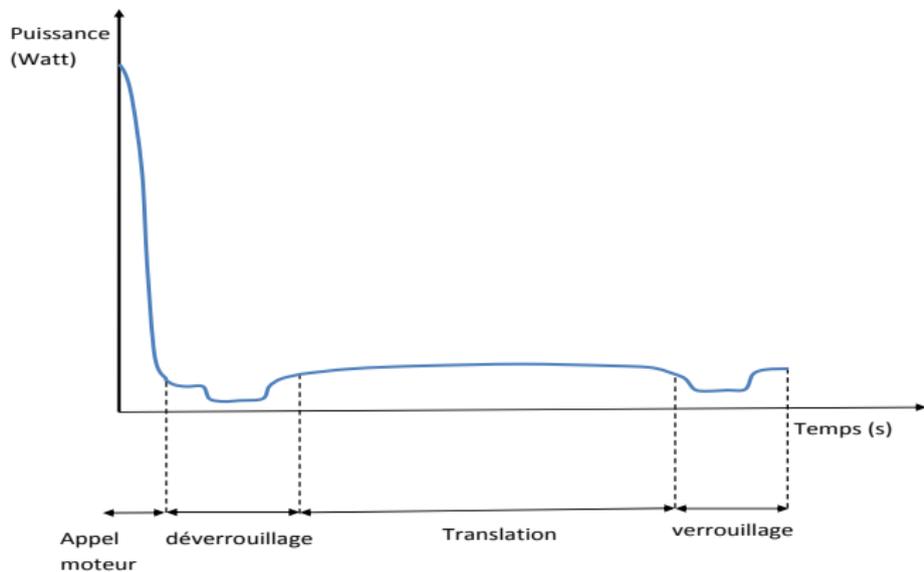
Données exploitées : Puissance consommée par le moteur durant la manoeuvre



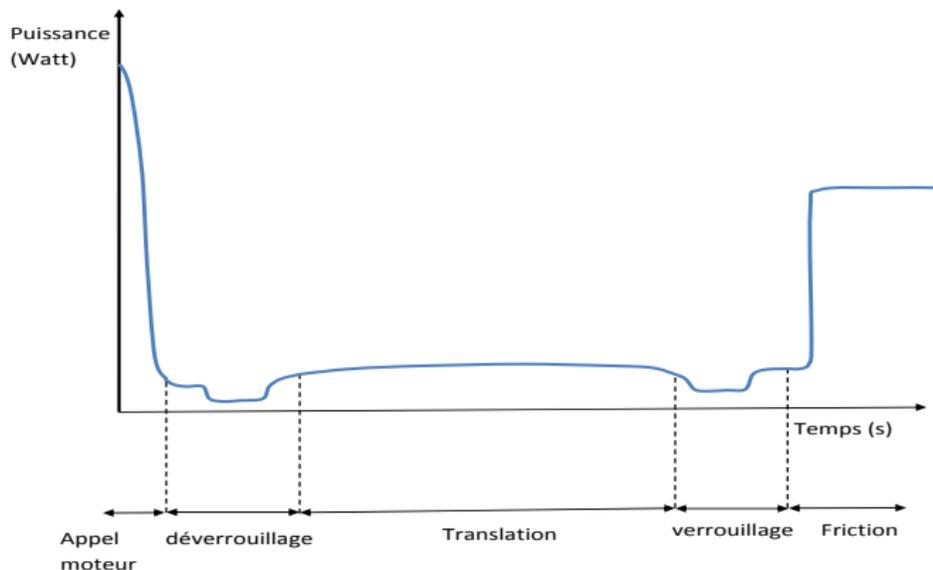
Données exploitées : Puissance consommée par le moteur durant la manoeuvre



Données exploitées : Puissance consommée par le moteur durant la manoeuvre

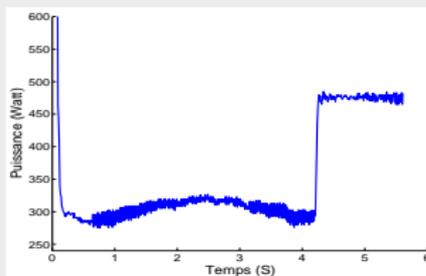


Données exploitées : Puissance consommée par le moteur durant la manoeuvre

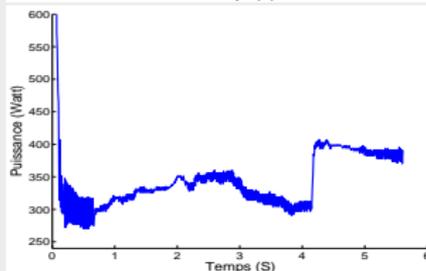


Exemples d'états de fonctionnement de l'aiguillage

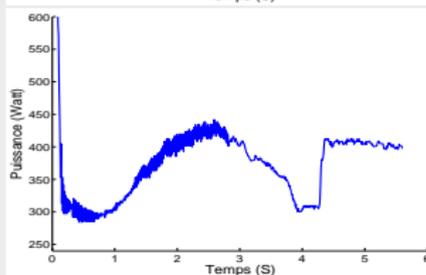
Manœuvre sans défaut



Manœuvre avec défaut mineur

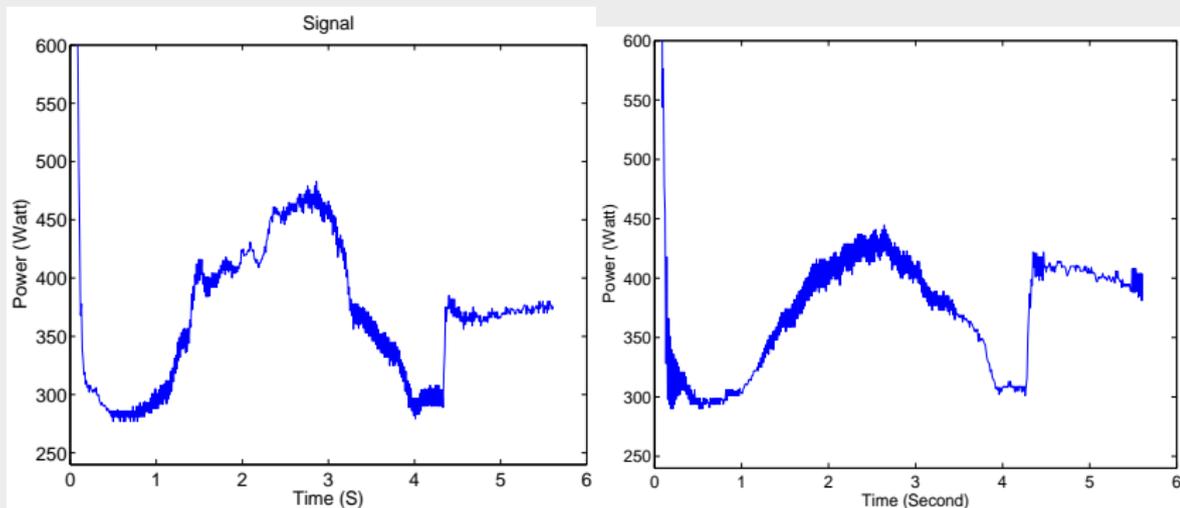


Manœuvre avec défaut critique



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.

Approche proposée

- ① 1ère étape : Caractérisation de chaque signal dans un espace de représentation spécifique
 - Estimation des paramètres d'un modèle probabiliste de signal
 - Avantages :
 - Apte à prendre en compte les différentes phases des signaux étudiés
 - Flexibilité
 - Possibilité de simuler facilement
- ② 2ème étape : Développement d'une méthode de suivi temporel s'appuyant sur les paramètres estimés (2ème partie de la thèse)

Modèle de régression par morceaux [McGee & Carleton 70]

- ▶ Échantillon $\{(x_1, t_1), \dots, (x_n, t_n)\}$
 - x_i : Variable dépendante réelle ;
 - t_i : Variable indépendante représentant le temps
- ▶ Le modèle de régression par morceaux générant le signal \mathbf{x} est :

$$\forall i = 1, \dots, n, \quad x_i = \beta_k^T \mathbf{r}_i + \sigma_k \epsilon_i \quad ; \quad \epsilon_i \sim \mathcal{N}(0, 1)$$

- 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 dans \mathbf{R}^{p+1}
- β_k : vecteur de coefficients de régression $\in \mathbf{R}^{(p+1)}$ pour le $k^{\text{ième}}$ segment

Paramètres du modèle

(ψ, γ) avec $\psi = (\beta_1, \dots, \beta_K, \sigma_1^2, \dots, \sigma_K^2)$ et $\gamma = (\gamma_1, \dots, \gamma_{K+1})$.

Estimation des paramètres du modèle de régression par morceaux

- ▶ Maximiser la vraisemblance de (ψ, γ) ou, par équivalence, minimiser par rapport à (ψ, γ) :

$$J(\psi, \gamma) = \sum_{k=1}^K \sum_{i \in I_k} \left[\log \sigma_k^2 + \frac{(x_i - \beta_k^T \mathbf{r}_i)^2}{\sigma_k^2} \right]$$

- ▶ Optimization globale par l'algorithme de Fisher [Fisher 58] qui est basé sur la programmation dynamique [Bellman 61 ; Lechevallier 90] puisque J est additif sur k

Estimation des paramètres du modèle de régression par morceaux

- ▶ Maximiser la vraisemblance de (ψ, γ) ou, par équivalence, minimiser par rapport à (ψ, γ) :

$$J(\psi, \gamma) = \sum_{k=1}^K \sum_{i \in I_k} \left[\log \sigma_k^2 + \frac{(x_i - \beta_k^T \mathbf{r}_i)^2}{\sigma_k^2} \right]$$

- ▶ Optimization globale par l'algorithme de Fisher [Fisher 58] qui est basé sur la programmation dynamique [Bellman 61 ; Lechevallier 90] puisque J est additif sur k

Approximation et segmentation d'un signal

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

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

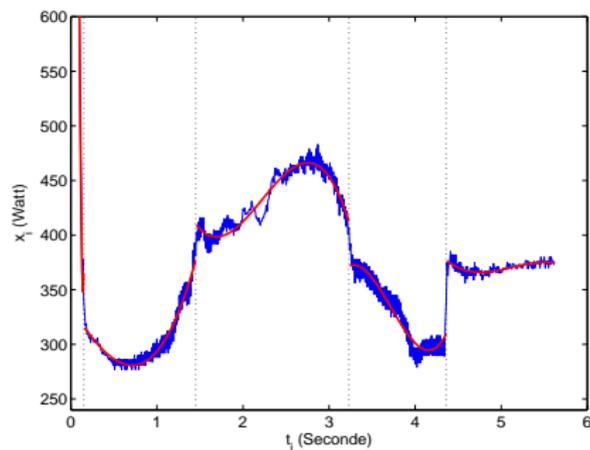


FIGURE – Exemple de résultat graphique obtenu avec le modèle de régression par morceaux

Modèle de régression régi par un HMM [Fridman 93]

Le modèle :

$$\forall i = 1, \dots, n, \quad x_i = \beta_{z_i}^T \mathbf{r}_i + \sigma_{z_i} \varepsilon_i \quad ; \quad \varepsilon_i \sim \mathcal{N}(0, 1),$$

- ▶ z_i une variable latente discrète qui prend ses valeurs dans $\{1, \dots, K\}$ représentant le label du modèle de régression de x_i
- ▶ La séquence cachée $\mathbf{z} = (z_1, \dots, z_n)$ est une chaîne de markov de matrice de transitions A et de loi initiale π

Contraintes sur la matrice de transition

- ▶ Les signaux d'aiguillages sont dus aux mouvements successifs liés à la mécanique du système
- ▶ On suppose une contrainte d'ordre sur les états du HMM

$$p(z_i = k | z_{i-1} = \ell) = 0 \quad \text{if } k < \ell ,$$

et

$$p(z_i = k | z_{i-1} = \ell) = 0 \quad \text{if } k > \ell + 1.$$

⇒ Pas de saut d'état et pas de retour à un état précédent

- ▶ Cas particulier du modèle gauche-droit [Rabiner 89]

Estimation des paramètres du modèle

- ▶ $\Psi = (\pi, A, \beta_1, \dots, \beta_K, \sigma_1^2, \dots, \sigma_K^2)$ vecteur paramètre du modèle
- ▶ Méthode du maximum de vraisemblance
- ▶ log-vraisemblance de Ψ :

$$\begin{aligned} L(\Psi; \mathbf{x}) &= \log p(\mathbf{x}; \Psi) \\ &= \log \sum_{\mathbf{z}} p(z_1; \pi) \prod_{i=2}^n p(z_i | z_{i-1}; A) \prod_{i=1}^n \mathcal{N}(x_i; \beta_{z_i}^T, \sigma_{z_i}^2). \end{aligned}$$

- ▶ Maximisation de la log-vraisemblance par l'algorithme EM (Baum-Welch)

Approximation et segmentation de signaux avec le modèle markovien

Approximation d'un signal

- Somme de polynômes pondérés par les probabilités de filtrage ω_{ik} :

$$\hat{x}_i = \sum_{k=1}^K \omega_{ik}(\hat{\Psi}) \hat{\beta}_k^T \mathbf{r}_i ; i = 1, \dots, n,$$

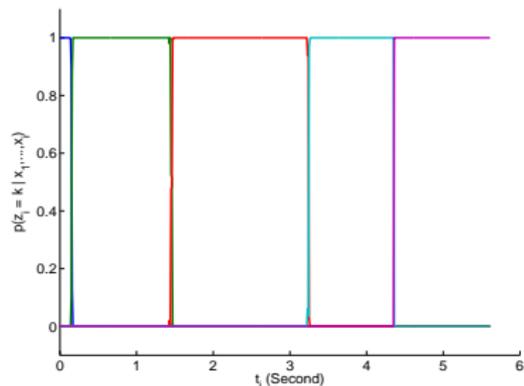
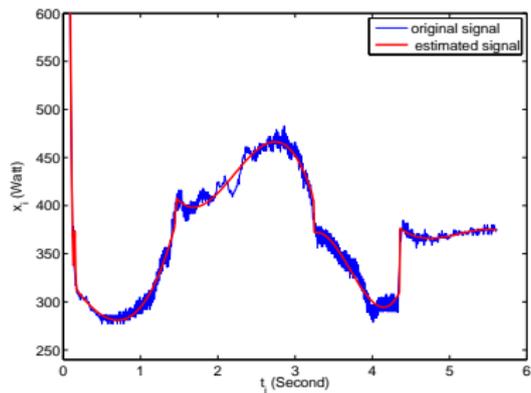
où $\hat{\Psi} = (\hat{\pi}, \hat{A}, \hat{\beta}_1, \dots, \hat{\beta}_K, \hat{\sigma}_1^2, \dots, \hat{\sigma}_K^2)$ et $\omega_{ik}(\hat{\Psi}) = p(z_i = k | x_1, \dots, x_i; \hat{\Psi})$

Segmentation d'un signal

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

$$\hat{z}_i = \arg \max_{1 \leq k \leq K} \hat{\tau}_{ik} ; \forall i = 1, \dots, n, \quad (1)$$

où $\tau_{ik} = p(z_i = k | \mathbf{x}; \Psi)$ est la probabilité a posteriori que x_i soit issu de la *kième* composante régressive.



Probabilités de filtrage

FIGURE – Exemple de résultat graphique obtenu avec le modèle markovien

Modèle de régression à processus latent proposé

Le modèle global

$$\forall i = 1, \dots, n, \quad x_i = \beta_{z_i}^T \mathbf{r}_i + \sigma_{z_i} \epsilon_i \quad ; \quad \epsilon_i \sim \mathcal{N}(0, 1),$$

$\mathbf{z} = (z_1, \dots, z_n)$ est un processus logistique caché

Modèle de régression à processus latent proposé

Le modèle global

$$\forall i = 1, \dots, n, \quad x_i = \beta_{z_i}^T \mathbf{r}_i + \sigma_{z_i} \epsilon_i \quad ; \quad \epsilon_i \sim \mathcal{N}(0, 1),$$

$\mathbf{z} = (z_1, \dots, z_n)$ est un processus logistique caché

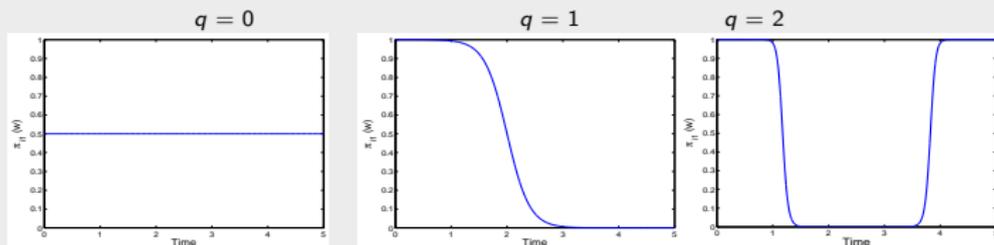
$z_i \sim \mathcal{M}(1, \pi_{i1}(\mathbf{w}), \dots, \pi_{iK}(\mathbf{w}))$; où

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

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

Flexibilité de la transformation logistique : Exemple pour $K = 2$:

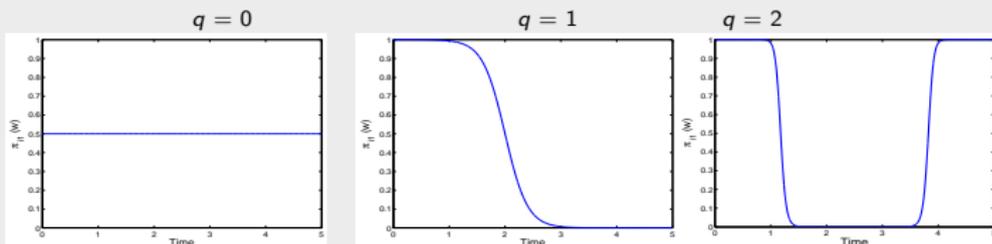
- ① $\pi_{ik}(\mathbf{w})$ en fonction de la dimension q de \mathbf{w}_k :



⇒ $q = 1$ garantit une segmentation en régions contiguës

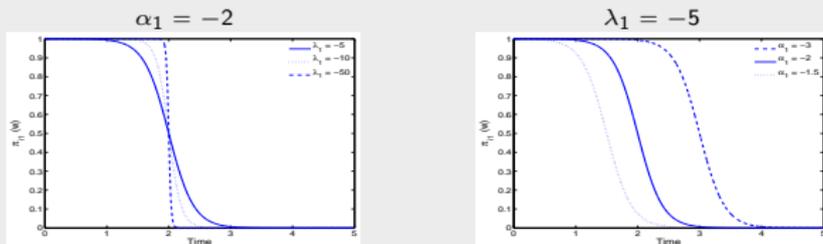
Flexibilité de la transformation logistique : Exemple pour $K = 2$:

- ❶ $\pi_{ik}(\mathbf{w})$ en fonction de la dimension q de \mathbf{w}_k :



$\Rightarrow q = 1$ garantit une segmentation en régions contiguës

- ❷ $\pi_{ik}(\mathbf{w})$ en fonction de \mathbf{w}_k pour $q = 1$. Paramétrisation de \mathbf{w}_k par $\mathbf{w}_k = \lambda_k(\alpha_k, 1)^T$



\Rightarrow Le paramètre λ_k contrôle la qualité des transitions entre les régimes

\Rightarrow Le paramètre α_k contrôle l'instant de transition

Estimation des paramètres du modèle par maximum de vraisemblance

- ▶ Densité mélange

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

- ▶ Vecteur paramètre du modèle :

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

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

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

- ▶ Maximisation de $L(\boldsymbol{\theta}; \mathbf{x})$ par un algorithme Expectation-Maximization (EM) dédié [Dempster et al. 77].

Algorithme EM dédié

Initialisation : $\theta^{(0)}$

Itérer jusqu'à la convergence les deux étapes suivantes :

Algorithme EM dédié

Initialisation : $\theta^{(0)}$

Itérer jusqu'à la convergence les deux étapes suivantes :

❶ **Étape E : Espérance** (à l'itération q)

Calcul de l'espérance conditionnelle de la log-vraisemblance de θ pour les données complétées $L(\theta; \mathbf{x}, \mathbf{z})$

Algorithme EM dédié

Initialisation : $\theta^{(0)}$

Itérer jusqu'à la convergence les deux étapes suivantes :

❶ **Étape E : Espérance** (à l'itération q)

Calcul de l'espérance conditionnelle de la log-vraisemblance de θ pour les données complétées $L(\theta; \mathbf{x}, \mathbf{z})$

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

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

Algorithme EM dédié

Initialisation : $\theta^{(0)}$

Itérer jusqu'à la convergence les deux étapes suivantes :

❶ **Étape E : Espérance** (à l'itération q)

Calcul de l'espérance conditionnelle de la log-vraisemblance de θ pour les données complétées $L(\theta; \mathbf{x}, \mathbf{z})$

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

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

❷ **Étape M : Maximisation** (à l'itération q)

Calcul de $\theta^{(q+1)} = \arg \max_{\theta} Q(\theta, \theta^{(q)})$

Détails de l'étape M

Maximisation séparée de $Q_1(\mathbf{w})$ et $Q_2(\beta_k, \sigma_k^2; k = 1, \dots, K)$:

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

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

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

Maximisation de Q_2 par rapport aux σ_k^2 :

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

Détails de l'étape M

Maximisation séparée de $Q_1(\mathbf{w})$ et $Q_2(\beta_k, \sigma_k^2; k = 1, \dots, K)$:

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

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

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

Maximisation de Q_2 par rapport aux σ_k^2 :

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

- 2 Maximisation de Q_1 par rapport \mathbf{w} : problème convexe de régression logistique multinomiale pondéré \Rightarrow Algorithme IRLS (Iterative Reweighted Least Squares) [Green 84]

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

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

Approximation et segmentation de signaux avec le modèle proposé

Approximation d'un signal

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

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

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

⇒ Adapté pour les changements brusques ou souples entre les différents composants

⇒ 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 entre la courbe de régression estimée et la courbe de régression simulée (erreur de débruitage) :

$$\frac{1}{n} \sum_{i=1}^n [E(x_i; \theta) - \hat{x}_i]^2$$

θ étant le vrai paramètre.

- ▶ Erreur de segmentation
- ▶ Comparaison de l'algorithme proposé avec
 - la méthode de régression par morceaux
 - le modèle de régression avec HMM

Protocole de simulation

Trois expériences effectuées à partir de signaux simulés selon le modèle de régression à processus logistique caché

- 1 Expérience 1 : effet du niveau de souplesse des transitions sur la qualité d'estimation.

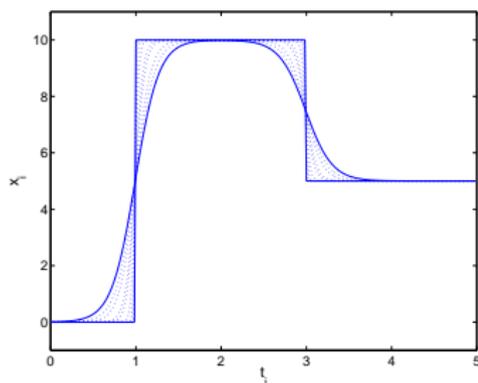
La qualité des transitions entre les différents régimes est contrôlée par les paramètres λ_k de la fonction logistique

- 2 Expérience 2 : effet de la taille d'échantillon n sur la qualité des estimations
- 3 Expérience 3 : effet du niveau de bruit σ sur la qualité des estimations.

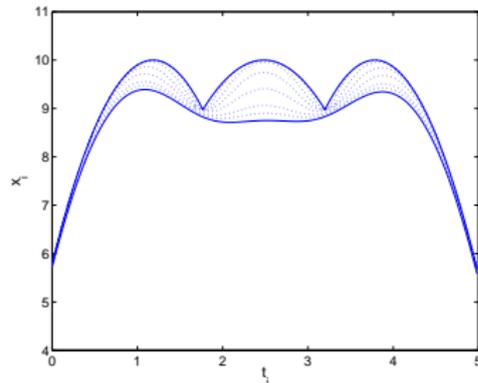
20 jeux de données différents ont été simulés pour chaque situation et les résultats ont été moyennés

Situation 1	$\beta_1 = 0$	$\mathbf{w}_1 = [3341.33, -1706.96]$
	$\beta_2 = 10$	$\mathbf{w}_2 = [2436.97, -810.07]$
	$\beta_3 = 5$	$\mathbf{w}_3 = [0, 0]$
Situation 2	$\beta_1 = [-0.64, 14.4, -6]$	$\mathbf{w}_1 = [3767.58, -1510.19]$
	$\beta_2 = [-21.25, 25, -5]$	$\mathbf{w}_2 = [2468.99, -742.55]$
	$\beta_3 = [-78.64, 45.6, -6]$	$\mathbf{w}_3 = [0, 0]$

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

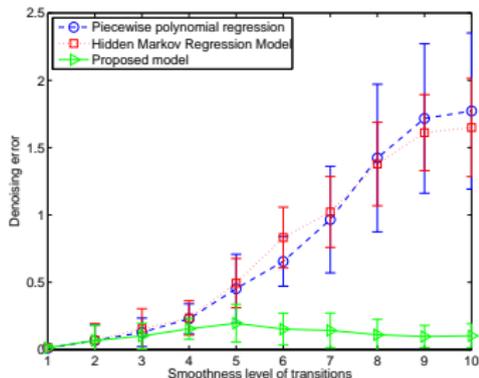


(a)

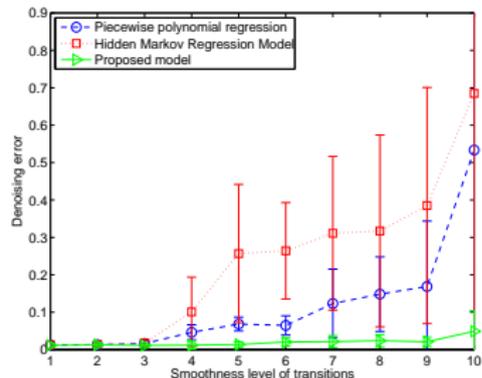


(b)

Résultats : EQM entre la courbe de régression estimée et la courbe de régression simulée en fonction de la qualité des transitions, obtenue pour la situation (a) et la situation (b) :

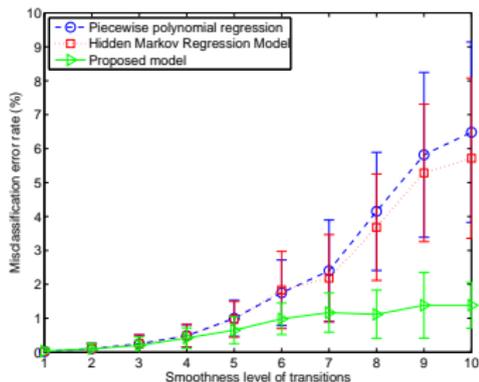


(a)

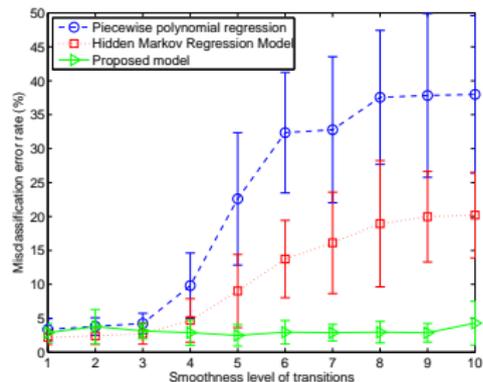


(b)

Résultats : erreur de segmentation en fonction de la qualité des transitions, obtenue pour la situation (a) et la situation (b) :

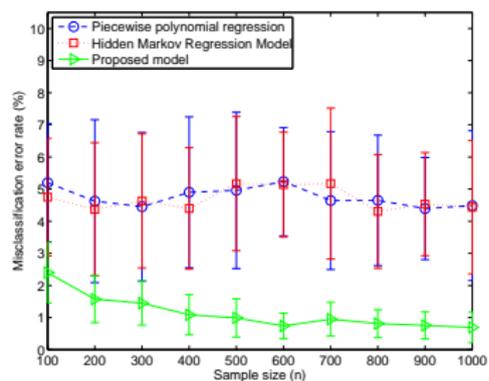
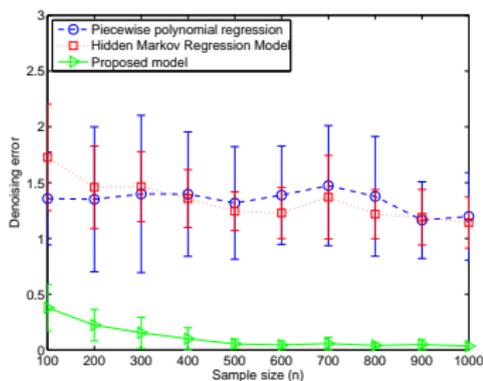


(a)

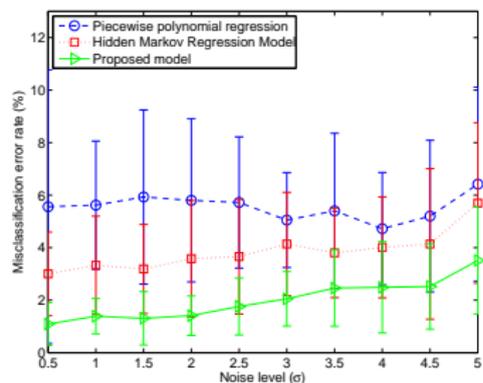
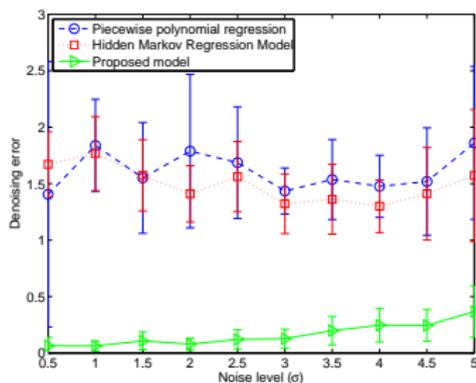


(b)

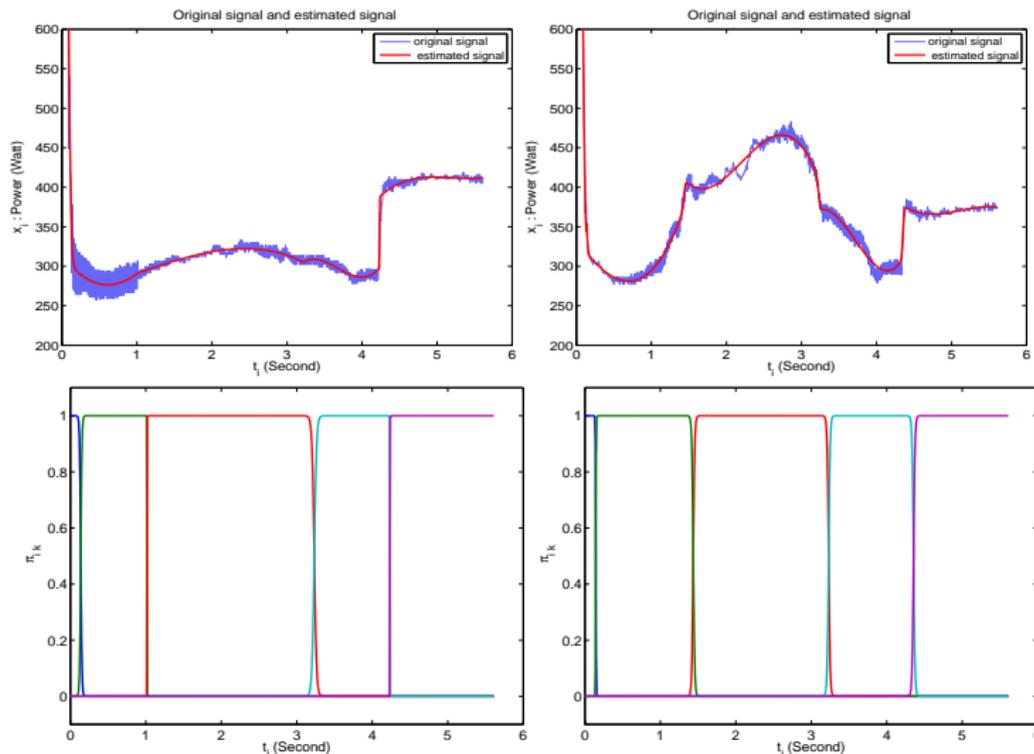
Résultats : EQM entre la courbe de régression estimée et la courbe de régression simulée, et erreur de segmentation en fonction de la taille d'échantillon



Résultats : EQM entre la courbe de régression estimée et la courbe de régression simulée, et erreur de segmentation en fonction du niveau de bruit



Exemple de signaux de manœuvres d'aiguillage : Modèles estimés et fonctions logistiques associées



Classement des signaux de manœuvres d'aiguillage

- ▶ $N = 120$ signaux de manœuvres étiquetés
- ▶ Paramétriser tous les signaux par les 3 approches
- ▶ 3 classes : sans défaut, avec défaut mineur, avec défaut critique
- ▶ Apprentissage des paramètres des classes par MDA (Mixture Discriminant Analysis) [Hastie & Tibshirani 96] sur les caractéristiques des signaux
- ▶ Classement des signaux par Maximum A Posteriori (MAP)

Classement des signaux de manœuvres d'aiguillage

- ▶ $N = 120$ signaux de manœuvres étiquetés
- ▶ Paramétriser tous les signaux par les 3 approches
- ▶ 3 classes : sans défaut, avec défaut mineur, avec défaut critique
- ▶ Apprentissage des paramètres des classes par MDA (Mixture Discriminant Analysis) [Hastie & Tibshirani 96] sur les caractéristiques des signaux
- ▶ Classement des signaux par Maximum A Posteriori (MAP)

Taux de bonne classification (%)

Modèle régression par morceaux	83
Modèle de régression par HMM	89
Modèle de régression proposé	91

Nombre de composant pour chaque
classe sélectionné par BIC

Classe	1	2	3
Nombre de composants	1	1	3

Extension : Apprentissage des paramètres de classes à partir de plusieurs signaux

Modèle dans le cas de plusieurs signaux

▶ Échantillon : $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ n signaux issus de la même classe

▶ $\mathbf{x}_i = \{(x_{i1}, t_1), \dots, (x_{ij}, t_j), \dots, (x_{im}, t_m)\}$, $i = 1, \dots, n$

▶ Modèle générant \mathbf{x}_i :

$$\forall j = 1, \dots, m, \quad x_{ij} = \beta_{z_j}^T \mathbf{r}_j + \sigma_{z_j} \epsilon_{ij} \quad ; \quad \epsilon_{ij} \sim \mathcal{N}(0, 1),$$

avec $z_j \in \{1, \dots, K\}$ est une variable latente discrète et $\mathbf{z} = (z_1, \dots, z_m)$ est un processus logistique.

Estimation des paramètres du modèle

- ▶ Maximum de vraisemblance mis en œuvre par EM

Formules de mise à jour :

- ▶ $\beta_k^{(q+1)} = \arg \min_{\beta_k} \sum_{i=1}^n \sum_{j=1}^m \tau_{ijk}^{(q)} (x_{ij} - \beta_k^T \mathbf{r}_j)^2 = (\Lambda^T \mathbf{W}_k^{(q)} \Lambda)^{-1} \Lambda^T \mathbf{W}_k^{(q)} \mathcal{X}$

Λ matrice de régression et $\mathbf{W}_k^{(q)} = (\tau_{11k}^{(q)}, \dots, \tau_{1mk}^{(q)}, \dots, \tau_{n1k}^{(q)}, \dots, \tau_{nmk}^{(q)})$

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

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

→ IRLS à l'épate M de l'algorithme EM

Évaluation en terme de modélisation de signaux

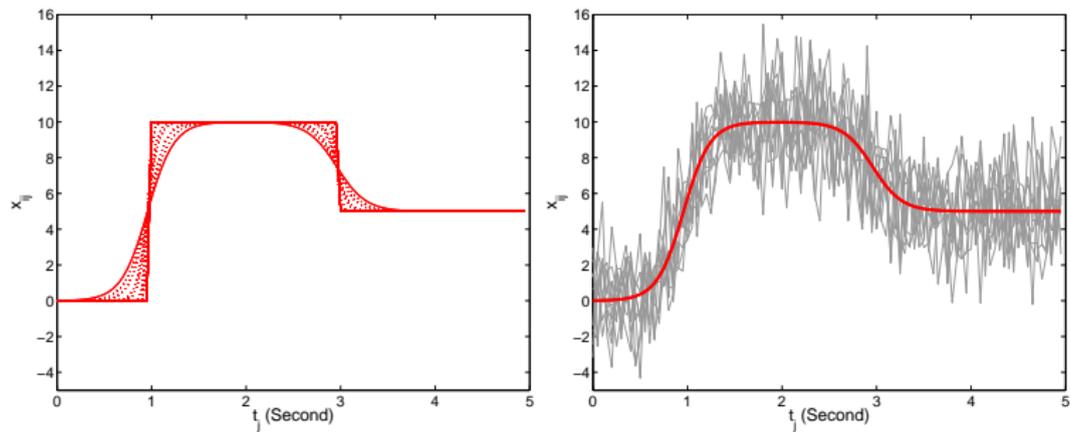
Protocole de simulation

Trois expérimentations :

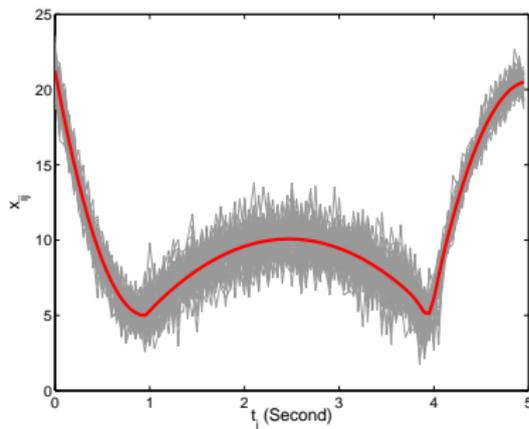
- ❶ Expérience 1 : effet de la qualité des transitions sur les estimations.
- ❷ Expérience 2 : effet du nombre de courbes n par échantillon sur la qualité des estimations
- ❸ Expérience 3 : effet du nombre d'observations m par courbe

20 échantillons différents pour chaque situation ont été simulés et les résultats ont été moyennés

$$\begin{array}{ll} \beta_1 = 0 & \mathbf{w}_1 = [3341.33, -1706.96] \\ \beta_2 = 10 & \mathbf{w}_2 = [2436.97, -810.07] \\ \beta_3 = 5 & \mathbf{w}_3 = [0, 0] \end{array}$$

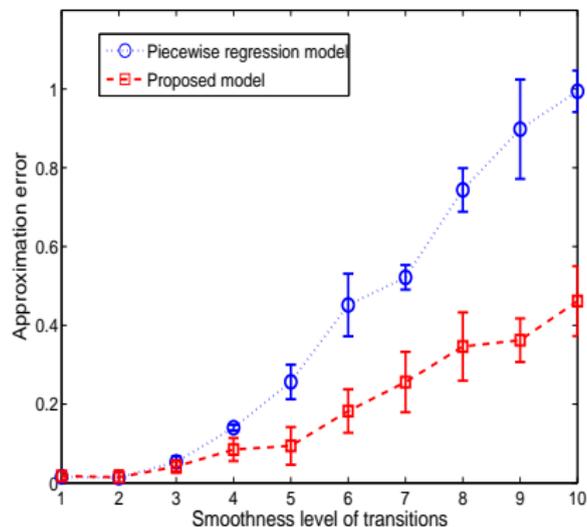
(a) ($n=10$, $m=100$, $\sigma = 2$)

$\beta_1 = [23, -36, 18]$	$\mathbf{w}_1 = [92.72, -46.72]$	$\sigma_1 = 1$
$\beta_2 = [-3.9, 11.08, -2.2]$	$\mathbf{w}_2 = [61.16, -15.28]$	$\sigma_2 = 1.25$
$\beta_3 = [-337, 141.5, -14]$	$\mathbf{w}_3 = [0, 0]$	$\sigma_3 = 0.75$

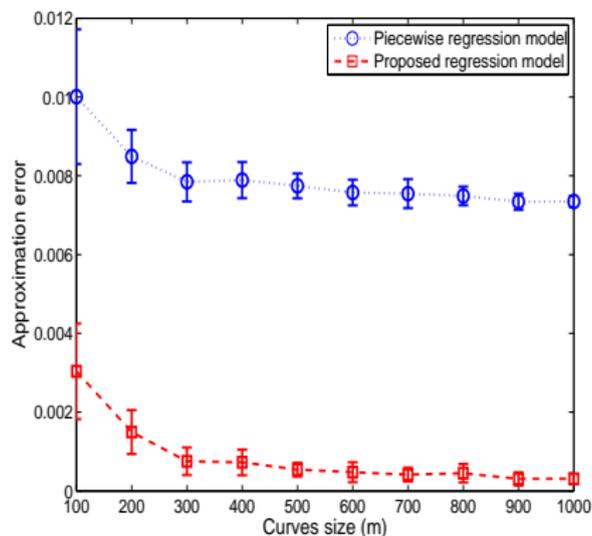


(b)(exemple : $n=50$, $m=100$)

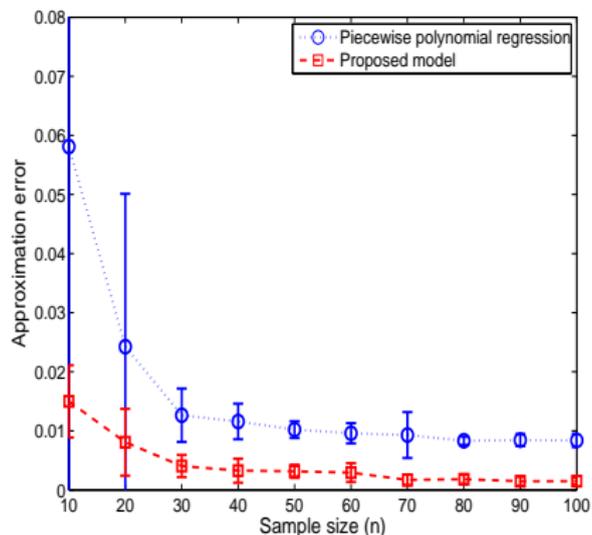
Résultats : erreur d'approximation des courbes en fonction de la qualité des transitions (situation (a))



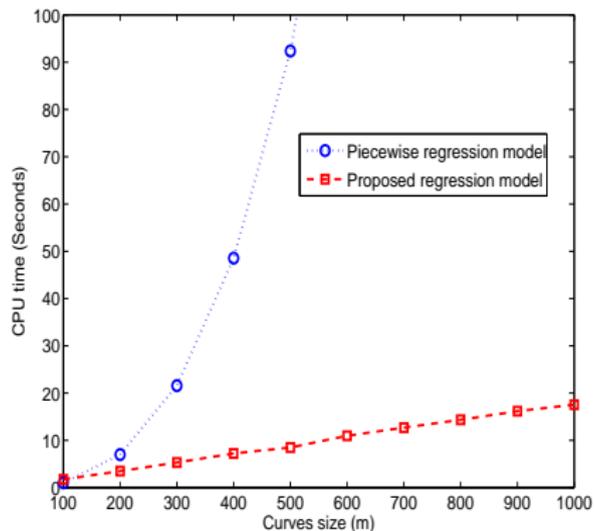
Résultats : erreur d'approximation des courbes en fonction du nombre de points par courbe m pour $n=50$ (situation (b)) :



Résultats : erreur d'approximation des courbes en fonction la taille d'échantillon n pour $m=100$ (situation (b)) :



Résultats : temps de calcul moyen en fonction de m pour $n=50$ (situation (b)) :

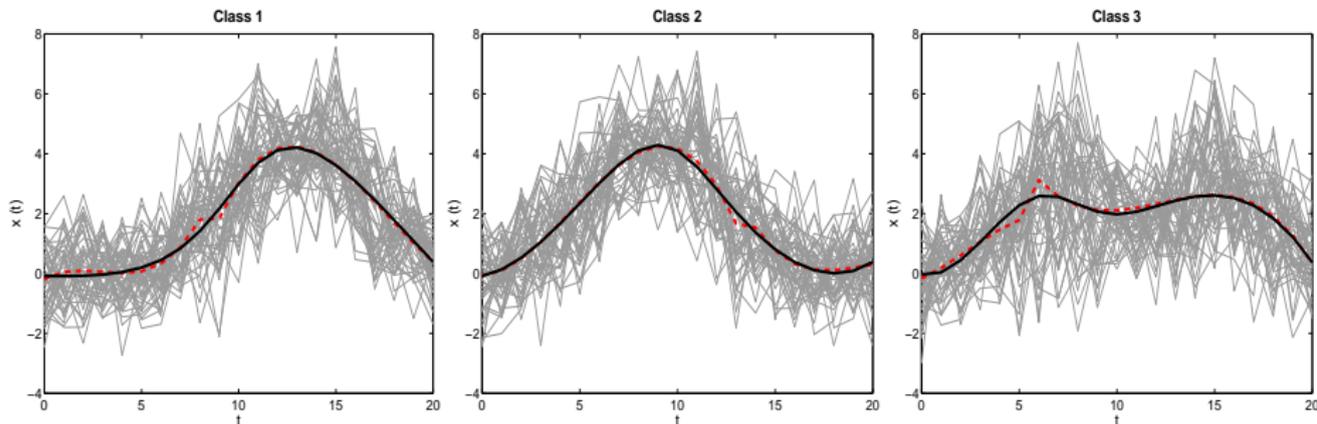


Vagues de Breiman : [Breiman et al. 84, Rossi & Conan-Guez 2005]

- ▶ Problème de 3 classes
- ▶ $m=21$ observations dans chaque courbe

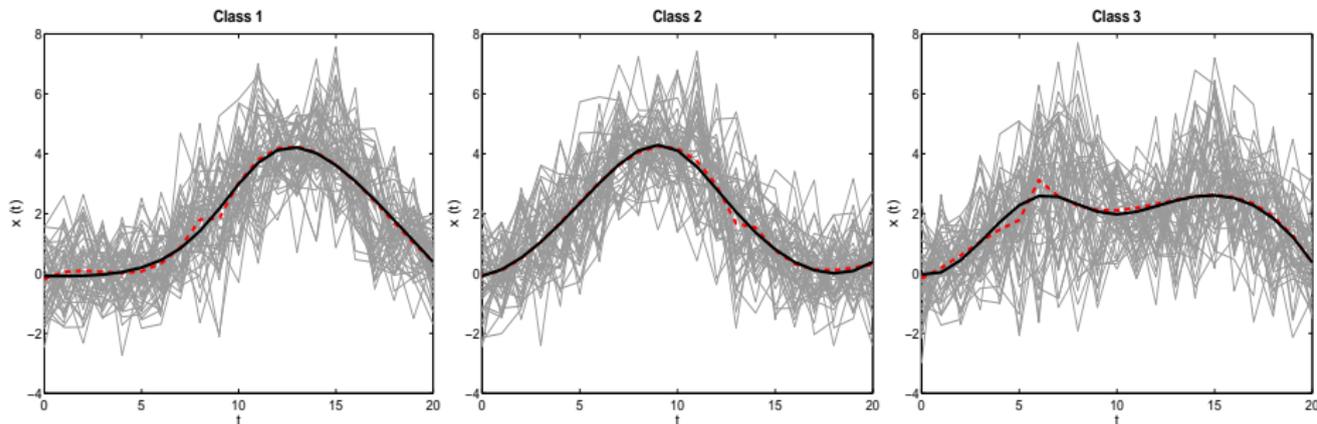
Vagues de Breiman : [Breiman et al. 84, Rossi & Conan-Guez 2005]

- ▶ Problème de 3 classes
- ▶ $m=21$ observations dans chaque courbe
- ▶ Exemple avec 50 courbes par classe ($K=2$ et $P=3$)



Vagues de Breiman : [Breiman et al. 84, Rossi & Conan-Guez 2005]

- ▶ Problème de 3 classes
- ▶ $m=21$ observations dans chaque courbe
- ▶ Exemple avec 50 courbes par classe ($K=2$ et $P=3$)

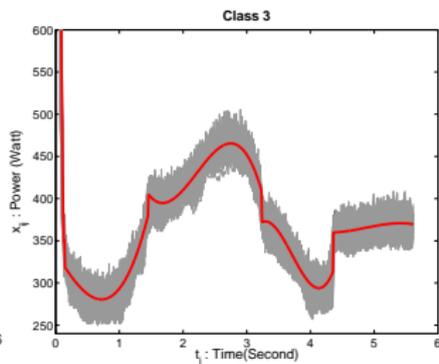
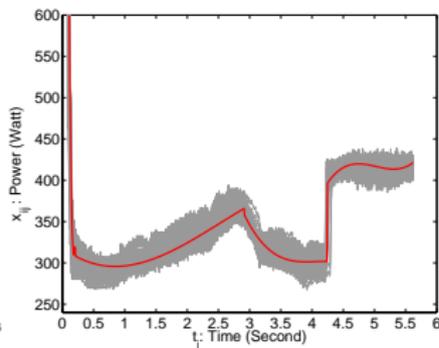
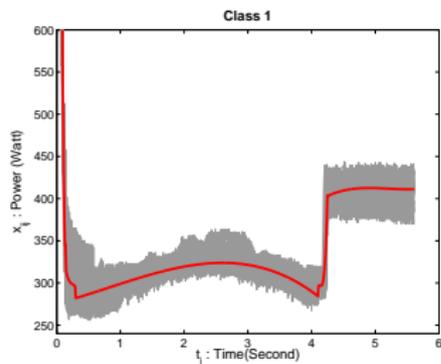


Approche de modélisation	taux d'erreur de test (%)
Régression par morceaux	2.4 (0.64)
Modèle proposé	1.67 (0.84)

TABLE – Taux d'erreur de classement de signaux

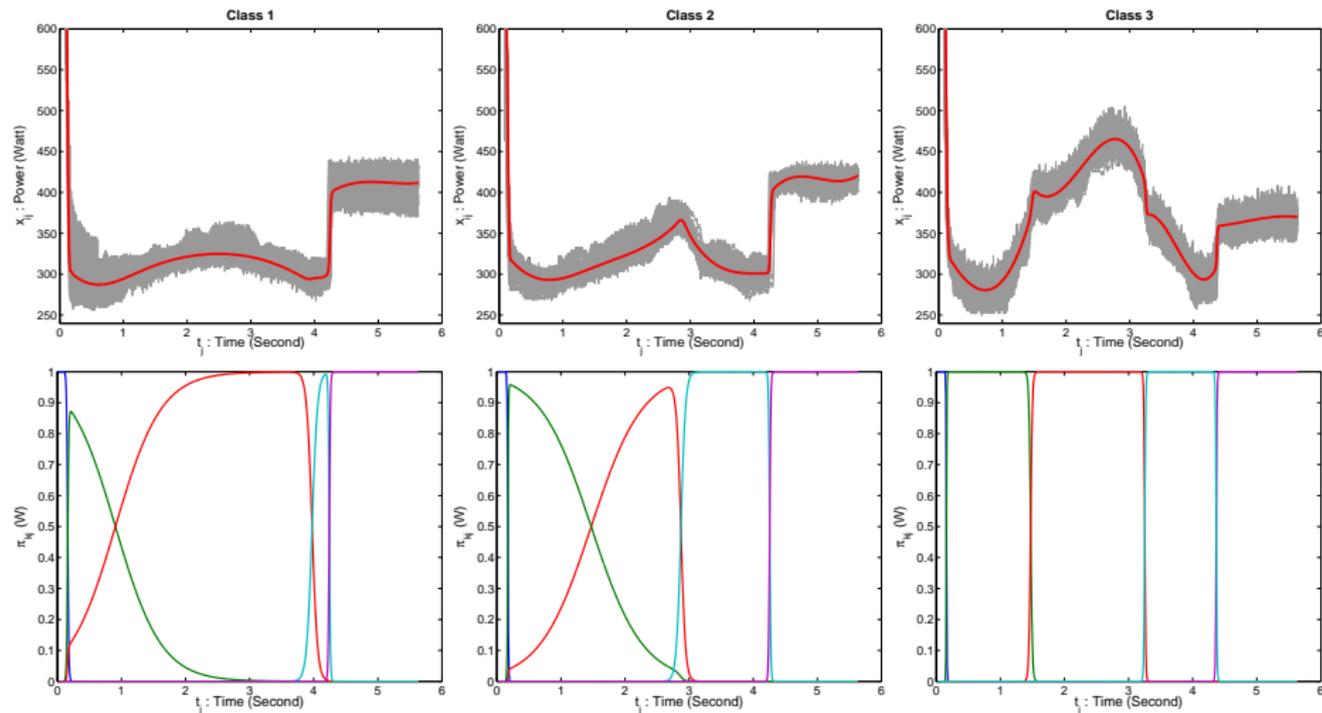
Signaux de manœuvres d'aiguillage

Les 3 classes de signaux de manœuvres d'aiguillage et leurs modèles estimés par le modèle de régression par morceaux :



Signaux de manoeuvres d'aiguillage

Les 3 classes de signaux de manoeuvres d'aiguillage, les modèles estimés et les fonctions logistiques associées obtenus par l'approche de régression proposée :



Approche de modélisation	Test error rates (%)
Régression par morceaux	1.82 (5.74)
Modèle proposé	1.67 (2.28)

TABLE – Taux d'erreur de classement des signaux de manœuvres d'aiguillage.

Conclusion

- ▶ Modèle de régression à processus latent proposé pour modéliser les signaux de manœuvres d'aiguillage.
- ▶ Extension pour apprendre à partir de plusieurs signaux
- ▶ Avantages :
 - Le modèle proposé intègre un processus logistique qui permet de basculer de manière souple ou brusque entre différents sous-modèles polynomiaux
 - L'algorithme EM incluant une procédure de régression logistique fournit un cadre très adapté pour estimer les paramètres du modèle proposé.
- ▶ On obtient des résultats précis en terme de discrimination de signaux

Perspectives

Mise au point d'une approche en ligne pour le suivi de l'état de fonctionnement du système

Pistes :

- ▶ Modèle AR à base de processus latents (HMM, logistique)
- ▶ les observations sont les paramètres caractéristiques des courbes
- ▶ Modélisation de la séquence des états de fonctionnement :
 - Chaîne de Markov
 - Modèle logistique

Merci pour votre attention !