

Modeèle de régression à processus logistique caché pour le diagnostic par suivi de point de fonctionnement des mécanismes d'aiguillages

Faïcel Chamroukhi

Direction de la thèse : Gérard Govaert (UTC), Patrice Aknin(INRETS)

Encadrement : Allou Samé (INRETS)



1 Introduction

- Contexte
- Objectifs visés
- Approche proposée

2 Modélisation probabiliste des signaux

- Modèle de mélange de régressions
- Estimation des paramètres du modèle

3 Expérimentations

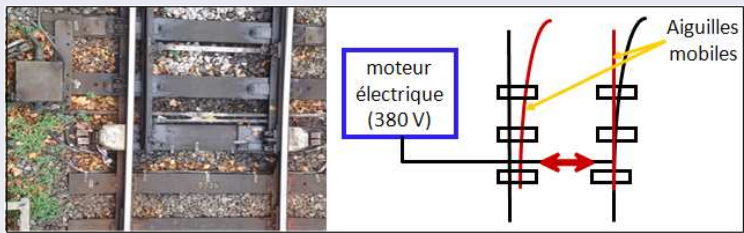
4 Conclusion

5 Perspectives

Contexte

- Collaboration avec la SNCF
- Diagnostic par suivi de point de fonctionnement du mécanisme d'aiguillage
- Application des méthodes développées au circuit de voie

Mécanisme d'aiguillage

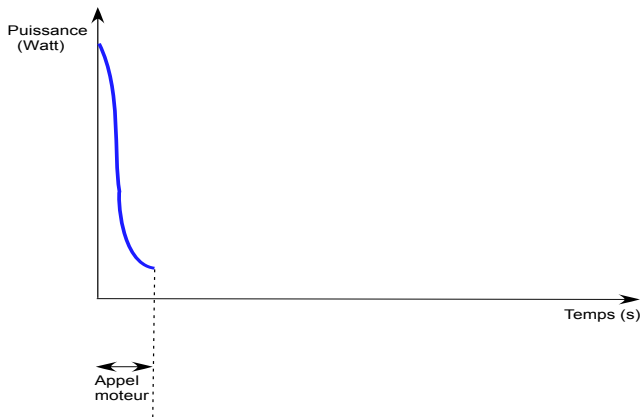


Objectifs visés

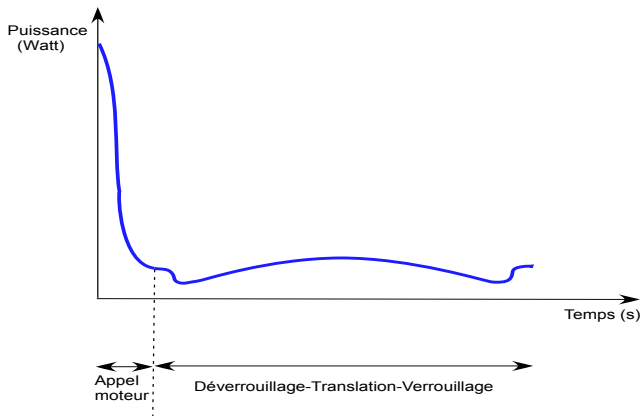
- Suivi du point de fonctionnement :
 - Estimer la dynamique d'évolution de l'état de fonctionnement du système
- Détecter précisément le type défaut (défaut mécanique, défaut électrique, défaut de graissage,...) afin de déployer l'équipe concernée

Données disponibles pour le diagnostic : Puissance consommée par le moteur durant la manœuvre

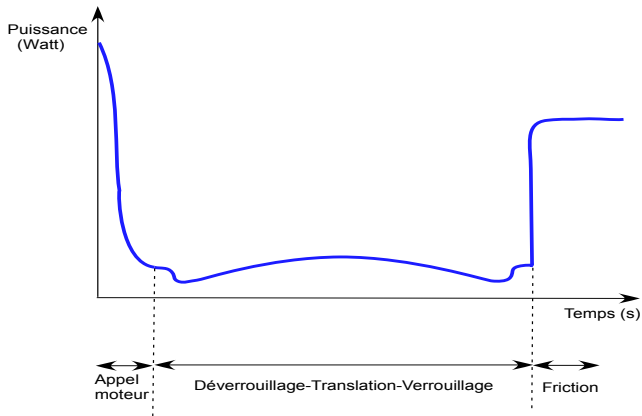
Données disponibles pour le diagnostic : Puissance consommée par le moteur durant la manœuvre



Données disponibles pour le diagnostic : Puissance consommée par le moteur durant la manœuvre

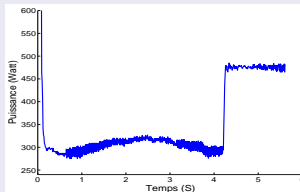


Données disponibles pour le diagnostic : Puissance consommée par le moteur durant la manœuvre

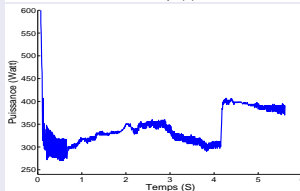


Exemples d'états de fonctionnement de l'aiguillage

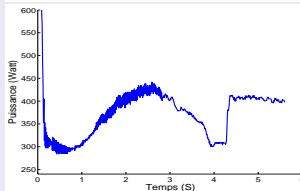
Manœuvre sans défaut



Manœuvre avec défaut mineur

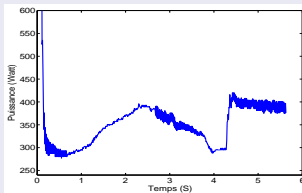


Manœuvre avec défaut

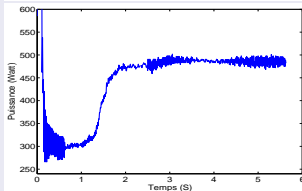


Variabilité des signatures des défauts

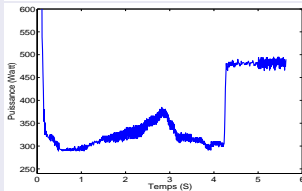
Défaut de graissage



Friction anticipée



Mauvais réglage de l'attaque multiple



Approche proposée

- 1 ère 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 ème étape : Développement d'une **méthode de suivi temporel** s'appuyant sur les paramètres estimés (2ème partie de la thèse)

Modèle de régression à variable Latente/ mélange de régressions

- Signal

- échantillon : $\{(t_1, x_1), \dots, (t_n, x_n)\}$
- t_i : le temps
- x_i : variable aléatoire réelle représentant le signal à l'instant t_i

- Modèle proposé

$$x_i = \beta_{z_i}^T \mathbf{t}_i + \varepsilon_i \quad ; \quad \varepsilon_i \sim \mathcal{N}(\mathbf{0}, \sigma_{z_i}^2) \quad \forall i = 1, \dots, n \quad (1)$$

- $z_i \in \{1, \dots, K\}$ **variable latente discrète** : le label du sous-modèle de régression de x_i
- $\Rightarrow z_i$ permet de basculer de façon automatique d'un modèle régressif à l'autre
- $\mathbf{t}_i = (1, t_i, \dots, t_i^p)$; p ordre de régression
- $\beta_k = (\beta_{k,1}, \dots, \beta_{k,p+1})$: coefficients de régression

Loi de probabilité des observations $(x_i, i = 1, \dots, n)$

En faisant intervenir **la variable latente discrète z** représentant la classe de chaque observation x_i , on montre que :

$$\begin{aligned}
 p(x_i; \theta) &= \sum_{k=1}^K p(x_i, z_i = k; \theta) \\
 &= \sum_{k=1}^K p(z_i = k) p(x_i | z_i = k; \alpha_k) \\
 &= \sum_{k=1}^K \pi_{ki} \mathcal{N}(x_i; \beta_k^T \mathbf{t}_i, \sigma_k^2) \quad \leftarrow \text{modèle de mélange de lois}
 \end{aligned}$$

Paramètres du modèle

$$\theta = \left(\pi_{11}, \dots, \pi_{ki}, \dots, \pi_{Kn}, \beta_1, \dots, \beta_K, \sigma_1^2, \dots, \sigma_K^2 \right)$$

Modélisation des proportions du mélange (π_{ki})

- Moyen de réduire le nombre de proportions à estimer
- Les proportions conditionnelles du mélange sont modélisées par un modèle de régression logistique [CAR 06] :

$$\pi_{ki}(\mathbf{w}) = \frac{\exp(\mathbf{w}_k^T \mathbf{t}_i)}{\sum_{\ell=1}^K \exp(\mathbf{w}_\ell^T \mathbf{t}_i)} \quad (2)$$

avec $\dim(\mathbf{w}_k) = \dim(\mathbf{t}_i) = p$
 $\sum_{k=1}^K \pi_{ki}(\mathbf{w}) = 1 \quad \forall i$

- Modèle concurrent : z_i distribuée suivant une chaîne de Markov.

⇒ Paramètres du modèle

$$\theta = \left(\mathbf{w}_1, \dots, \mathbf{w}_K, \beta_1, \dots, \beta_K, \sigma_1^2, \dots, \sigma_K^2 \right)$$

Estimation du vecteur θ des paramètres du modèle

- Méthode du maximum de vraisemblance

- $x_i = \beta_{z_i}^T \mathbf{t}_i + \varepsilon_i$

- $\varepsilon_1, \dots, \varepsilon_n$ sont supposés indépendants cond. aux z_i et de loi $\mathcal{N}(0, \sigma_{z_i}^2)$

$$\Rightarrow P(x_i | z_i = k; \theta) = \mathcal{N}(x_i; \beta_{z_i}^T \mathbf{t}_i, \sigma_{z_i}^2)$$

- Log-vraisemblance

$$\begin{aligned} L(\theta) &= \log P(x_1, \dots, x_n; \theta) \\ &= \sum_{i=1}^n \log \left[\sum_{k=1}^K \pi_{ki}(\mathbf{w}) \mathcal{N}(x_i; \beta_k^T \mathbf{t}_i, \sigma_k^2) \right] \end{aligned}$$

La maximisation directe de la log-vraisemblance ne peut pas être effectuée de manière analytique

- Maximisation par **l'algorithme EM (Expectation Maximization)**
[DEM 77 , GOV 03]

Algorithme EM

Maximiser à chaque itération $Q(\theta; \theta^{(q)}) = E(L_c(\theta; \mathbf{Y}) | \mathbf{x}, \theta^{(q)})$ avec :

- données observées : $\mathbf{x} = (x_1, \dots, x_n)$
- données complétées : $\mathbf{y} = (x_1, \dots, x_n, z_1, \dots, z_n)$
- $L_c(\theta; \mathbf{y})$: log-vraisemblance complétée (plus facile à maximiser)

$$L_c(\theta; \mathbf{y}) = \log P(x_1, \dots, x_n, z_1, \dots, z_n | \theta)$$

$$L(\theta; \mathbf{x}) = \underbrace{E(L_c(\theta; \mathbf{Y}) | \mathbf{x}, \theta^{(q)})}_{Q(\theta; \theta^{(q)})} - \underbrace{E(\log P(\mathbf{Y} | \mathbf{x}, \theta) | \mathbf{x}, \theta^{(q)})}_{H(\theta; \theta^{(q)})}$$

⇒ La maximisation de $Q(\theta; \theta^{(q)})$ à chaque itération q fait croître la vraisemblance $L(\theta; \mathbf{x})$

Détails de l'algorithme EM obtenu

Détails de l'algorithme EM obtenu

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

Détails de l'algorithme EM obtenu

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

Itérer jusqu'à la convergence les deux étapes suivantes :

Détails de l'algorithme EM obtenu

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

Itérer jusqu'à la convergence les deux étapes suivantes :

- 1 Étape E : Espérance (à l'itération q)

Calcul de l'espérance conditionnelle de la vraisemblance complétée $Q(\theta; \theta^{(q)})$

⇒ nécessite simplement le calcul des probabilités a posteriori $t_{ik}^{(q)}$, $\forall i, \forall k$

$$t_{ik}^{(q)} = \frac{\pi_{ki}(\mathbf{w}^{(q)}) \mathcal{N}(x_i; \beta_k^T \mathbf{t}_i, \sigma_k^2(q))}{\sum_{\ell=1}^K \pi_{\ell i}(\mathbf{w}^{(q)}) \mathcal{N}(x_i; \beta_{\ell}^T \mathbf{t}_i, \sigma_{\ell}^2(q))}$$

Détails de l'algorithme EM obtenu

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

Itérer jusqu'à la convergence les deux étapes suivantes :

- 1 Étape E : Espérance (à l'itération q)

Calcul de l'espérance conditionnelle de la vraisemblance complétée $Q(\theta; \theta^{(q)})$

⇒ nécessite simplement le calcul des probabilités a posteriori $t_{ik}^{(q)}$, $\forall i, \forall k$

$$t_{ik}^{(q)} = \frac{\pi_{ki}(\mathbf{w}^{(q)}) \mathcal{N}(x_i; \beta_k^T \mathbf{t}_i, \sigma_k^2(q))}{\sum_{\ell=1}^K \pi_{\ell i}(\mathbf{w}^{(q)}) \mathcal{N}(x_i; \beta_{\ell}^T \mathbf{t}_i, \sigma_{\ell}^2(q))}$$

- 2 Étape M : Maximisation (à l'itération q)

Calcul de $\theta^{(q+1)}$ qui maximise $Q(\theta; \theta^{(q)})$ par rapport au paramètre θ

Détails de l'étape M

Maximisation de $Q(\theta; \theta^{(q)}) = Q_1(\mathbf{w}) + Q_2(\beta, \sigma_1^2, \dots, \sigma_K^2)$:

Détails de l'étape M

Maximisation de $Q(\theta; \theta^{(q)}) = Q_1(\mathbf{w}) + Q_2(\beta, \sigma_1^2, \dots, \sigma_K^2)$:

- Calcul du $\mathbf{w}^{(q+1)}$ qui maximise Q_1 :
 - problème de régression logistique multinomial pondéré par les $t_{ik}^{(q)}$ (Algorithme IRLS (Iterative Reweighted Least Square [GRE 84, KRI 05])

Détails de l'étape M

Maximisation de $Q(\theta; \theta^{(q)}) = Q_1(\mathbf{w}) + Q_2(\beta, \sigma_1^2, \dots, \sigma_K^2)$:

- Calcul du $\mathbf{w}^{(q+1)}$ qui maximise Q_1 :
 - problème de régression logistique multinomial pondéré par les $t_{ik}^{(q)}$ (Algorithme IRLS (Iterative Reweighted Least Square [GRE 84, KRI 05])

- Calcul du $(\beta_1, \dots, \beta_K)$ qui maximise Q_2 :

- K problèmes de régression linéaire pondérés par les $t_{ik}^{(q)}$
- $$(\beta_k^T)^{(q+1)} = \underset{\beta_k}{\operatorname{argmin}} \sum_{i=1}^n t_{ik}^{(q)} (x_i - \beta_k^T \mathbf{t}_i)^2 \quad ; \quad \forall k = 1, \dots, K$$

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

Premières séries d'expérimentations

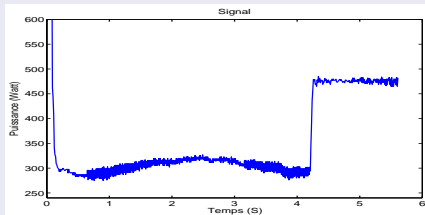
Données réelles

- Signaux réels couvrant un large panel de défauts
- Nombre de composantes du mélange choisi :
 $K = 4 \simeq$ nombre de phases d'une manœuvre
- Ordre de régression linéaire $p = 3$ (splines cubiques)
- Lancer de l'algorithme sur chaque signal
nombre de paramètres à estimer par signal : $K * (2p + 2) = 32$
(Vs. 560 pts initiaux)

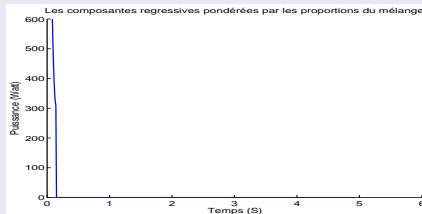
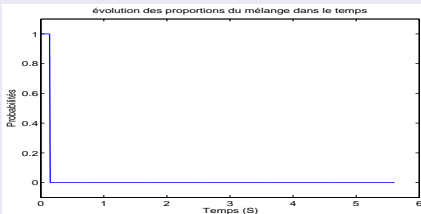
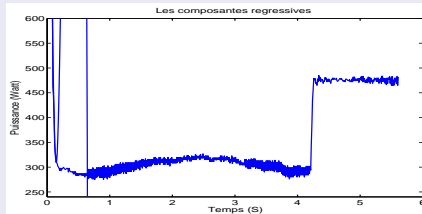
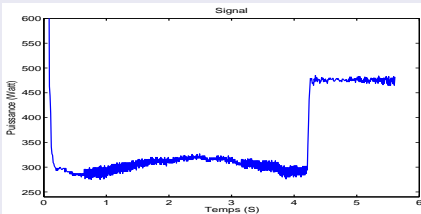
Données simulées

- Générer des signaux à partir de paramètres $\hat{\theta}$ estimés par notre algorithme, et les comparer avec les vrais signaux
- Qualité des estimateurs mesurée par l'erreur de segmentation par rapport à la segmentation de référence (simulée)

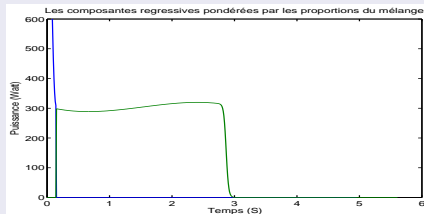
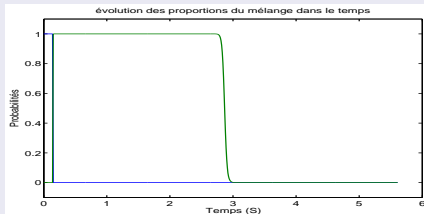
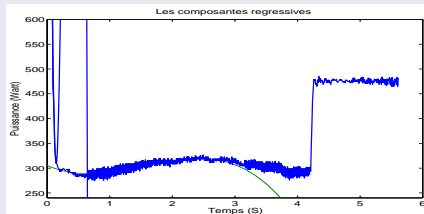
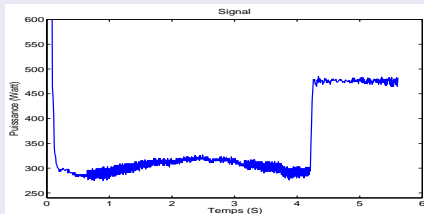
Cas d'un signal sans défaut



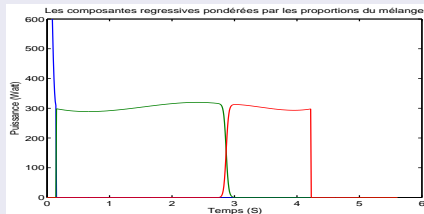
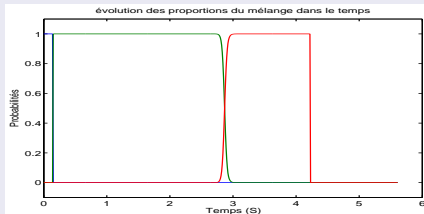
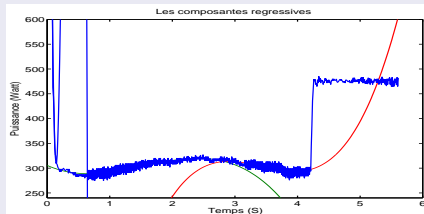
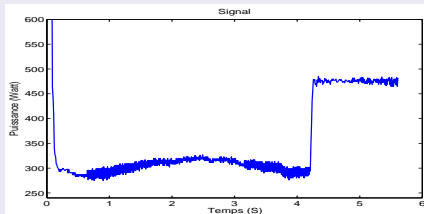
Cas d'un signal sans défaut



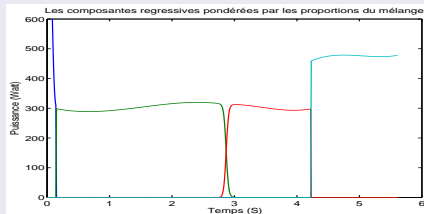
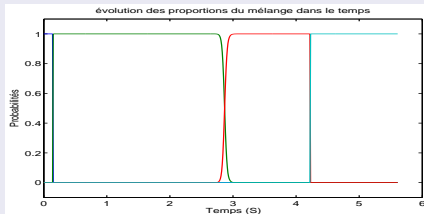
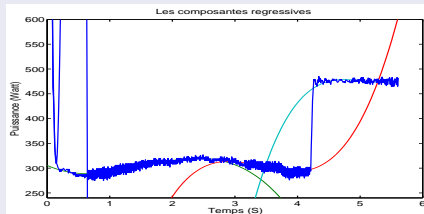
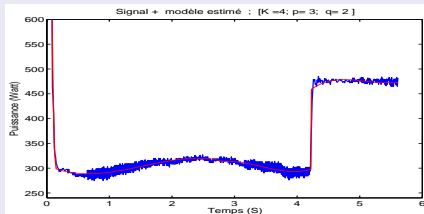
Cas d'un signal sans défaut



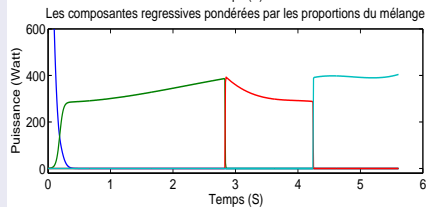
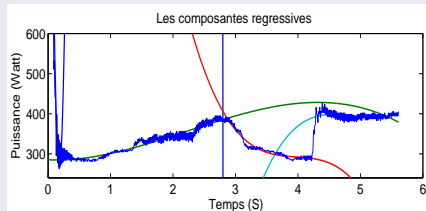
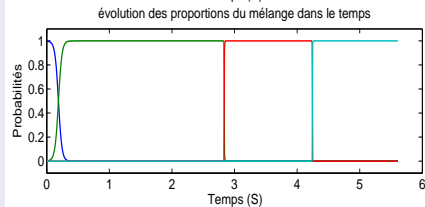
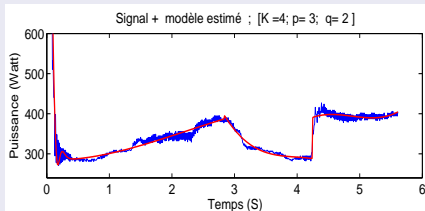
Cas d'un signal sans défaut



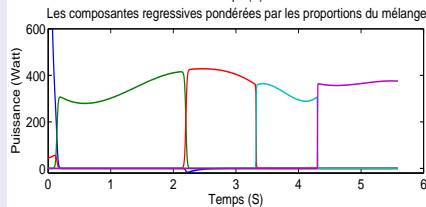
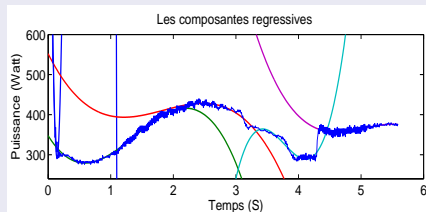
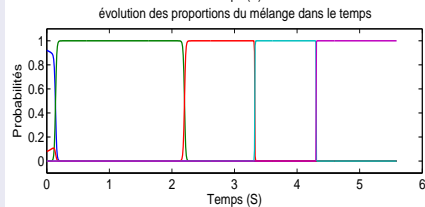
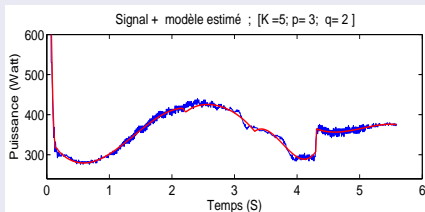
Cas d'un signal sans défaut



Cas d'un signal avec défaut léger

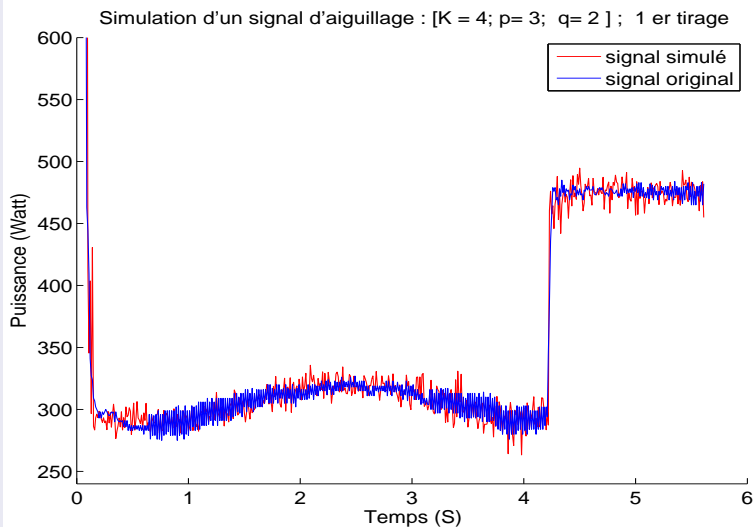


Cas d'un signal avec défaut critique

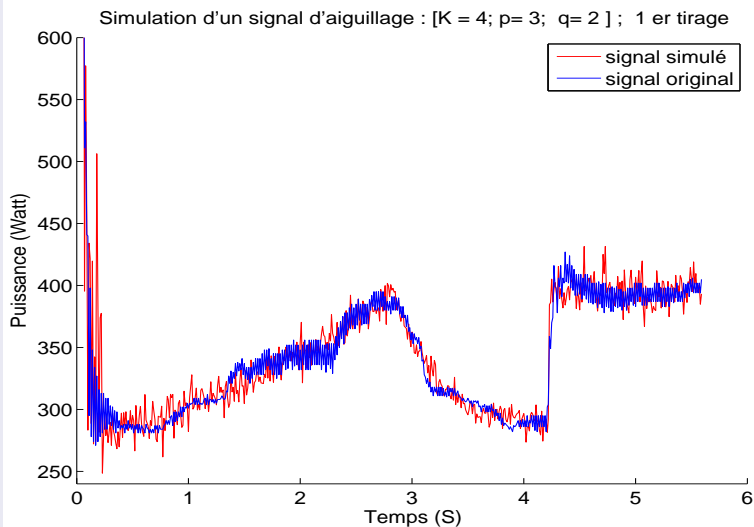


Génération des signaux

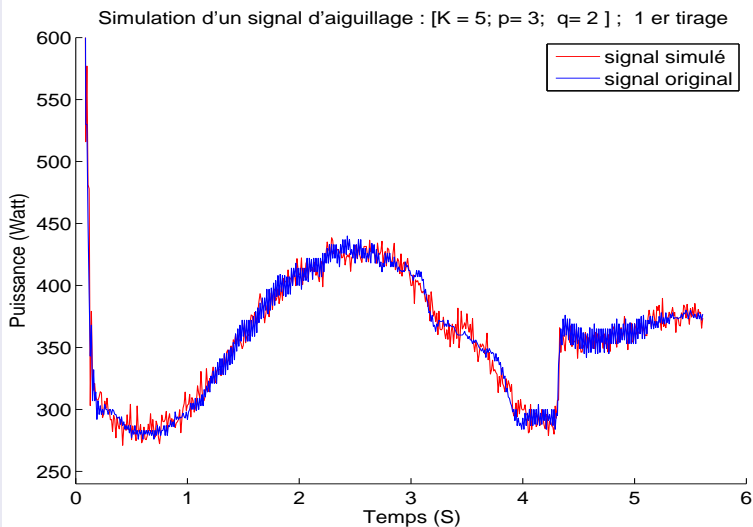
Génération des signaux



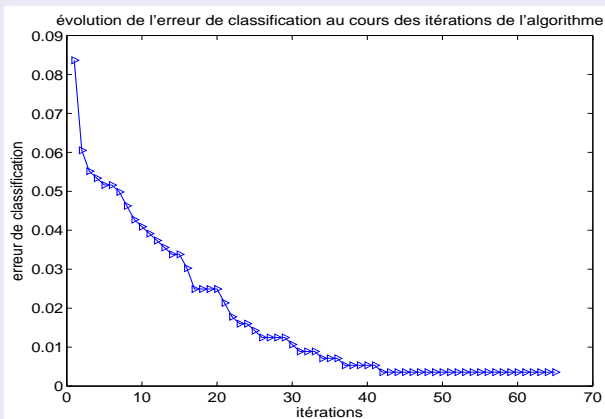
Génération des signaux



Génération des signaux



Résultats : erreur de classification au cours des itérations (en pourcentage de mal classés)



Erreur (à la convergence) moyennée sur 20 tirages de signaux : **2.3 %**

Conclusion

Avantages du modèle probabiliste de mélange de régression par rapport à une approche de régression classique :

- Autorise les paramètres régression **à évoluer aux cours du temps**
→ meilleure modélisation des signaux
- Particularité de l'algorithme de type EM permettant d'estimer les paramètres du modèle :
 - inclut une procédure de régression logistique pour le calcul des proportions du mélange, qui évoluent au cours du temps
 - moyennes des composantes : courbes polynomiales
- Premiers résultats encourageants

Perspectives

- 1 Mise au point d'un protocole de simulation :
 - évaluer l'approche
 - optimiser la complexité du mélangePistes : BIC, AIC, ICL

- 2 Étude comparative avec d'autres approches
 - Approche markovienne (Z_i suit une chaîne de Markov)
 - Modèle autorégressif à variable latente

- 3 Suivi de point de fonctionnement
 - Une observation (signal) sera désormais représentée par le vecteur θ de ses paramètres
 - Approche régressive, HMM



P. Dempster, N. M. Laird, and D. B. Rubin.

Maximum likelihood from incomplete data via the EM algorithm (with discussion).
Journal of the Royal Statistical Society B , pages 39, PP. 138, 1977.



G. Govaert

Analyse des données.
Hermès, Lavoisier, 2003.



A. X. Carvalho and M. A. Tanner.

Modeling nonlinearities with mixtures of-experts of time series models.
International Journal of Mathematics and Mathematical Sciences, 2006.



Krishnapuram, B. ; Carin, L. ; Figueiredo, M.A.T. ; Hartemink, A.J.

Sparse multinomial logistic regression : fast algorithms and generalization bounds
IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 27, Issue 6, Page(s) : 957 - 968
June 2005



P. Green.

Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives.
Journal of the Royal Statistical Society B ,46(2), pp149-192, 1984.



G. Celeux, J. S. Marques and J. Nascimento

Learning Switching Dynamic Models for Objects Tracking.
Rapport de recherche INRIA, Juillet 2003.



Dimitrios V. Paliouras

Comparing regime-switching models in time series : Logistic mixtures vs. Markov Switching
Master Thesis, University of Maryland, Department of mathematics, 2007

Publications

- *"Switch mechanism diagnosis using a pattern recognition approach"*
The 4th IET International Conference on Railway Condition Monitoring RCM 2008, 18 - 20 June 2008, Derby, UK.
- *"Détection séquentielle de défauts sur des signaux de manœuvres d'aiguillage"*
Workshop Surveillance, Sûreté et Sécurité des Grands Systèmes 3SGS'08 - 4 et 5 juin 2008. Université de Technologie de Troyes.
- *"Régression à variable latente pour la modélisation de signaux de manœuvres d'aiguillage"*
à paraître dans les actes INRETS.

Merci pour votre attention !