

Modèle à processus latent et algorithme EM pour la régression non linéaire

A. Samé, F. Chamroukhi, G. Govaert



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

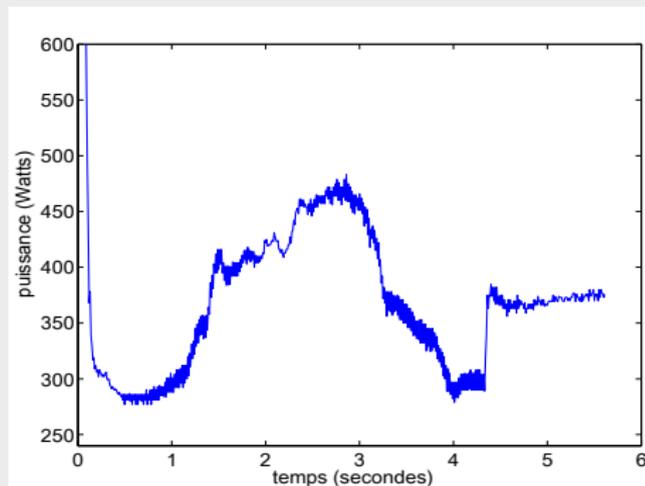
Université de Technologie de Compiègne (UTC)
Laboratoire HEUDIASYC, UMR CNRS 6599

29 Mai 2009

- 1 Contexte applicatif de l'étude
- 2 Régression non linéaire et modèle à processus latent
 - Cadre général de la régression non linéaire
 - Modèle à processus latent proposé
 - Estimation des paramètres du modèle
- 3 Expérimentations
- 4 Conclusion

Contexte applicatif de l'étude

- ▶ Caractériser au mieux des signaux d'effort de manœuvres d'aiguillages



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

Cadre général de la régression non linéaire

- ▶ Échantillon $((x_1, t_1), \dots, (x_n, t_n))$
 - x_i : variable dépendante réelle ;
 - t_i : variable indépendante représentant le temps
- ▶ Estimer une fonction g de paramètre θ , en considérant le modèle

$$M1 : x_i = g(t_i; \theta) + \epsilon_i$$

où les ϵ_i sont des bruits gaussiens centrés *i.i.d* de variance σ^2

- ▶ $g(t_i; \theta)$ représente l'espérance de x_i conditionnellement à t_i
- ▶ Estimation du paramètre θ par minimisation du critère des moindres carrés :

$$C_1(\theta) = \sum_{i=1}^n (x_i - g(t_i; \theta))^2$$

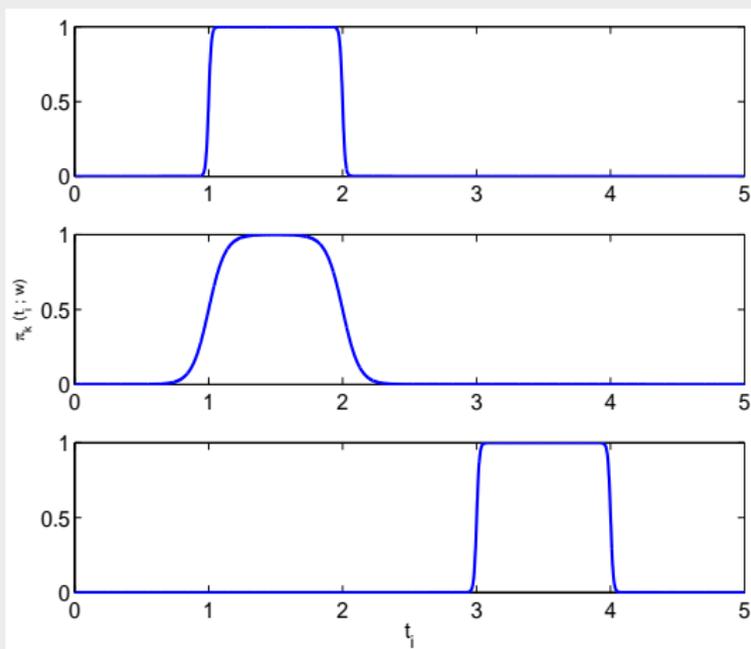
- ▶ Minimisation du critère C_1 : algorithmes itératifs d'optimisation numérique qui convergent localement (Gauss-Newton, Newton-Raphson, quasi-Newton, ...)

Fonctions de régression adoptées

Somme finie de polynômes pondérés par des fonctions logistiques

$$g(t_i; \theta) = \sum_{k=1}^K \pi_k(t_i; \mathbf{w}) \beta_k^T \mathbf{t}_i, \quad \text{avec} \quad \pi_k(t_i; \mathbf{w}) = \frac{\exp(w_{k0} + w_{k1} t_i)}{\sum_{\ell=1}^K \exp(w_{\ell 0} + w_{\ell 1} t_i)},$$

- ▶ $\mathbf{w} = (w_{10}, w_{11}, \dots, w_{K0}, w_{K1}) \in \mathbf{R}^{2K}$
- ▶ $\beta_k = (\beta_{k0}, \dots, \beta_{kp})^T \in \mathbf{R}^{p+1}$: coefficients d'un polynôme de degré p
- ▶ $\mathbf{t}_i = (1, t_i, t_i^2, \dots, t_i^p)^T \in \mathbf{R}^{p+1}$
- ▶ $\theta = (\mathbf{w}, \beta_1, \dots, \beta_K)$: vecteur paramètres

Exemple de fonctions logistiques pondérant les polynômes ($K=3$)

Modèle à processus latent proposé (1/2)

Pour estimer la fonction g , on propose de s'appuyer sur un modèle alternatif à processus latent discret $\mathbf{z} = (z_1, \dots, z_n)$, $z_i \in \{1, \dots, K\}$:

$$M2 : x_i = \sum_{k=1}^K z_{ik} \beta_k^T \mathbf{t}_i + \epsilon_i \quad ; \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

- ▶ $z_{ik} = 1$ si $z_i = k$ et $z_{ik} = 0$ sinon
- ▶ Cond. à t_i , $z_i \sim \mathcal{M}(1, \pi_1(t_i; \mathbf{w}), \dots, \pi_K(t_i; \mathbf{w}))$
 où $\pi_k(t_i; \mathbf{w}) = p(z_i = k | t_i; \mathbf{w}) = \frac{\exp(w_{k0} + w_{k1} t_i)}{\sum_{\ell=1}^K \exp(w_{\ell 0} + w_{\ell 1} t_i)}$

Modèle à processus latent proposé (2/2)

- ▶ Les modèles $M1$ et $M2$ conduisent à la même espérance conditionnelle

$$E(x_i | t_i) = \sum_{k=1}^K \pi_k(t_i; \mathbf{w}) \beta_k^T \mathbf{t}_i = g(t_i; \boldsymbol{\theta})$$

Ainsi, grâce aux propriétés asymptotiques de l'estimateur du maximum de vraisemblance, la fonction estimée à partir du modèle $M2$ sera asymptotiquement identique à celle estimée par $M1$

- ▶ On peut montrer que cond. à t_i , x_i est distribué suivant un modèle de mélange

$$p(x_i | t_i; \boldsymbol{\Phi}) = \sum_{k=1}^K \pi_k(t_i; \mathbf{w}) \mathcal{N}(x_i; \beta_k^T \mathbf{t}_i, \sigma^2), \quad \text{avec } \boldsymbol{\Phi} = (\boldsymbol{\theta}, \sigma^2)$$

→ Mélange d'experts [Jacobs and Jordan. 1991] avec contraintes sur les variances.

- ▶ Dans ce contexte, l'algorithme EM fournit un cadre adapté pour estimer le paramètre $\boldsymbol{\theta}$

Estimation des paramètres du modèle par maximum de vraisemblance

- Densité mélange conditionnelle

$$p(x_i | t_i; \Phi) = \sum_{k=1}^K \pi_k(t_i; \mathbf{w}) \mathcal{N}(x_i; \beta_k^T \mathbf{t}_i, \sigma^2)$$

- Vecteur paramètre du modèle :

$$\Phi = (\mathbf{w}, \beta_1, \dots, \beta_K, \sigma^2)$$

- Log-vraisemblance :

$$L(\Phi; \mathbf{x}) = \sum_{i=1}^n \log \sum_{k=1}^K \pi_k(t_i; \mathbf{w}) \mathcal{N}(x_i; \beta_k^T \mathbf{t}_i, \sigma^2)$$

- Maximisation de $L(\Phi; \mathbf{x})$ par un algorithme EM dédié

Algorithme EM dédié

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

Itérer jusqu'à la convergence les deux étapes suivantes :

Algorithme EM dédié

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

Itérer jusqu'à la convergence les deux étapes suivantes :

Algorithme EM dédié

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

Itérer jusqu'à la convergence les deux étapes suivantes :

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

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

$$\begin{aligned}
 Q(\Phi, \Phi^{(q)}) &= E \left[L(\Phi; \mathbf{x}, \mathbf{z}) | \mathbf{x}, \Phi^{(q)} \right] \\
 &= \underbrace{\sum_{i=1}^n \sum_{k=1}^K \tau_{ik}^{(q)} \log \pi_k(t_i; \mathbf{w})}_{Q_1(\mathbf{w})} + \underbrace{\sum_{i=1}^n \sum_{k=1}^K \tau_{ik}^{(q)} \log \mathcal{N}(x_i; \beta_k^T \mathbf{t}_i, \sigma^2)}_{Q_2(\beta_1, \dots, \beta_K, \sigma^2)},
 \end{aligned}$$

$$\text{avec } \tau_{ik}^{(q)} = p(z_i = k | x_i; \Phi^{(q)}) = \frac{\pi_k(t_i; \mathbf{w}^{(q)}) \mathcal{N}(x_i; \beta_k^T \mathbf{t}_i, \sigma^2(q))}{\sum_{\ell=1}^K \pi_\ell(t_i; \mathbf{w}^{(q)}) \mathcal{N}(x_i; \beta_\ell^T \mathbf{t}_i, \sigma^2(q))}$$

Algorithme EM dédié

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

Itérer jusqu'à la convergence les deux étapes suivantes :

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

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

$$\begin{aligned}
 Q(\Phi, \Phi^{(q)}) &= E \left[L(\Phi; \mathbf{x}, \mathbf{z}) | \mathbf{x}, \Phi^{(q)} \right] \\
 &= \underbrace{\sum_{i=1}^n \sum_{k=1}^K \tau_{ik}^{(q)} \log \pi_k(t_i; \mathbf{w})}_{Q_1(\mathbf{w})} + \underbrace{\sum_{i=1}^n \sum_{k=1}^K \tau_{ik}^{(q)} \log \mathcal{N}(x_i; \beta_k^T \mathbf{t}_i, \sigma^2)}_{Q_2(\beta_1, \dots, \beta_K, \sigma^2)},
 \end{aligned}$$

$$\text{avec } \tau_{ik}^{(q)} = p(z_i = k | x_i; \Phi^{(q)}) = \frac{\pi_k(t_i; \mathbf{w}^{(q)}) \mathcal{N}(x_i; \beta_k^T \mathbf{t}_i, \sigma^2(q))}{\sum_{\ell=1}^K \pi_\ell(t_i; \mathbf{w}^{(q)}) \mathcal{N}(x_i; \beta_\ell^T \mathbf{t}_i, \sigma^2(q))}$$

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

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

Détails de l'étape M

Maximisation séparée de $Q_1(\mathbf{w})$ et $Q_2(\beta_1, \dots, \beta_K, \sigma^2)$:

- 1 Maximisation de Q_2 par rapport aux β_k : résolution analytique de K problèmes de régression polynomiale pondérés

- $\beta_k^{(q+1)} = (\mathbf{M}^T \Gamma_k^{(q)} \mathbf{M})^{-1} \mathbf{M}^T \Gamma_k^{(q)} \mathbf{x}$

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 à σ^2 :

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

Détails de l'étape M

Maximisation séparée de $Q_1(\mathbf{w})$ et $Q_2(\beta_1, \dots, \beta_K, \sigma^2)$:

- ❶ Maximisation de Q_2 par rapport aux β_k : résolution analytique de K problèmes de régression polynomiale pondérés

- $\beta_k^{(q+1)} = (\mathbf{M}^T \Gamma_k^{(q)} \mathbf{M})^{-1} \mathbf{M}^T \Gamma_k^{(q)} \mathbf{x}$

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 à σ^2 :

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

- ❷ Maximisation de Q_1 par rapport \mathbf{w} : problème convexe de régression logistique multinomiale pondéré

⇒ algorithme IRLS [Green 84]

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

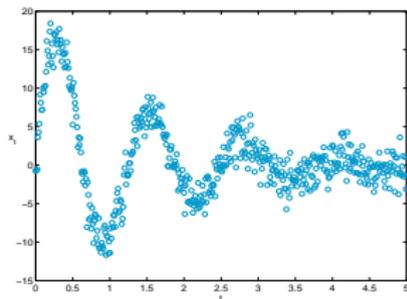
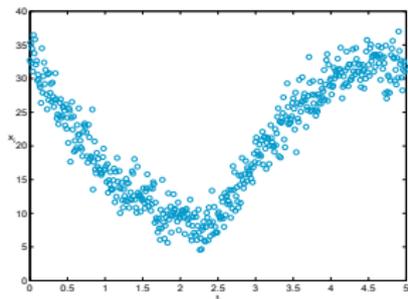
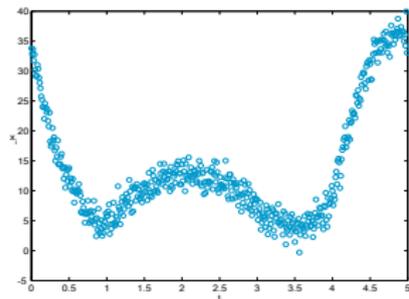
⇒ L'algorithme IRLS fournit le paramètre $\mathbf{w}^{(q+1)}$ de l'itération courante.

Expérimentations sur des données simulées :

Trois situations de fonctions non linéaires simulées

- ▶ selon le modèle proposé
- ▶ selon un modèle de régression polynomiale par morceaux
- ▶ selon une fonction non linéaire (sinus amorti)

Fonctions	Paramètres
$g_1(t_i, \theta) = \sum_{k=1}^4 \pi_k(t_i; \mathbf{w}) \beta_k^T \mathbf{t}_i$	$\begin{aligned} \beta_1 &= [34, -60, 30] & \mathbf{w}_1 &= [547, -154] \\ \beta_2 &= [-17, 29, -7] & \mathbf{w}_2 &= [526, -135] \\ \beta_3 &= [185, -104, 15] & \mathbf{w}_3 &= [464, -115] \\ \beta_4 &= [-804, 343, -35] & \mathbf{w}_4 &= [0, 0] \end{aligned}$
$g_2(t_i, \theta) = \beta_1^T \mathbf{t}_i \mathbf{1}_{[0;2.5]}(t_i) + \beta_2^T \mathbf{t}_i \mathbf{1}_{[2.5;5]}(t_i)$	$\begin{aligned} \beta_1 &= [33, -20, 4] \\ \beta_2 &= [-78, 47, -5] \end{aligned}$
$g_3(t_i) = 20 \sin(1.6\pi t_i) \exp(-0.7 t_i)$	

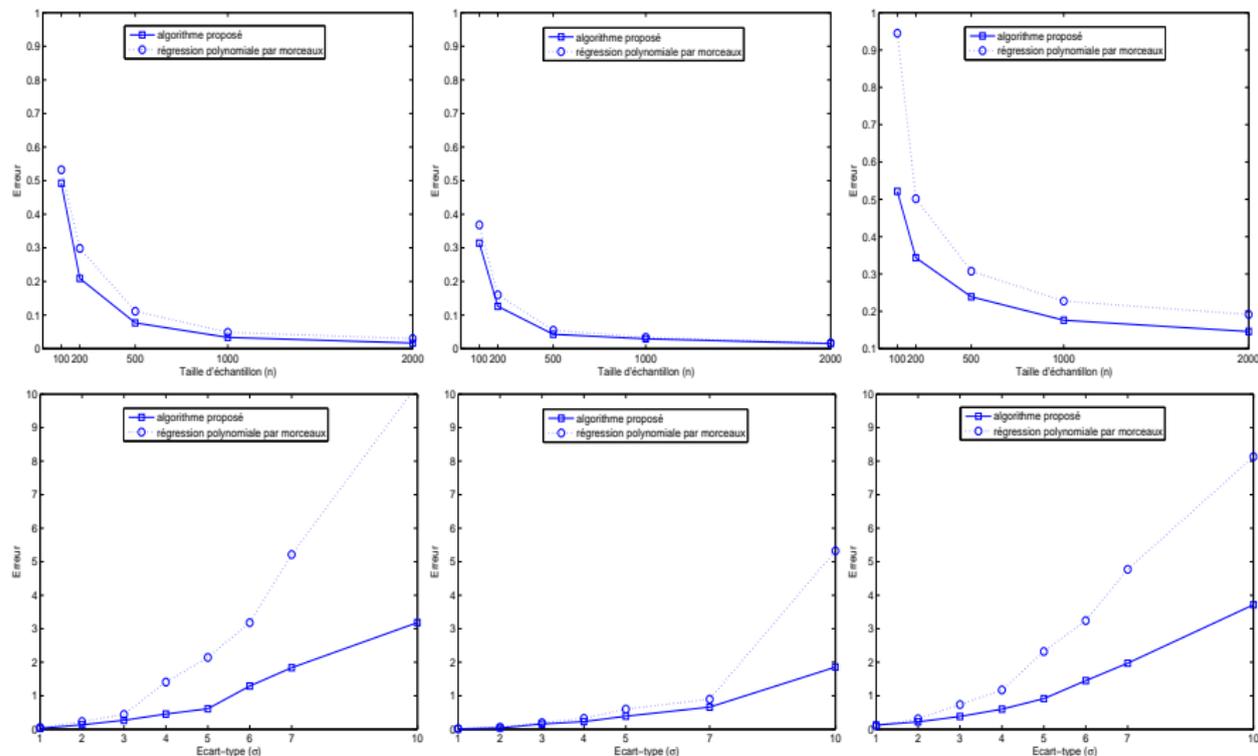


- ▶ **Critère d'évaluation** : écart quadratique moyen entre la courbe de régression estimée et la courbe de régression simulée :

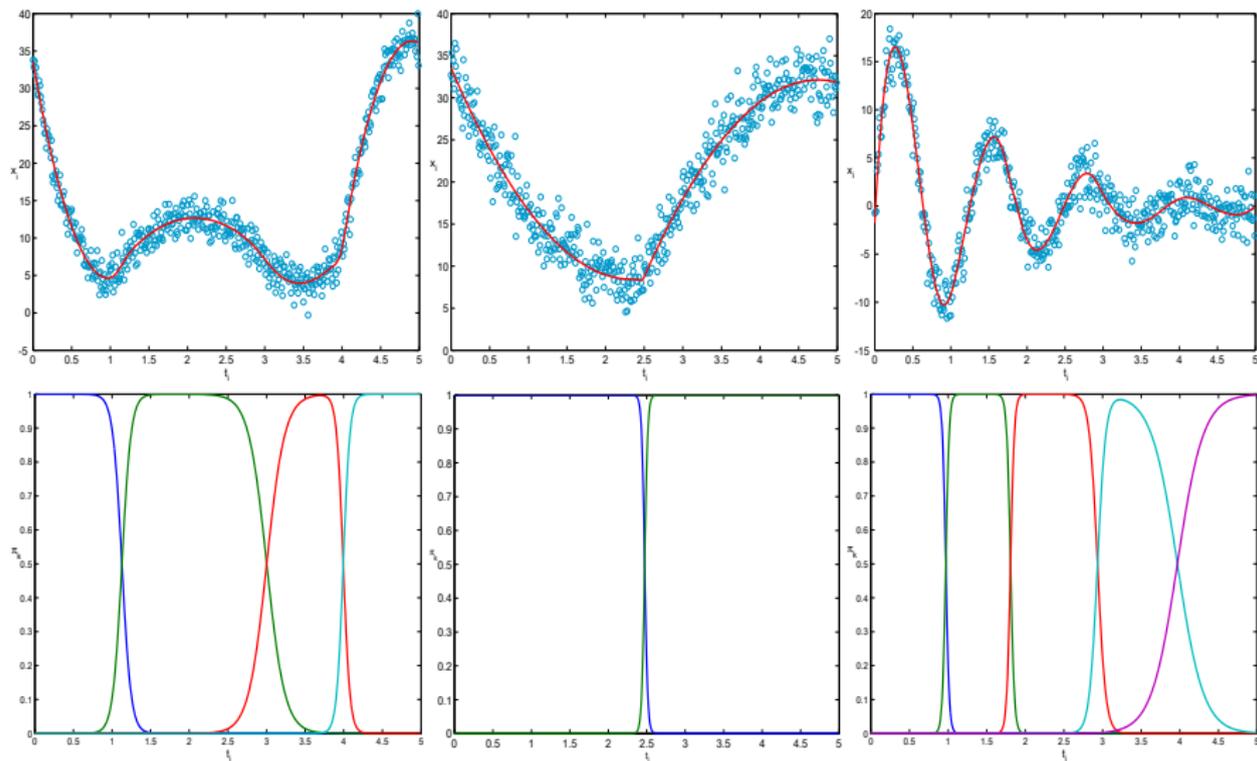
$$\frac{1}{n} \sum_{i=1}^n (g_{sim}(t_i; \theta) - g_{est}(t_i; \hat{\theta}))^2$$

- ▶ Comparaison de l'algorithme proposé avec avec la méthode de régression par morceaux [McGee et Carleton. 65].
- ▶ Simulation de Monte Carlo : 20 jeux de données différents sont simulés pour chaque situation et les résultats sont moyennés

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

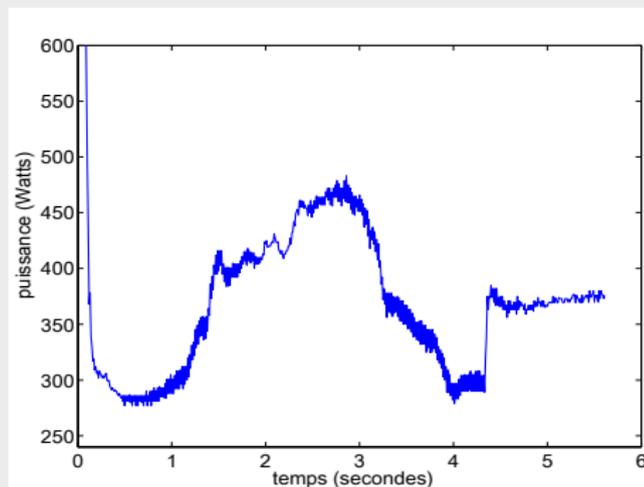


Résultats : Courbes de régression estimées et fonctions logistiques associées

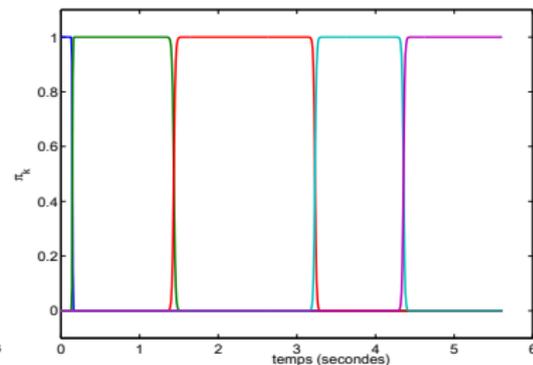
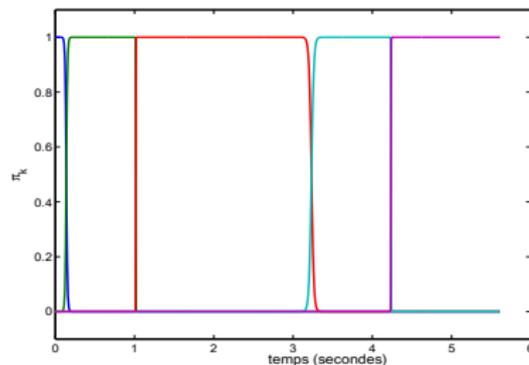
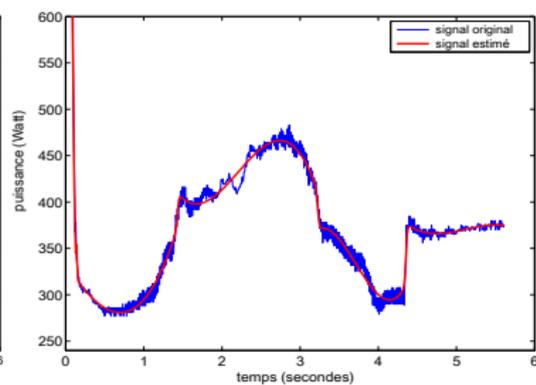
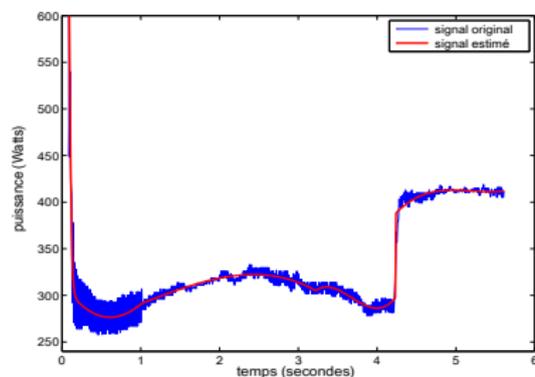


Données réelles

- ▶ Signaux de manœuvres d'aiguillage
- ▶ Nombre de composantes du mélange choisi : $K = 5$ (nombre de phases mécaniques d'une manœuvre)
- ▶ Ordre de régression polynomiale : $p = 3$;



Résultats : Fonctions de régression estimées et fonctions logistiques associées



Conclusion

- ▶ Modèle de régression à variable latente proposé pour résoudre le problème de régression non linéaire.
- ▶ 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.

Merci pour votre attention !