

Détection statistique de rupture de modèle dans les systèmes dynamiques - Application à la supervision de procédés de dépollution

Ghislain VERDIER

UMR Analyse des Systèmes et Biométrie - SUPAGRO-INRA - Montpellier

Laboratoire de Probabilité et Statistiques - Université Montpellier II

30 Novembre 2007



- 1 Contexte et Objectifs de la Thèse
- 2 La règle du CUSUM
 - Définition
 - Introduction d'un seuil adaptatif
- 3 Approximation du CUSUM
 - Définition et Résultats
 - Le CUSUM non paramétrique
 - Algorithmes de filtrage-détection
- 4 Applications
- 5 Conclusions et Perspectives

- 1 Contexte et Objectifs de la Thèse
- 2 La règle du CUSUM
 - Définition
 - Introduction d'un seuil adaptatif
- 3 Approximation du CUSUM
 - Définition et Résultats
 - Le CUSUM non paramétrique
 - Algorithmes de filtrage-détection
- 4 Applications
- 5 Conclusions et Perspectives

- ⇒ Mettre au point des approches statistiques de détection de rupture pour des systèmes dynamiques.
- ⇒ Application à la détection de panne sur un procédé de dépollution biologique.

Plusieurs particularités pour ce type de procédés :

- modélisation incertaine
 - ⇒ Approche non paramétrique
- observation indirecte
 - ⇒ Approche par filtrage particulière

- Type de modèle étudié :

$$X_n = g(X_{n-1}, \theta) + \epsilon_n$$

- X_n variable d'état,
 - ϵ_n bruit blanc (généralement gaussien).
- θ paramètre caractérisant le changement :
 - avant l'instant de changement, noté t_p (inconnu) : $\theta = \theta_0$ (régime H_0)
 - à partir de l'instant de changement : $\theta = \theta_i, i = 1, \dots, K$ (régime H_i)
 - Détection : $K = 1$, Diagnostic : $K \geq 2$.
 - Construction d'une statistique de test et comparaison avec un seuil h pour détecter le changement de paramètre (ou la rupture de modèle).
 - Critères de qualité d'une règle : faible retard à la détection, faible taux de fausses alarmes.

- Type de modèle étudié :

$$X_n = g(X_{n-1}, \theta) + \epsilon_n$$

- X_n variable d'état,
 - ϵ_n bruit blanc (généralement gaussien).
- θ paramètre caractérisant le changement :
 - avant l'instant de changement, noté t_p (inconnu) : $\theta = \theta_0$ (régime H_0)
 - à partir de l'instant de changement : $\theta = \theta_i$, $i = 1, \dots, K$ (régime H_i)
 - Détection : $K = 1$, Diagnostic : $K \geq 2$.
 - Construction d'une statistique de test et comparaison avec un seuil h pour détecter le changement de paramètre (ou la rupture de modèle).
 - Critères de qualité d'une règle : faible retard à la détection, faible taux de fausses alarmes.

- Type de modèle étudié :

$$X_n = g(X_{n-1}, \theta) + \epsilon_n$$

- X_n variable d'état,
 - ϵ_n bruit blanc (généralement gaussien).
- θ paramètre caractérisant le changement :
 - avant l'instant de changement, noté t_p (inconnu) : $\theta = \theta_0$ (régime H_0)
 - à partir de l'instant de changement : $\theta = \theta_i$, $i = 1, \dots, K$ (régime H_i)
 - Détection : $K = 1$, Diagnostic : $K \geq 2$.
 - Construction d'une statistique de test et comparaison avec un seuil h pour détecter le changement de paramètre (ou la rupture de modèle).
 - Critères de qualité d'une règle : faible retard à la détection, faible taux de fausses alarmes.

- Type de modèle étudié :

$$X_n = g(X_{n-1}, \theta) + \epsilon_n$$

- X_n variable d'état,
 - ϵ_n bruit blanc (généralement gaussien).
- θ paramètre caractérisant le changement :
 - avant l'instant de changement, noté t_p (inconnu) : $\theta = \theta_0$ (régime H_0)
 - à partir de l'instant de changement : $\theta = \theta_i$, $i = 1, \dots, K$ (régime H_i)
 - Détection : $K = 1$, Diagnostic : $K \geq 2$.
 - Construction d'une statistique de test et comparaison avec un seuil h pour détecter le changement de paramètre (ou la rupture de modèle).
 - Critères de qualité d'une règle : faible retard à la détection, faible taux de fausses alarmes.

- Type de modèle étudié :

$$X_n = g(X_{n-1}, \theta) + \epsilon_n$$

- Une partie du modèle pourra être inconnue :

$$X_n = f(X_{n-1}) + F(X_{n-1}, \theta) + \epsilon_n$$

f inconnue \Rightarrow Estimation non paramétrique

- La variable d'état pourra être indirectement observée :

$$\begin{cases} X_n = g(X_{n-1}, \theta, \epsilon_n) \\ Y_n = h(X_n, \theta, v_n) \end{cases}$$

\Rightarrow Méthode de filtrage

- Type de modèle étudié :

$$X_n = g(X_{n-1}, \theta) + \epsilon_n$$

- Une partie du modèle pourra être inconnue :

$$X_n = f(X_{n-1}) + F(X_{n-1}, \theta) + \epsilon_n$$

f inconnue \Rightarrow Estimation non paramétrique

- La variable d'état pourra être indirectement observée :

$$\begin{cases} X_n = g(X_{n-1}, \theta, \epsilon_n) \\ Y_n = h(X_n, \theta, v_n) \end{cases}$$

\Rightarrow Méthode de filtrage

- Type de modèle étudié :

$$X_n = g(X_{n-1}, \theta) + \epsilon_n$$

- Une partie du modèle pourra être inconnue :

$$X_n = f(X_{n-1}) + F(X_{n-1}, \theta) + \epsilon_n$$

f inconnue \Rightarrow Estimation non paramétrique

- La variable d'état pourra être indirectement observée :

$$\begin{cases} X_n = g(X_{n-1}, \theta, \epsilon_n) \\ Y_n = h(X_n, \theta, v_n) \end{cases}$$

\Rightarrow Méthode de filtrage

- 1 Contexte et Objectifs de la Thèse
- 2 La règle du CUSUM
 - Définition
 - Introduction d'un seuil adaptatif
- 3 Approximation du CUSUM
 - Définition et Résultats
 - Le CUSUM non paramétrique
 - Algorithmes de filtrage-détection
- 4 Applications
- 5 Conclusions et Perspectives

- 1 Contexte et Objectifs de la Thèse
- 2 La règle du CUSUM
 - Définition
 - Introduction d'un seuil adaptatif
- 3 Approximation du CUSUM
 - Définition et Résultats
 - Le CUSUM non paramétrique
 - Algorithmes de filtrage-détection
- 4 Applications
- 5 Conclusions et Perspectives

Quelques notations

- $X_{1:n} := X_{1,\dots,n}$
- $P^{(t_p)}$: mesure de probabilité associée à X_1, \dots, X_n, \dots avec un changement de régime en t_p (passage de H_0 à H_1).
- $E^{(t_p)}$: espérance associée à la mesure $P^{(t_p)}$.
- si $t_p = \infty$, on utilise P_{θ_0} et E_{θ_0} .

La règle du CUSUM (Page (1954))

- Permet la détection d'un changement dans la densité conditionnelle **connue** des observations $p_{\theta}(X_i|X_1, \dots, X_{i-1})$.
- Suivant le type de modèle, il existe plusieurs façons de générer un changement :

- Cas 1 : La densité conditionnelle sous H_1 ne dépend pas de t_p :

$$\text{si } n < t_p, \quad X_n|X_1, \dots, X_{n-1} \sim p_{\theta_0}(X_n|X_1, \dots, X_{n-1})$$

$$\text{si } n \geq t_p, \quad X_n|X_1, \dots, X_{n-1} \sim p_{\theta_1}(X_n|X_1, \dots, X_{n-1}).$$

- Cas 2 : La densité conditionnelle sous H_1 dépend de t_p :

$$\text{si } n < t_p, \quad X_n|X_1, \dots, X_{n-1} \sim p_{\theta_0}(X_n|X_1, \dots, X_{n-1})$$

$$\text{si } n \geq t_p, \quad X_n|X_1, \dots, X_{n-1} \sim p_{\theta_0, \theta_1, t_p}(X_n|X_1, \dots, X_{n-1}).$$

La règle du CUSUM (Page (1954))

- Permet la détection d'un changement dans la densité conditionnelle **connue** des observations $p_{\theta}(X_i|X_1, \dots, X_{i-1})$.
- Suivant le type de modèle, il existe plusieurs façons de générer un changement :
- Cas 1 : La densité conditionnelle sous H_1 ne dépend pas de t_p :

$$\text{si } n < t_p, \quad X_n|X_1, \dots, X_{n-1} \sim p_{\theta_0}(X_n|X_1, \dots, X_{n-1})$$

$$\text{si } n \geq t_p, \quad X_n|X_1, \dots, X_{n-1} \sim p_{\theta_1}(X_n|X_1, \dots, X_{n-1}).$$

- Cas 2 : La densité conditionnelle sous H_1 dépend de t_p :

$$\text{si } n < t_p, \quad X_n|X_1, \dots, X_{n-1} \sim p_{\theta_0}(X_n|X_1, \dots, X_{n-1})$$

$$\text{si } n \geq t_p, \quad X_n|X_1, \dots, X_{n-1} \sim p_{\theta_0, \theta_1, t_p}(X_n|X_1, \dots, X_{n-1}).$$

La règle du CUSUM (Page (1954))

- Permet la détection d'un changement dans la densité conditionnelle **connue** des observations $p_{\theta}(X_i|X_1, \dots, X_{i-1})$.
- Suivant le type de modèle, il existe plusieurs façons de générer un changement :

- Cas 1 : La densité conditionnelle sous H_1 ne dépend pas de t_p :

$$\text{si } n < t_p, \quad X_n|X_1, \dots, X_{n-1} \sim p_{\theta_0}(X_n|X_1, \dots, X_{n-1})$$

$$\text{si } n \geq t_p, \quad X_n|X_1, \dots, X_{n-1} \sim p_{\theta_1}(X_n|X_1, \dots, X_{n-1}).$$

- Cas 2 : La densité conditionnelle sous H_1 dépend de t_p :

$$\text{si } n < t_p, \quad X_n|X_1, \dots, X_{n-1} \sim p_{\theta_0}(X_n|X_1, \dots, X_{n-1})$$

$$\text{si } n \geq t_p, \quad X_n|X_1, \dots, X_{n-1} \sim p_{\theta_0, \theta_1, t_p}(X_n|X_1, \dots, X_{n-1}).$$

- Conditions d'applications : valeurs θ_0 et θ_1 connues.
- Statistique de test :

$$g_n = \max_{1 \leq j \leq n} \sum_{i=j}^n Z_{i,j} \quad \text{avec} \quad Z_{i,j} = \log \frac{p_{\theta_0, \theta_1, j}(X_i | X_{1:i-1})}{p_{\theta_0}(X_i | X_{1:i-1})}$$

- Temps d'arrêt de la règle :

$$t_a = \inf\{n : g_n \geq h\}$$

où h (seuil) fixé par expérimentateur.

- Possibilité d'écriture récursive de la règle dans le cas de génération 1.

La règle du CUSUM : application

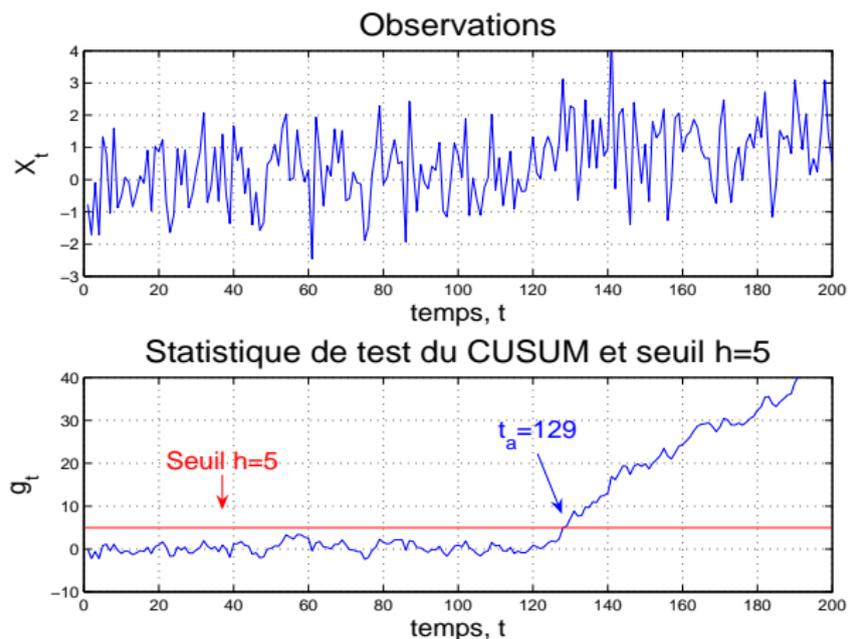


Figure: Mise en place du test CUSUM sur des données suivant une loi $N(0,1)$ avant l'instant de changement $t_p = 120$ et suivant une loi $N(1,1)$ après. Pour $h = 5$, la détection a lieu à l'instant $t_a = 129$.

Choix du seuil pour des algorithmes de type CUSUM

- Le seuil h réalise un compromis entre le taux de fausses alarmes et le retard à la détection.
- Il est généralement choisi tel que la contrainte suivante soit vérifiée :

$$E_{\theta_0}(t_a) = \gamma ,$$

où γ est fixée par l'expérimentateur.

- Plusieurs travaux théoriques dans le cas de données *iid*.
- Des méthodes à partir de simulations de Monte Carlo mais qui fournissent un seuil fixe.

- 1 Contexte et Objectifs de la Thèse
- 2 La règle du CUSUM
 - Définition
 - Introduction d'un seuil adaptatif
- 3 Approximation du CUSUM
 - Définition et Résultats
 - Le CUSUM non paramétrique
 - Algorithmes de filtrage-détection
- 4 Applications
- 5 Conclusions et Perspectives

Le seuil adaptatif h_n : Principe

- Objectif : obtenir un seuil qui garantit un taux de fausses alarmes constant au cours du temps.
- Basé sur une estimation de la densité de la statistique de test g_n .
- A chaque instant, nous calculons h_n tel que :

$$P_{\theta_0}[g_n \geq h_n | g_1 < h_1, \dots, g_{n-1} < h_{n-1}] = \alpha$$

où α est fixée par l'expérimentateur.

- Si la contrainte précédente est vérifiée, alors : $E_{\theta_0}[t_a] = 1/\alpha$.
- Procédure étendue au diagnostic.

Nombreux résultats d'optimalité

- Données indépendantes : Lorden (1971), Moustakides (1986),...
- Données dépendantes : Mei (2006), Tartakovsky (2007),...

Lai (1998)

La règle du CUSUM minimise asymptotiquement ($\gamma \rightarrow \infty$) le critère du pire retard moyen à la détection (Lorden) :

$$\bar{E}_{\theta_1}(T) = \sup_{t_p \geq 1} \sup_{\text{ess}} E^{(t_p)}[(T - t_p + 1)^+ | X_1, \dots, X_{t_p-1}]$$

parmi toutes les règles T vérifiant :

$$E_{\theta_0}(T) \geq \gamma.$$

⇒ "Optimalité au sens de Lorden".

- 1 Contexte et Objectifs de la Thèse
- 2 La règle du CUSUM
 - Définition
 - Introduction d'un seuil adaptatif
- 3 Approximation du CUSUM**
 - Définition et Résultats
 - Le CUSUM non paramétrique
 - Algorithmes de filtrage-détection
- 4 Applications
- 5 Conclusions et Perspectives

- 1 Contexte et Objectifs de la Thèse
- 2 La règle du CUSUM
 - Définition
 - Introduction d'un seuil adaptatif
- 3 Approximation du CUSUM
 - Définition et Résultats
 - Le CUSUM non paramétrique
 - Algorithmes de filtrage-détection
- 4 Applications
- 5 Conclusions et Perspectives

Inconvénient du CUSUM

- Pour appliquer la règle du CUSUM, il est nécessaire de connaître l'écriture analytique des densités conditionnelles :

$$p_{\theta_0}(X_n|X_1, \dots, X_{n-1}) \text{ et } p_{\theta_0, \theta_{1,j}}(X_n|X_1, \dots, X_{n-1}).$$

- Cette écriture n'est pas accessible :
 - lorsqu'une partie du modèle est inconnue
 - dans un modèle à espace d'état

⇒ Dans les deux cas, on va pouvoir estimer les vraisemblances conditionnelles...

Inconvénient du CUSUM

- Pour appliquer la règle du CUSUM, il est nécessaire de connaître l'écriture analytique des densités conditionnelles :

$$p_{\theta_0}(X_n|X_1, \dots, X_{n-1}) \text{ et } p_{\theta_0, \theta_{1,j}}(X_n|X_1, \dots, X_{n-1}).$$

- Cette écriture n'est pas accessible :
 - lorsqu'une partie du modèle est inconnue
 - dans un modèle à espace d'état

⇒ Dans les deux cas, on va pouvoir estimer les vraisemblances conditionnelles...

On se place dans le cas où :

$$\text{si } n < t_p, \quad X_n | X_1, \dots, X_{n-1} \sim p_{\theta_0}(X_n | X_1, \dots, X_{n-1})$$

$$\text{si } n \geq t_p, \quad X_n | X_1, \dots, X_{n-1} \sim p_{\theta_0, \theta_1, t_p}(X_n | X_1, \dots, X_{n-1}),$$

Les densités conditionnelles p_{θ_0} et $p_{\theta_0, \theta_1, j}$, $j = 1, \dots, n$ sont inconnues mais on peut les estimer respectivement par :

$$\widehat{l}_{0,n}^N \quad \text{et} \quad \widehat{l}_{1,n,j}^N, \quad j = 1, \dots, n$$

La règle du "CUSUM approché"

Le CUSUM approché

Le temps d'arrêt de la règle du CUSUM approché est défini par :

$$\hat{t} = \inf \left\{ n : \max_{1 \leq j \leq n} \sum_{i=j}^n \hat{Z}_{i,j}^N \geq h \right\},$$

avec

$$\hat{Z}_{i,j}^N = \log \frac{\hat{l}_{1,n,j}^N}{\hat{l}_{0,n}^N}.$$

Question : Sous quelles conditions le CUSUM approché est-il optimal?

La règle du "CUSUM approché"

Le CUSUM approché

Le temps d'arrêt de la règle du CUSUM approché est défini par :

$$\hat{t} = \inf \left\{ n : \max_{1 \leq j \leq n} \sum_{i=j}^n \hat{Z}_{i,j}^N \geq h \right\},$$

avec

$$\hat{Z}_{i,j}^N = \log \frac{\hat{l}_{1,n,j}^N}{\hat{l}_{0,n}^N}.$$

Question : Sous quelles conditions le CUSUM approché est-il optimal?

Résultat d'optimalité

Hypothèse 1

Les hypothèses de stabilité suivantes sont vérifiées : $\exists l$ telle que :

$$\lim_{n \rightarrow \infty} \sup_{t_p \geq 1} \sup \text{ess } P^{(t_p)} \left\{ \max_{t \leq n} \sum_{i=t_p}^{t_p+t} Z_{i,t_p} \geq l(1 + \delta)n | X_{1:t_p-1} \right\} = 0 \quad \forall \delta > 0, \quad (1)$$

$$\lim_{n \rightarrow \infty} \sup_{t_p \geq 1} \sup_{j \geq t_p} \sup \text{ess } P^{(t_p)} \left\{ n^{-1} \sum_{i=j}^{j+n} Z_{i,j} \leq l(1 - \delta) | X_{1:j-1} \right\} = 0, \quad \forall \delta > 0, \quad (2)$$

Hypothèse 2

Les estimations des vraisemblances sont convergentes : lorsque $N \rightarrow \infty$,

$$\begin{cases} \widehat{l}_{0,n}^N \rightarrow p_{\theta_0}(X_n | X_{1:n-1}) \text{ p.s.} \\ \widehat{l}_{1,n,j}^N \rightarrow p_{\theta_0, \theta_{1,j}}(X_n | X_{1:n-1}) \text{ p.s. } \forall j \in [1; n], \end{cases}$$

Résultat d'optimalité

Hypothèse 1

Les hypothèses de stabilité suivantes sont vérifiées : $\exists l$ telle que :

$$\lim_{n \rightarrow \infty} \sup_{t_p \geq 1} \sup \text{ess } P^{(t_p)} \left\{ \max_{t \leq n} \sum_{i=t_p}^{t_p+t} Z_{i,t_p} \geq l(1 + \delta)n | X_{1:t_p-1} \right\} = 0 \quad \forall \delta > 0, \quad (1)$$

$$\lim_{n \rightarrow \infty} \sup_{t_p \geq 1} \sup_{j \geq t_p} \sup \text{ess } P^{(t_p)} \left\{ n^{-1} \sum_{i=j}^{j+n} Z_{i,j} \leq l(1 - \delta) | X_{1:j-1} \right\} = 0, \quad \forall \delta > 0, \quad (2)$$

Hypothèse 2

Les estimations des vraisemblances sont convergentes : lorsque $N \rightarrow \infty$,

$$\begin{cases} \widehat{l}_{0,n}^N \rightarrow p_{\theta_0}(X_n | X_{1:n-1}) \text{ p.s.} \\ \widehat{l}_{1,n,j}^N \rightarrow p_{\theta_0, \theta_{1,j}}(X_n | X_{1:n-1}) \text{ p.s. } \forall j \in [1; n], \end{cases}$$

Théorème 1

Si les hypothèses 1 et 2 sont vérifiées, alors la règle du CUSUM approché \hat{t} est asymptotiquement (sur N et γ) optimale au sens de Lorden.

Autrement dit, \hat{t} minimise le critère du pire retard moyen à la détection :

$$\bar{E}_{\theta_1}(T) = \sup_{t_p \geq 1} \sup_{\text{ess}} E^{(t_p)}[(T - t_p + 1)^+ | X_1, \dots, X_{t_p-1}]$$

parmi toutes les règles T vérifiant :

$$E_{\theta_0}(T) \geq \gamma,$$

lorsque $\gamma \rightarrow \infty$.

- 1 Contexte et Objectifs de la Thèse
- 2 La règle du CUSUM
 - Définition
 - Introduction d'un seuil adaptatif
- 3 Approximation du CUSUM
 - Définition et Résultats
 - Le CUSUM non paramétrique
 - Algorithmes de filtrage-détection
- 4 Applications
- 5 Conclusions et Perspectives

- On considère le modèle :

$$X_{n+1} = f(X_n) + F(X_n, \theta) + \epsilon_{n+1}$$

où f est inconnue et ϵ_n est un bruit blanc gaussien.

- REMARQUE : Densité conditionnelle sous H_1 indépendante de t_p .
- Les incréments du CUSUM s'écrivent :

$$Z_i = \frac{(\eta_i^0)^2 - (\eta_i^1)^2}{2\sigma^2}$$

avec $\eta_i^d = X_i - f(X_{i-1}) - F(X_{i-1}, \theta_d)$, $d = 0, 1$.

⇒ On ne peut plus appliquer le CUSUM.

- On estime alors la fonction f ⇒ CUSUM non paramétrique.

$$X_{n+1} = f(X_n) + F(X_n, \theta) + \epsilon_{n+1}$$

- Echantillon d'apprentissage issu du modèle : $(\bar{X}_i)_{1 \leq i \leq N}$
- L'estimateur :

$$\hat{f}_N(x) = \frac{\sum_{i=1}^{N-1} k_i^{-1} K\left(\frac{x - \bar{X}_i}{k_i}\right) (\bar{X}_{i+1} - F(\bar{X}_i, \theta))}{\sum_{i=1}^{N-1} k_i^{-1} K\left(\frac{x - \bar{X}_i}{k_i}\right)}, \quad \forall x.$$

$\Rightarrow K$ est un noyau et k_i sa fenêtre.

- Consistance de l'estimateur sous de "bonnes hypothèses".
- Possibilité d'écriture récursive de l'estimateur.

$$X_{n+1} = f(X_n) + F(X_n, \theta) + \epsilon_{n+1}$$

- Echantillon d'apprentissage issu du modèle : $(\bar{X}_i)_{1 \leq i \leq N}$
- L'estimateur :

$$\hat{f}_N(x) = \frac{\sum_{i=1}^{N-1} k_i^{-1} K\left(\frac{x - \bar{X}_i}{k_i}\right) (\bar{X}_{i+1} - F(\bar{X}_i, \theta))}{\sum_{i=1}^{N-1} k_i^{-1} K\left(\frac{x - \bar{X}_i}{k_i}\right)}, \quad \forall x.$$

$\Rightarrow K$ est un noyau et k_i sa fenêtre.

- Consistance de l'estimateur sous de "bonnes hypothèses".
- Possibilité d'écriture récursive de l'estimateur.

$$X_{n+1} = f(X_n) + F(X_n, \theta) + \epsilon_{n+1}$$

- Echantillon d'apprentissage issu du modèle : $(\bar{X}_i)_{1 \leq i \leq N}$
- L'estimateur :

$$\hat{f}_N(x) = \frac{\sum_{i=1}^{N-1} k_i^{-1} K\left(\frac{x - \bar{X}_i}{k_i}\right) (\bar{X}_{i+1} - F(\bar{X}_i, \theta))}{\sum_{i=1}^{N-1} k_i^{-1} K\left(\frac{x - \bar{X}_i}{k_i}\right)}, \quad \forall x.$$

$\Rightarrow K$ est un noyau et k_i sa fenêtre.

- Consistance de l'estimateur sous de "bonnes hypothèses".
- Possibilité d'écriture récursive de l'estimateur.

Ecriture

La statistique de test du CUSUM non paramétrique s'écrit :

$$\hat{g}_n^N = \max_{1 \leq k \leq n} \sum_{i=k}^n \hat{Z}_i$$

avec
$$\hat{Z}_i^N = \frac{(\hat{\eta}_i^0)^2 - (\hat{\eta}_i^1)^2}{2\sigma^2}$$

et
$$\hat{\eta}_i^d = X_i - \hat{f}_N(X_{i-1}) - F(X_{i-1}, \theta_d), \quad d = 0, 1.$$

Nécessite un échantillon d'apprentissage avec des données en régime de panne.

Théorème 2

La règle du CUSUM non paramétrique :

$$\hat{t}_{NP} = \inf\{n : \hat{g}_n^N \geq h\},$$

est asymptotiquement optimale (sur N et γ) au sens de Lorden.

- Modèle de dépollution biologique :

$$\begin{cases} X_{n+1} = (TX_n)\mu_n + (X_n - T(1 - \theta)U_nX_n) + \epsilon_{n+1}^{(1)} \\ S_{n+1} = (-\frac{TX_n}{\tau})\mu_n + (S_n + (1 - \theta)U_n(S_{ini} - S_n)T) + \epsilon_{n+1}^{(2)} \end{cases}$$

- $\epsilon_{n+1}^{(1)}, \epsilon_{n+1}^{(2)}$ bb gaussiens.
- $T = 0.17h, \tau = 1, S_{ini} = 50mg.l^{-1}$ et $\mu_{max} = 0.05h^{-1}$.
- $\theta_0 = 0$ et différentes valeurs pour θ_1 .
- $\mu_n = \mu(S_n)$ (taux de croissance de la biomasse) est souvent mal connue pour ce type de procédé.
⇒ Utilisation de l'estimateur à noyau.
- Comparaison entre CUSUM et CUSUM non paramétrique.

- On applique le CUSUM en supposant que $\mu(S_n)$ suit une loi de Monod :

$$\mu(S_n) = \mu^{Monod}(S_n) = (\mu_{max} \cdot S_n) / (S_n + K_s).$$

- En réalité, on simule le modèle avec :

$$\mu(S_n) = (1 - a)\mu^{Monod}(S_n) + a\mu^{Tessier}(S_n),$$

où a est le taux de mélange.

- On applique d'autre part, le CUSUM non paramétrique, avec K noyau gaussien, $k_i = 10 \cdot i^{-4}$, et une “panne d'apprentissage”.
- On simule une panne et on s'intéresse au retard à la détection sachant que $E_{\theta_0}(t_a) = 200$.
- Retard moyen à la détection sur 2000 répétitions.

Valeur de a	Tests	$\theta_1 = 0.02$	$\theta_1 = 0.01$	$\theta_1 = 0.005$
0%	t_{CUSUM}	15.28	34.00	66.08
	\hat{t}_{NP}	15.41	34.13	65.04
1%	t_{CUSUM}	15.38	35.64	68.05
	\hat{t}_{NP}	15.10	34.93	65.11
2%	t_{CUSUM}	15.55	40.80	81.09
	\hat{t}_{NP}	15.11	35.98	64.42
5%	t_{CUSUM}	27.01	82.45	129.42
	\hat{t}_{NP}	15.31	33.16	66.00

Retard moyen à la détection des algorithmes CUSUM et CUSUM non paramétrique pour différentes pannes et en fonction du taux de mélange de la loi de Monod.

- 1 Contexte et Objectifs de la Thèse
- 2 La règle du CUSUM
 - Définition
 - Introduction d'un seuil adaptatif
- 3 Approximation du CUSUM
 - Définition et Résultats
 - Le CUSUM non paramétrique
 - Algorithmes de filtrage-détection
- 4 Applications
- 5 Conclusions et Perspectives

A présent, la variable d'intérêt X n'est pas mesurée. On considère le modèle :

$$\begin{cases} X_n = f(X_{n-1}, \theta, v_n) \\ Y_n = g(X_n, \theta, w_n) \end{cases}$$

Méthodes statistiques de détection :

- Dans le cas linéaire : procédure de Willsky et Jones (filtre de Kalman)
- Dans le cas non linéaire :
 - Willsky et Jones avec Kalman étendu (aucune justification théorique)
 - **Approximation du CUSUM** :
 $p_\theta(Y_t | Y_{1:t-1})$ inconnue \Rightarrow Estimation par filtrage particulière + intégration de Monte Carlo.

Le filtre à noyaux de convolution (Rossi 2004)

Objectif des méthodes de filtrage : à partir d'un modèle

$$\begin{cases} X_n = f(X_{n-1}, v_n) \\ Y_n = g(X_n, w_n), \end{cases}$$

⇒ déterminer le filtre optimal $p(X_t | Y_{1:t})$

Filtre à noyaux

- filtre particulière
- conditions d'application moins restrictives
- résultats de convergence

- Le modèle :

$$\begin{cases} X_n = f(X_{n-1}, \theta, v_n) \\ Y_n = h(X_n, \theta, w_n). \end{cases}$$

Pour $n < t_p$, $\theta = \theta_0$ et pour $n \geq t_p$, $\theta = \theta_1$.

- On va considérer tout un ensemble de modèles qui vont caractériser tous les comportements possibles du système de sorte à estimer :

$$p_{\theta_0}(Y_n | Y_{1:n-1})$$

et $p_{\theta_0, \theta_{1,j}}(Y_n | Y_{1:n-1})$ pour tout j dans $\{1, \dots, n\}$.

- Fonctionnement normal du système (1 modèle): $\forall i \geq 1$:

$$\begin{cases} X_i = f(X_{i-1}, \theta_0, v_i) \\ Y_i = h(X_i, \theta_0, w_i). \end{cases}$$

⇒ Estimation de $p_{\theta_0}(Y_t | Y_{1:t-1})$

- Fonctionnement sous régime de panne (t modèles à l'instant t) :
modèle : $1 \leq j \leq t$

$$\begin{cases} X_i = f(X_{i-1}, \theta_0, v_i) \\ Y_i = h(X_i, \theta_0, w_i) , \end{cases} \quad \text{pour } i < j$$

$$\begin{cases} X_i = f(X_{i-1}, \theta_1, v_i) \\ Y_i = h(X_i, \theta_1, w_i) , \end{cases} \quad \text{pour } i \geq j$$

⇒ Estimation de $p_{\theta_0, \theta_1, j}(Y_t | Y_{1:t-1})$

- Fonctionnement normal du système (1 modèle): $\forall i \geq 1$:

$$\begin{cases} X_i = f(X_{i-1}, \theta_0, v_i) \\ Y_i = h(X_i, \theta_0, w_i). \end{cases}$$

⇒ Estimation de $p_{\theta_0}(Y_t | Y_{1:t-1})$

- Fonctionnement sous régime de panne (t modèles à l'instant t) :
modèle : $1 \leq j \leq t$

$$\begin{cases} X_i = f(X_{i-1}, \theta_0, v_i) \\ Y_i = h(X_i, \theta_0, w_i) \end{cases}, \quad \text{pour } i < j$$

$$\begin{cases} X_i = f(X_{i-1}, \theta_1, v_i) \\ Y_i = h(X_i, \theta_1, w_i) \end{cases}, \quad \text{pour } i \geq j$$

⇒ Estimation de $p_{\theta_0, \theta_1, j}(Y_t | Y_{1:t-1})$

- On remarque que :

$$p_{\theta_0}(Y_t|Y_{1:t-1}) = \int p_{\theta_0}(Y_t|x_t) \cdot p_{\theta_0}(x_t|Y_{1:t-1}) dx_t ,$$

- On estime le filtre prédictif $p_{\theta_0}(x_t|Y_{1:t-1})$
 \Rightarrow filtre particulière à noyaux de convolution.

- Nous pouvons estimer $p_{\theta_0}(Y_t|Y_{1:t-1})$ par :

$$\hat{l}_{0,t}^{N,m} = \frac{1}{m} \sum_{k=1}^m p_{\theta_0}(Y_t|x_{t,0}(k)) ,$$

avec $x_{t,0}(k) \sim \hat{p}_{\theta_0}(x_t|Y_{1:t-1})$, $k = 1, \dots, m$.

- On remarque que :

$$p_{\theta_0}(Y_t|Y_{1:t-1}) = \int p_{\theta_0}(Y_t|x_t) \cdot p_{\theta_0}(x_t|Y_{1:t-1}) dx_t ,$$

- On estime le filtre prédictif $p_{\theta_0}(x_t|Y_{1:t-1})$
⇒ filtre particulière à noyaux de convolution.

- Nous pouvons estimer $p_{\theta_0}(Y_t|Y_{1:t-1})$ par :

$$\hat{l}_{0,t}^{N,m} = \frac{1}{m} \sum_{k=1}^m p_{\theta_0}(Y_t|x_{t,0}(k)) ,$$

avec $x_{t,0}(k) \sim \hat{p}_{\theta_0}(x_t|Y_{1:t-1})$, $k = 1, \dots, m$.

- On remarque que :

$$p_{\theta_0}(Y_t|Y_{1:t-1}) = \int p_{\theta_0}(Y_t|x_t) \cdot p_{\theta_0}(x_t|Y_{1:t-1}) dx_t ,$$

- On estime le filtre prédictif $p_{\theta_0}(x_t|Y_{1:t-1})$
 \Rightarrow filtre particulière à noyaux de convolution.

- Nous pouvons estimer $p_{\theta_0}(Y_t|Y_{1:t-1})$ par :

$$\hat{l}_{0,t}^{N,m} = \frac{1}{m} \sum_{k=1}^m p_{\theta_0}(Y_t|x_{t,0}(k)) ,$$

avec $x_{t,0}(k) \sim \hat{p}_{\theta_0}(x_t|Y_{1:t-1})$, $k = 1, \dots, m$.

L'algorithme (modèle sous H_0)

Pour $i = 1, \dots, N$ et $j = 1, \dots, m$,

Etape 1 : Génération de $\bar{X}_{t-1, H_0}(i) \sim \hat{p}_{\theta_0}(x_{t-1} | Y_{1:t-1})$ (filtre optimal)

$$\begin{aligned} \text{modèle sous } H_0 \quad \Rightarrow \quad & \tilde{X}_{t, H_0}(i) = f(\bar{X}_{t-1, H_0}(i), \theta_0, w_t) \\ & \tilde{Y}_{t, H_0}(i) = h(\tilde{X}_{t, H_0}(i), \theta_0, v_t) \end{aligned}$$

L'algorithme (modèle sous H_0)

Etape 2 :

- On estime la densité prédictive :

$$\hat{p}_{\theta_0}(x_t | Y_{1:t-1}) = \frac{\sum_{i=1}^N K_{\delta_y}^1(\tilde{Y}_{t-1, H_0}(i) - Y_{t-1}) \cdot K_{\delta_x}^2(\tilde{X}_{t, H_0}(i) - x_t)}{\sum_{i=1}^N K_{\delta_y}^1(\tilde{Y}_{t-1, H_0}(i) - Y_{t-1})}$$

- Génération de $x_{t,0}(k) \sim \hat{p}_{\theta_0}(x_t | Y_{1:t-1})$, $k = 1, \dots, m$
- A partir de la mesure Y_t :

$$\begin{aligned} \Rightarrow \hat{l}_{0,t}^{N,m} &= \frac{1}{m} \sum_{k=1}^m p_{\theta_0}(Y_t | x_{t,0}(k)) \\ &\rightarrow p_{\theta_0}(Y_t | Y_{1:t-1}) \quad \text{quand } N \text{ et } m \rightarrow \infty \end{aligned}$$

Parallèlement, on obtient $\hat{l}_{1,t,j}^{N,m}$ pour $j = 1, \dots, t$.

Etape 3 : Statistique de test :

$$\hat{g}_t^{N,m} = \max_{1 \leq j \leq t} \sum_{i=j}^t \log \frac{\hat{l}_{1,t,j}^{N,m}}{\hat{l}_{0,t}^{N,m}}$$

⇒ Alarme si $\hat{g}_t^{N,m} \geq h$. Sinon, on passe à l'étape 4.

Etape 4 : Estimation du filtre optimal à partir de l'observation Y_t

$$\hat{p}_{\theta_0}(x_t | Y_{1:t}) = \frac{\sum_{i=1}^N K_{\delta_y}^1(\tilde{Y}_{t,H_0}(i) - Y_t) \cdot K_{\delta_x}^2(\tilde{X}_{t,H_0}(i) - x)}{\sum_{i=1}^N K_{\delta_y}^1(\tilde{Y}_{t,H_0}(i) - Y_t)}$$

où K^1 et K^2 sont des noyaux et δ_y et δ_x leurs fenêtres.

- $t=t+1$ puis retour à l'étape 1.

Etape 3 : Statistique de test :

$$\hat{g}_t^{N,m} = \max_{1 \leq j \leq t} \sum_{i=j}^t \log \frac{\hat{l}_{1,t,j}^{N,m}}{\hat{l}_{0,t}^{N,m}}$$

⇒ Alarme si $\hat{g}_t^{N,m} \geq h$. Sinon, on passe à l'étape 4.

Etape 4 : Estimation du filtre optimal à partir de l'observation Y_t

$$\hat{p}_{\theta_0}(x_t | Y_{1:t}) = \frac{\sum_{i=1}^N K_{\delta_y}^1(\tilde{Y}_{t,H_0}(i) - Y_t) \cdot K_{\delta_x}^2(\tilde{X}_{t,H_0}(i) - x)}{\sum_{i=1}^N K_{\delta_y}^1(\tilde{Y}_{t,H_0}(i) - Y_t)}$$

où K^1 et K^2 sont des noyaux et δ_y et δ_x leurs fenêtres.

- $t=t+1$ puis retour à l'étape 1.

Théorème 4

Sous les hypothèses du Théorème de convergence du filtre à noyaux de convolution (Rossi et Vila, 2005), on a lorsque m et N tendent vers l'infini:

$$\widehat{l}_{0,n}^{N,m} \rightarrow p_{\theta_0}(Y_n | Y_{1:n-1}) \text{ p.s.}$$

$$\widehat{l}_{1,n,j}^{N,m} \rightarrow p_{\theta_0, \theta_{1,j}}(Y_n | Y_{1:n-1}, j) \text{ p.s.} \quad \forall 1 \leq j \leq n.$$

Corollaire

Si les conditions de stabilité du rapport de vraisemblance sont vérifiées, la règle de décision

$$\widehat{t} = \inf \left\{ n : \max_{1 \leq j \leq n} \sum_{i=j}^n \log \frac{\widehat{l}_{1,i,j}^{N,m}}{\widehat{l}_{0,i}^{N,m}} \geq h \right\},$$

est asymptotiquement optimale (sur N , m et γ) au sens de Lorden.

Théorème 4

Sous les hypothèses du Théorème de convergence du filtre à noyaux de convolution (Rossi et Vila, 2005), on a lorsque m et N tendent vers l'infini:

$$\widehat{l}_{0,n}^{N,m} \rightarrow p_{\theta_0}(Y_n | Y_{1:n-1}) \text{ p.s.}$$

$$\widehat{l}_{1,n,j}^{N,m} \rightarrow p_{\theta_0, \theta_{1,j}}(Y_n | Y_{1:n-1}, j) \text{ p.s.} \quad \forall 1 \leq j \leq n.$$

Corollaire

Si les conditions de stabilité du rapport de vraisemblance sont vérifiées, la règle de décision

$$\widehat{t} = \inf \left\{ n : \max_{1 \leq j \leq n} \sum_{i=j}^n \log \frac{\widehat{l}_{1,i,j}^{N,m}}{\widehat{l}_{0,i}^{N,m}} \geq h \right\},$$

est asymptotiquement optimale (sur N , m et γ) au sens de Lorden.

Problème

A chaque pas de temps, un filtre supplémentaire.

Deux solutions envisagées :

- Considérer une règle de décision à fenêtre limitée M : Ainsi, on considère $M + 2$ filtres en parallèle qui fournissent $\hat{l}_{0,t}^{N,m}$ et $\hat{l}_{1,t,j}^{N,m}$ pour $j = t - M, \dots, t$.
⇒ Approche optimale (via le CUSUM approché à fenêtre limitée).
- On néglige la dépendance de la densité conditionnelle sous H_1 vis à vis de l'instant de panne.
Plus d'optimalité mais procédure beaucoup plus rapide (2 filtres en parallèle).
⇒ En simulation, résultats proches de la procédure précédente.

- On peut adapter l'algorithme au cas où θ_0 et θ_1 sont inconnus en considérant des modèles à espace d'état du type :

$$\left\{ \begin{array}{l} \begin{bmatrix} X_t \\ \theta_t \end{bmatrix} = \begin{bmatrix} f(X_{t-1}, \theta_{t-1}, w_t) \\ \theta_{t-1} \end{bmatrix} \\ Y_t = h(X_t, \theta_t, v_t) \end{array} \right.$$

où le paramètre θ est alors vu comme une variable aléatoire.

- Cas le plus fréquemment rencontré en pratique.
- Possibilité de poser un modèle de panne.

Cas d'un modèle autorégressif :

$$X_t = f(X_{t-1}, \theta) + \epsilon_t, \quad \text{où } \theta_1 \text{ est inconnu.}$$

Algorithme de filtrage-détection dans le seul but d'estimer θ

⇒ Modèle simplifié où la variable d'état est observée ($Y_t = X_t$) :

$$\left\{ \begin{array}{l} \begin{bmatrix} X_t \\ \theta_t \end{bmatrix} = \begin{bmatrix} f(X_{t-1}, \theta_{t-1}) + \epsilon_t \\ \theta_{t-1} \end{bmatrix} \\ Y_t = X_t \end{array} \right.$$

⇒ **Méthode concurrente directe du GLR :**

- Temps de calculs beaucoup plus rapide
- Résultats en simulations comparables
- Optimalité de la procédure au sens de Lorden

- 1 Contexte et Objectifs de la Thèse
- 2 La règle du CUSUM
 - Définition
 - Introduction d'un seuil adaptatif
- 3 Approximation du CUSUM
 - Définition et Résultats
 - Le CUSUM non paramétrique
 - Algorithmes de filtrage-détection
- 4 Applications
- 5 Conclusions et Perspectives

Le procédé de méthanisation

- Procédé de dépollution biologique : digestion anaérobie.
- Réaction entre deux substrats S_1 , S_2 , et deux biomasses X_1 et X_2 :

$$\begin{cases} \dot{X}_1 = (\mu_1(S_1) - \alpha D)X_1 \\ \dot{X}_2 = (\mu_2(S_2) - \alpha D)X_2 \\ \dot{S}_1 = D(S_1^{in} - S_1) - k_1\mu_1(S_1)X_1 \\ \dot{S}_2 = D(S_2^{in} - S_2) + k_2\mu_1(S_1)X_1 - k_3\mu_2(S_2)X_2 \\ \dot{Z} = D(Z_{in} - Z) \\ \dot{C}_{TI} = D(C_{TI}^{in} - C_{TI}) + k_4\mu_1(S_1)X_1 + k_5\mu_2(S_2)X_2 - q_{CO_2} \end{cases}$$

- Données provenant du LBE de Narbonne : 16 jours de mesure, 1 mesure toutes les deux minutes.
- Type de panne : panne de capteur (apparition d'un biais).

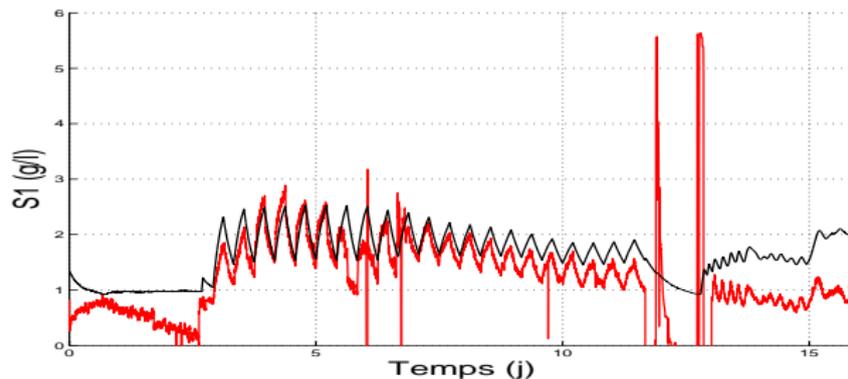
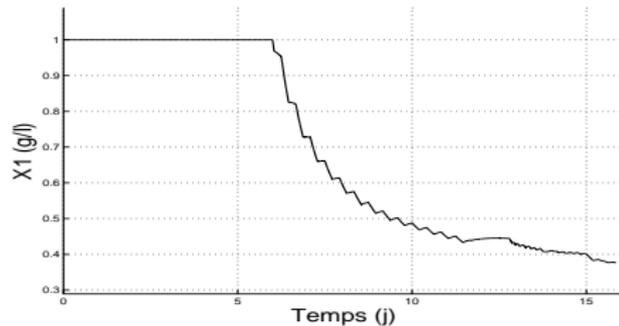
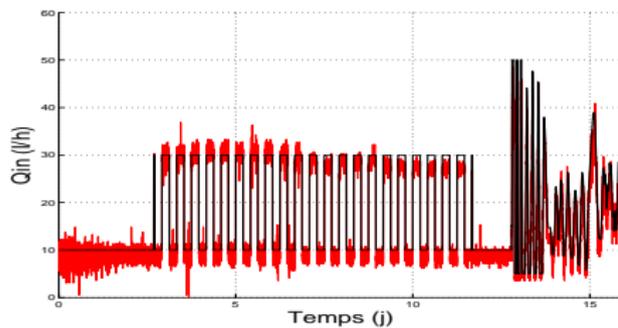
- Le modèle :

$$\begin{cases} \dot{X}_1 = (\mu_1 - \alpha D)X_1 \\ \dot{S}_1 = D(S_1^{in} - S_1) - k_1\mu_1 X_1 \end{cases}$$

- avec

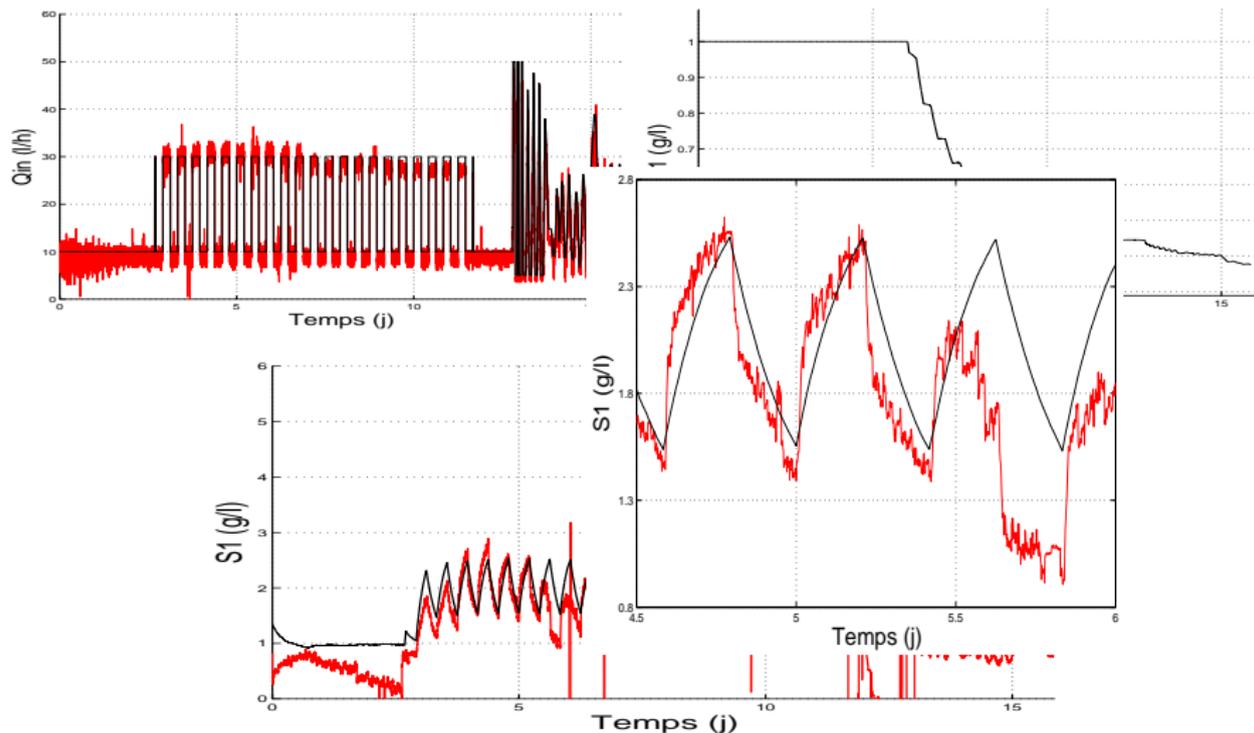
- X_1 concentration en biomasse.
- S_1 concentration en substrat.
- μ_1 taux de croissance de la biomasse et suit une loi de Monod.
- $D = Q_{in}/V$ joue le rôle de la variable de contrôle.
- valeurs des paramètres : Steyer et Bernard (2003).

Les données



Débit d'alimentation Q_{in} et concentrations en biomasse X_1 et en substrat S_1 (modèle en noir et données en rouge)

Les données et le biais sur le capteur



Débit d'alimentation Q_{in} et concentrations en biomasse X_1 et en substrat S_1 (modèle en noir et données en vert)

- Le modèle :

$$\left\{ \begin{array}{l} \left[\begin{array}{l} X_1(t+1) \\ S_1(t+1) \end{array} \right] = \left[\begin{array}{l} X_1(t) + T(\mu_1 - \alpha D(t))X_1(t) + v_{t+1} \\ S_1(t) + T(D(t)(S_1^{in} - S_1(t)) - k_1\mu_1 X_1(t)) + w_{t+1} \end{array} \right] \\ C_{S_1}(t+1) = S_1(t+1) - \theta + \epsilon_{t+1}, \end{array} \right.$$

v , w et ϵ sont des bruits blancs gaussiens indépendants de variances respectives : $\sigma_v^2 = 10^{-6}$, $\sigma_w^2 = 10^{-4}$ et $\sigma_\epsilon^2 = 10^{-3}$.

- Sous H_0 , $\theta = \theta_0 = 0$ et sous H_1 , $\theta = \theta_1 \in [0.6, 1.5]$.
- Algorithme de filtrage-détection avec $N = m = 1000$.

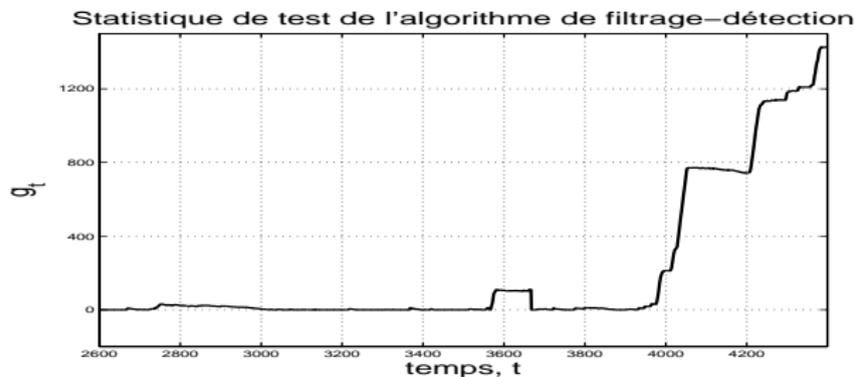
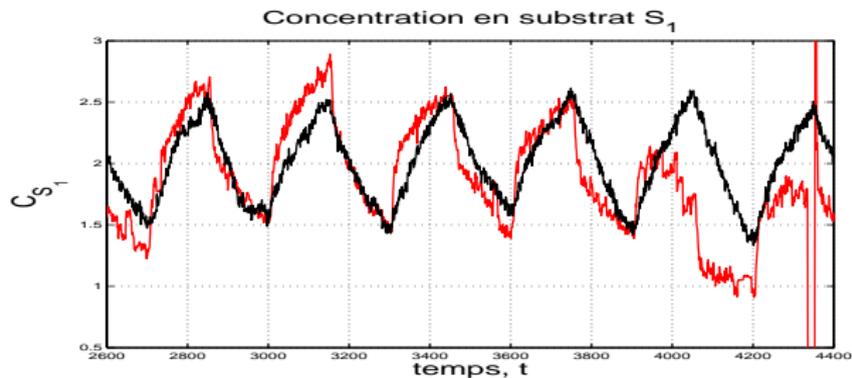
- Le modèle :

$$\left\{ \begin{array}{l} \left[\begin{array}{l} X_1(t+1) \\ S_1(t+1) \end{array} \right] = \left[\begin{array}{l} X_1(t) + T(\mu_1 - \alpha D(t))X_1(t) + v_{t+1} \\ S_1(t) + T(D(t)(S_1^{in} - S_1(t)) - k_1\mu_1 X_1(t)) + w_{t+1} \end{array} \right] \\ C_{S_1}(t+1) = S_1(t+1) - \theta + \epsilon_{t+1}, \end{array} \right.$$

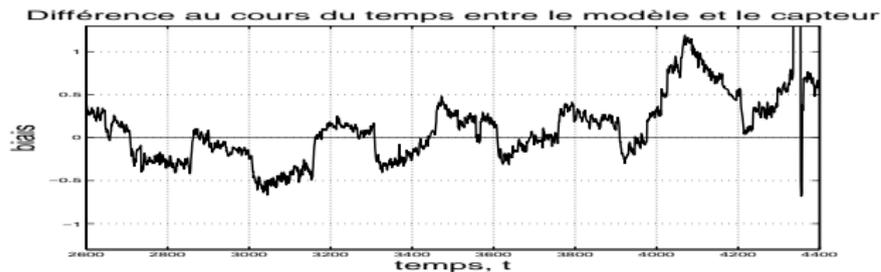
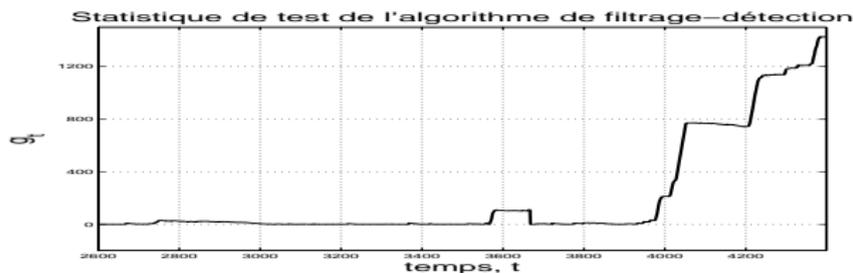
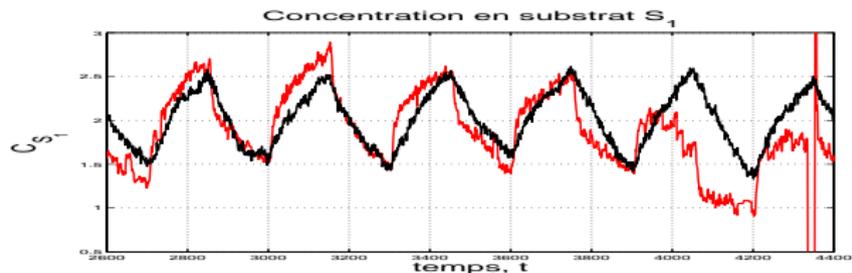
v , w et ϵ sont des bruits blancs gaussiens indépendants de variances respectives : $\sigma_v^2 = 10^{-6}$, $\sigma_w^2 = 10^{-4}$ et $\sigma_\epsilon^2 = 10^{-3}$.

- Sous H_0 , $\theta = \theta_0 = 0$ et sous H_1 , $\theta = \theta_1 \in [0.6, 1.5]$.
- Algorithme de filtrage-détection avec $N = m = 1000$.

Algorithme de filtrage-détection



Algorithme de filtrage-détection



Plan de l'exposé

- 1 Contexte et Objectifs de la Thèse
- 2 La règle du CUSUM
 - Définition
 - Introduction d'un seuil adaptatif
- 3 Approximation du CUSUM
 - Définition et Résultats
 - Le CUSUM non paramétrique
 - Algorithmes de filtrage-détection
- 4 Applications
- 5 Conclusions et Perspectives

- Mise au point d'un seuil adaptatif.
- Optimalité d'un CUSUM approché : applicable à de nombreuses situations.
- Mise au point d'une approche non paramétrique.
- Utilisation de filtres à noyaux de convolution pour les modèles à espace d'état.
- Comportement des méthodes proposées satisfaisant en pratique.

- Conditions de stabilité du rapport de vraisemblance.
- Fenêtre des estimateurs à noyaux.
- Prise en compte d'une variable de contrôle.
- Traiter le problème du diagnostic :
 - sur le plan théorique (Nikiforov (2003), Lai (2000))
 - sur le plan pratique (méthode avec moins de calculs notamment dans le cas du filtrage)
- Valorisation des procédures auprès des biotechnologues