

# Approches Particulières de la Statistique des HMM

## Applications en Biologie

Jean-Pierre Vila

INRA - UMR MISTEA

- 1 A propos de quelques problèmes en biologie / écologie
- 2 Une structure commune
- 3 Approches actuelles
- 4 Applications Microbiologiques
- 5 Extensions Inférentielles
- 6 Perspectives

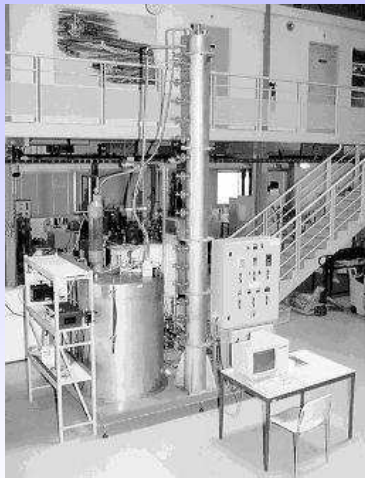
# Dépollution anaérobie (collaboration LBE-INRA)

## Objectifs :

- ▶ *Identification du système*
- ▶ *Contrôle prédictif*
- ▶ *Surveillance (dysfonctionnement)*

## Informations

disponibles : *mesures indirectes par capteur*



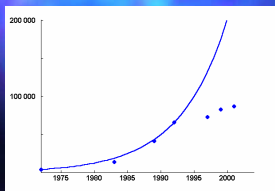
# Dynamique de Populations Animales (col. : CNRS)

## Objectifs :

- ▶ *Identification de systèmes*
- ▶ *Inférence paramétrique*
- ▶ *Sélection statistique de modèle*

### Gérer : Grands Cormorans hivernant en France

- Nombre d'individus hivernants en France : + 15 % par an, 1970-1990
- Augmentation similaire ailleurs (~18%)
- Impact sur les réserves de pêches



- Comprendre le fonctionnement de la dynamique
- Recommandations pour la gestion

Informations disponibles :

*comptages après capture-bagage-recapture*



# Un modèle commun : HMM

Systeme à espace d'état non linéaire  
observé indirectement :

(S)

$$\begin{cases} x_t \sim q_t(\cdot | x_{t-1}, \theta) \\ y_t \sim g_t(\cdot | x_t, \theta) \end{cases} \quad \text{ou} \quad \begin{cases} x_t = f_t(x_{t-1}, \theta, \varepsilon_t) \\ y_t = r_t(x_t, \theta, \eta_t) \end{cases}$$

$$x \in \mathbb{R}^d, \quad \theta \in \mathbb{R}^p, \quad y \in \mathbb{R}^s$$

Problème :

$$g_t(\cdot | x_t, \theta) : ?$$

# Un modèle commun : HMM

Systeme à espace d'état non linéaire  
observé indirectement :

(S)

$$\begin{cases} x_t \sim q_t(\cdot | x_{t-1}, \theta) \\ y_t \sim g_t(\cdot | x_t, \theta) \end{cases} \quad \text{ou} \quad \begin{cases} x_t = f_t(x_{t-1}, \theta, \varepsilon_t) \\ y_t = r_t(x_t, \theta, \eta_t) \end{cases}$$

$$x \in \mathbf{R}^d, \quad \theta \in \mathbf{R}^p, \quad y \in \mathbf{R}^s$$

Problème :

$$g_t(\cdot | x_t, \theta) : ?$$

# Un modèle commun : HMM

Systeme à espace d'état non linéaire  
observé indirectement :

(S)

$$\begin{cases} x_t \sim q_t(\cdot | x_{t-1}, \theta) \\ y_t \sim g_t(\cdot | x_t, \theta) \end{cases} \quad \text{ou} \quad \begin{cases} x_t = f_t(x_{t-1}, \theta, \varepsilon_t) \\ y_t = r_t(x_t, \theta, \eta_t) \end{cases}$$

$$x \in \mathbf{R}^d, \quad \theta \in \mathbf{R}^p, \quad y \in \mathbf{R}^s$$

Problème :

$$g_t(\cdot | x_t, \theta) : ?$$



## ⇒ inadaptation des techniques classiques

- ▶ *d'identification*  
(MC, MV, EM, MCMC, Filtrage...)
- ▶ *d'inférence*  
(test de rapport de vraisemblances,...)
- ▶ *de comparaison/sélection de modèles*  
(AIC, BIC, Facteur de Bayes,...)
- ▶ *de détection de rupture de modèle*  
(test CUSUM,...)

# Des approches actuelles :

# Des approches actuelles :

## Méthodes SMC-ABC :

Del Moral et al. 2011, Jasra et al. 2011

Principe de base :

*Echantillonnages SMC dans des approximations ABC  
d'une succession de densités cibles*

# Des approches actuelles :

## Filtrage à Convolution de Particules

Principe de base :

*Estimations récursives non paramétriques des densités conditionnelles des variables d'état et des paramètres*

# Filtrage à Convolution de Particules :

# Filtrage à Convolution de Particules :

Objectif 1 : estimation en tout temps  $t$  de

▶  $p_t(x_t|y_1, \dots, y_t)$  et  $E(x_t|y_1, \dots, y_t)$

▶  $p_t(\theta_t|y_1, \dots, y_t)$  et  $E(\theta_t|y_1, \dots, y_t)$

# Filtrage à Convolution de Particules :

## Objectif 2 : combinaison avec les méthodes

- ▶ *d'inférence statistique*
- ▶ *de comparaison/sélection de modèles*
- ▶ *de détection statistique de rupture de modèle*
- ▶ *de contrôle stochastique*

# Filtrage et extensions statistiques :

Projet en développement :

Plate-forme logicielle *FILTREX*



# Principe général : estimation fonctionnelle non paramétrique des densités conditionnelles d'état

# Principe général : estimation fonctionnelle non paramétrique des densités conditionnelles d'état

## Exemple :

*Filtre à convolution de particules avec ré-échantillonnage*

# Principe général : estimation fonctionnelle non paramétrique des densités conditionnelles d'état

A chaque instant  $t > 0$  :

1. *Prédiction* : simulation de  $n$  particules

$$(\tilde{x}_t^i, \tilde{\theta}_t^i) \sim p_t^n(x, \theta | y_{1:t-1}), \quad i = 1, \dots, n:$$

▶ si  $t = 1$  :

$$\tilde{x}_1^i \sim q_1(x_0^i, \theta_0^i), \quad \tilde{\theta}_1^i = \theta_0^i, \quad \tilde{y}_1^i \sim g_1(\cdot | \tilde{x}_1^i, \tilde{\theta}_1^i)$$

▶ si  $t > 1$  :

$$(x_{t-1}^i, \theta_{t-1}^i) \sim p_{t-1}^n(x, \theta | y_{1:t-1})$$

$$\tilde{x}_t^i \sim q_t(x_{t-1}^i, \theta_{t-1}^i), \quad \tilde{\theta}_t^i = \theta_{t-1}^i, \quad \tilde{y}_t^i \sim g_t(\cdot | \tilde{x}_t^i, \tilde{\theta}_t^i)$$

# Principe général : estimation fonctionnelle non paramétrique des densités conditionnelles d'état

A chaque instant  $t > 0$  :

2. Correction :

$$p_t^n(x|y_{1:t}) = \frac{\sum_{i=1}^n K_h^y(\tilde{y}_t^i - y_t) K_h^x(\tilde{x}_t^i - x)}{\sum_{i=1}^n K_h^y(\tilde{y}_t^i - y_t)}$$

$$p_t^n(\theta|y_{1:t}) = \frac{\sum_{i=1}^n K_h^y(\tilde{y}_t^i - y_t) K_h^\theta(\tilde{\theta}_t^i - \theta)}{\sum_{i=1}^n K_h^y(\tilde{y}_t^i - y_t)}$$

$t = t + 1$

# Propriétés principales

# Propriétés principales

## Convergences fonctionnelles $L_1$ p.s

$$\left\{ \begin{array}{l} \lim_{n \rightarrow \infty} \frac{nh_n^{s+d+p}}{\log n} = \infty \\ \lim_{n \rightarrow \infty} h_n = 0 \end{array} \right. \implies$$

$$\lim_{n \rightarrow \infty} \|p_t^n(x|y_{1:t}) - p_t(x|y_{1:t})\|_{L_1} = 0 \quad \text{p.s.}$$

$$\lim_{n \rightarrow \infty} \|p_t^n(\theta|y_{1:t}) - p_t(\theta|y_{1:t})\|_{L_1} = 0 \quad \text{p.s.}$$

(Rossi & Vila 2005, 2006)

# Propriétés principales

## Convergences ponctuelles p.s.

$$\begin{cases} h_n = O(n^{-\beta/(s+d+2p)}) \\ 0 < \beta < 1/2 \end{cases} \implies$$

$$\lim_{n \rightarrow \infty} p_t^n(x|y_{1:t}) = p_t(x|y_{1:t}) \quad p.s.$$

$$\lim_{n \rightarrow \infty} p_t^n(\theta|y_{1:t}) = p_t(\theta|y_{1:t}) \quad p.s.$$

(Vila 2011)

# Propriétés principales

## Convergences ponctuelles p.s.

$$\forall t \text{ soit } \hat{x}_t^n = \frac{1}{n} \sum_{i=1}^n x_t^i \quad \text{et} \quad \hat{\theta}_t^n = \frac{1}{n} \sum_{i=1}^n \theta_t^i$$

alors

$$\lim_{n \rightarrow \infty} |\hat{x}_t^n - \mathbb{E}[x_t | y_{1:t}]| = 0 \quad \text{p.s.}$$

$$\lim_{n \rightarrow \infty} |\hat{\theta}_t^n - \mathbb{E}[\theta_t | y_{1:t}]| = 0 \quad \text{p.s.}$$

De plus :

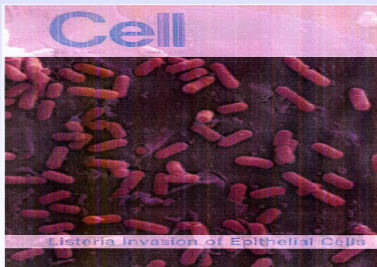
$$\lim_{t \rightarrow \infty} \mathbb{E}[\theta | y_{1:t}] = \theta^* \quad \text{p.s.} \quad \implies \quad \lim_{n, t \rightarrow \infty} |\hat{\theta}_t^n - \theta^*| = 0 \quad \text{p.s.}$$



# PROBLEMATIQUE MICROBIOLOGIQUE

Milieu d'origine : substrat alimentaire

Ex : contamination par *Listeria Monocytogenes*



# Recueil de données

## Séries de prélèvements et de dilutions successives

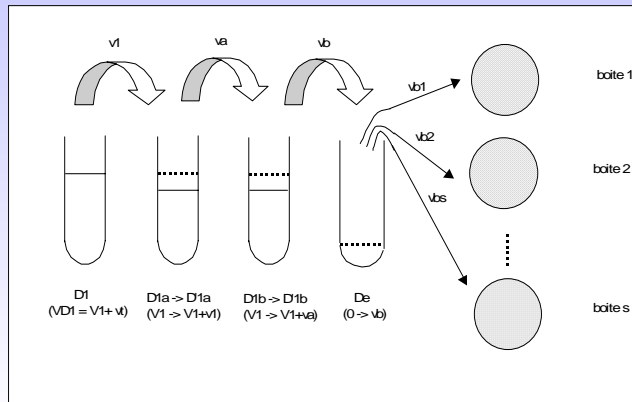


Schéma de dilution d'un tube  $D_1$ .

# Recueil de données - suite



## Comptage bactérien en milieu de culture



# Dynamique de *Listeria Monocytogenes*

## Modèle Primaire : Modèle de Baranyi II

$$\text{▶ } X_{t+1} = \delta x_0 \exp(\mu_{max} A_t) \frac{1}{B_t} \left( \mu_{max} \frac{dA_t}{dt} - \frac{dB_t}{dt} \frac{1}{B_t} \right) + X_t + \Phi_t$$

avec

$$\text{▶ } A_t = t + \frac{1}{\mu_{max}} \ln(\exp(-\mu_{max} t) + \exp(-\mu_{max} \lambda) - \exp(-\mu_{max} t - \mu_{max} \lambda))$$

$$\text{▶ } B_t = 1 + \frac{\exp(\mu_{max} A_t) - 1}{\frac{x_{max}}{x_0}}$$

Comptage :  $y_{t+1} \sim g_{t+1}(\cdot | x_{t+1}, x_0, x_{max}, \mu_{max}, \lambda)$

# Dynamique de *Listeria Monocytogenes*

## Modèle Primaire : Modèle de Baranyi II

$$\triangleright x_{t+1} = \delta x_0 \exp(\mu_{max} A_t) \frac{1}{B_t} \left( \mu_{max} \frac{dA_t}{dt} - \frac{dB_t}{dt} \frac{1}{B_t} \right) + x_t + \phi_t$$

avec

$$\triangleright A_t = t + \frac{1}{\mu_{max}} \ln(\exp(-\mu_{max} t) + \exp(-\mu_{max} \lambda) - \exp(-\mu_{max} t - \mu_{max} \lambda))$$

$$\triangleright B_t = 1 + \frac{\exp(\mu_{max} A_t) - 1}{\frac{x_{max}}{x_0}}$$

Comptage :  $y_{t+1} \sim g_{t+1}(\cdot | x_{t+1}, x_0, x_{max}, \mu_{max}, \lambda)$

# Dynamique de *Listeria Monocytogenes*

## Modèle Primaire : Modèle de Baranyi II

$$\triangleright x_{t+1} = \delta x_0 \exp(\mu_{max} A_t) \frac{1}{B_t} \left( \mu_{max} \frac{dA_t}{dt} - \frac{dB_t}{dt} \frac{1}{B_t} \right) + x_t + \phi_t$$

avec

$$\triangleright A_t = t + \frac{1}{\mu_{max}} \ln(\exp(-\mu_{max} t) + \exp(-\mu_{max} \lambda) - \exp(-\mu_{max} t - \mu_{max} \lambda))$$

$$\triangleright B_t = 1 + \frac{\exp(\mu_{max} A_t) - 1}{\frac{x_{max}}{x_0}}$$

Comptage :  $y_{t+1} \sim g_{t+1}(\cdot | x_{t+1}, x_0, x_{max}, \mu_{max}, \lambda)$

# Dynamique de *Listeria Monocytogenes*

## Modèle Primaire : Modèle de Baranyi II

$$\triangleright x_{t+1} = \delta x_0 \exp(\mu_{max} A_t) \frac{1}{B_t} \left( \mu_{max} \frac{dA_t}{dt} - \frac{dB_t}{dt} \frac{1}{B_t} \right) + x_t + \phi_t$$

avec

$$\triangleright A_t = t + \frac{1}{\mu_{max}} \ln(\exp(-\mu_{max} t) + \exp(-\mu_{max} \lambda) - \exp(-\mu_{max} t - \mu_{max} \lambda))$$

$$\triangleright B_t = 1 + \frac{\exp(\mu_{max} A_t) - 1}{\frac{x_{max}}{x_0}}$$

Comptage :  $y_{t+1} \sim g_{t+1}(\cdot | x_{t+1}, x_0, x_{max}, \mu_{max}, \lambda)$

# Dynamique de *Listeria Monocytogenes*

## Modèle Primaire : Modèle de Baranyi II

$$\triangleright x_{t+1} = \delta x_0 \exp(\mu_{max} A_t) \frac{1}{B_t} \left( \mu_{max} \frac{dA_t}{dt} - \frac{dB_t}{dt} \frac{1}{B_t} \right) + x_t + \phi_t$$

avec

$$\triangleright A_t = t + \frac{1}{\mu_{max}} \ln(\exp(-\mu_{max} t) + \exp(-\mu_{max} \lambda) - \exp(-\mu_{max} t - \mu_{max} \lambda))$$

$$\triangleright B_t = 1 + \frac{\exp(\mu_{max} A_t) - 1}{\frac{x_{max}}{x_0}}$$

Comptage :  $y_{t+1} \sim g_{t+1}(\cdot | x_{t+1}, x_0, x_{max}, \mu_{max}, \lambda)$



# Données : ENV-Alfort (J.C. Augustin)

# Données : ENV-Alfort (J.C. Augustin)

## Protocole :

- ▶ 10 instants de prélèvement (heures) :  
1, 72, 120, 168, 240, 264, 288, 336, 408,  
504.
- ▶ 3 prélèvements par instants.
- ▶ Facteurs de dilution :  
 $1(1 - 120)$ ,  $0.1(168)$ ,  $10^{-3}(240 - 264)$ ,  
 $10^{-4}(288 - 408)$ ,  $10^{-5}(504)$ .

# Données : ENV-Alfort (J.C. Augustin)

## Lois a priori :

- ▶  $\mu_{max} \sim U [0.01 - 2.]$
- ▶  $\lambda \sim U [20. - 50.]$
- ▶  $x_0 \sim U [100 - 400]$
- ▶  $x_{max} \sim U [10^8 - 10^9]$

# Données : ENV-Alfort (J.C. Augustin)

Pas de temps du filtrage :

$$\delta = 24h.$$

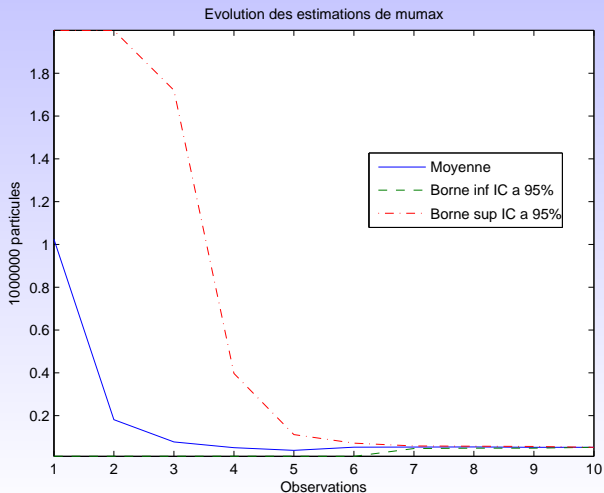
# Données : ENV-Alfort (J.C. Augustin)

Nombre de particules :

$$n = 10^5.$$

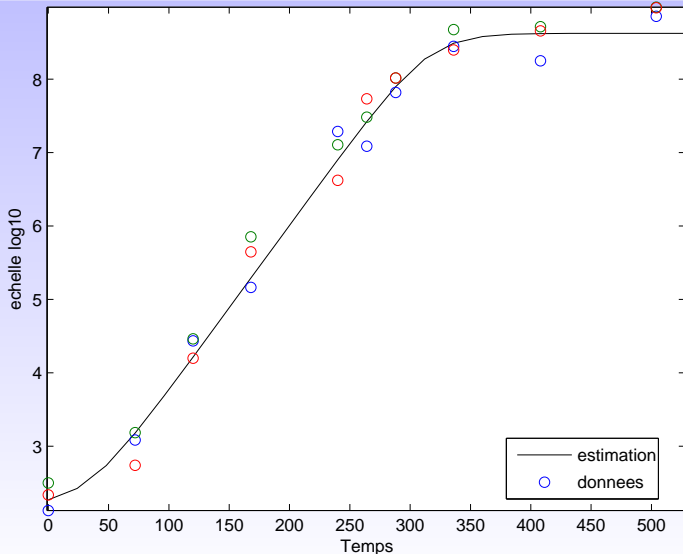
# Identification du modèle de Baranyi :

## Estimation par filtrage de $\mu_{max}$ : 0.051780



# Résultats du filtrage : estimation de la cinétique de Baranyi

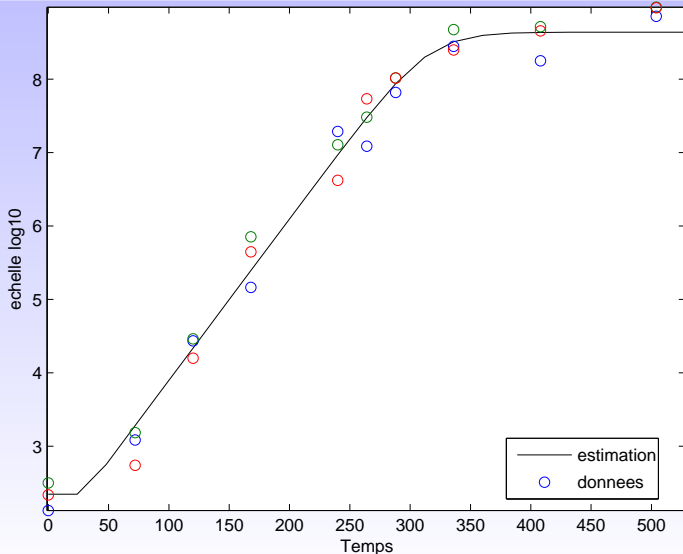
concentration



Temps d'exécution : 66 s. (pentium)

# Résultats du filtrage : estimation de la cinétique de Rosso

concentration

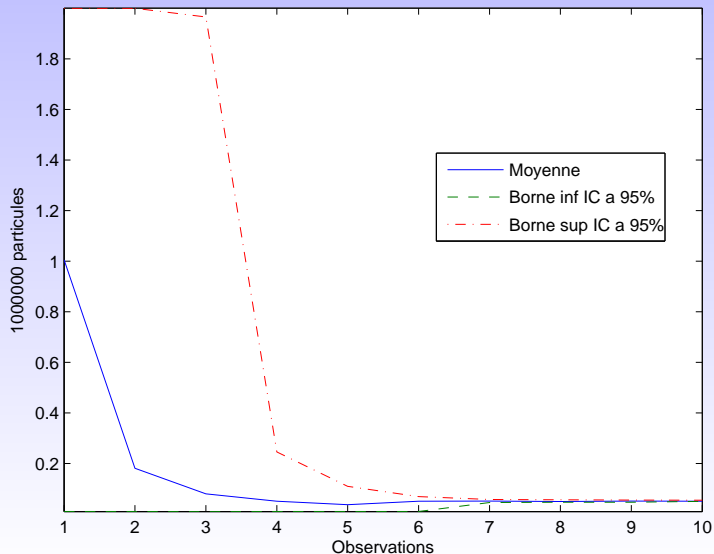


Temps d'exécution : 64 s. (pentium)



# Estimation finale de $\mu_{max}$ : 0.050686

Evolution des estimations de mumax



# Extensions inférentielles du filtrage à convolution

# Extensions inférentielles du filtrage à convolution

- ▶ *Sélection bayésienne de modèles :  
estimation particulière de Facteurs de Bayes*
- ▶ *Construction de prédicteurs particulières à  $k$  pas  
en avant :*  
$$p_{t+k}^n(x|y_1, \dots, y_t) \quad \text{et} \quad p_{t+k}^n(\theta|y_1, \dots, y_t)$$
- ▶ *Détection de rupture de modèle  
(test séquentiel CUSUM,...)*

# Comparaison / Sélection bayésienne particulière de modèles dynamiques

- ▶ Soit :  $M_1, M_2$  deux modèles probabilistes candidats.
- ▶ Soit :  $Y = (y_1, \dots, y_T) = y_{1:T}$ ,  $T$  observations.

# Comparaison / Sélection bayésienne particulière de modèles dynamiques

- ▶ Soit :  $M_1, M_2$  deux modèles probabilistes candidats.
- ▶ Soit :  $Y = (y_1, \dots, y_T) = y_{1:T}$ ,  $T$  observations.

Rappel : *Facteur de Bayes*

$$B_{12} = \frac{p_1(Y)}{p_2(Y)}$$

avec  $p_i(Y) = \int_{\Theta_i} p_i(Y|\theta) p_i(\theta) d\theta \quad i = 1, 2.$

# Comparaison / Sélection de Modèles - suite

Remarque :

$$p(Y) = p(y_{1:T}) = p(y_1) \prod_{t=1}^{T-1} p(y_{t+1} | y_{1:t})$$

# Comparaison / Sélection de Modèles - suite

Remarque :

$$p(Y) = p(y_{1:T}) = p(y_1) \prod_{t=1}^{T-1} p(y_{t+1} | y_{1:t})$$

⇒ *Estimation convergente du Facteur de Bayes :*

$$B_{12}^n = \frac{p_1^n(Y)}{p_2^n(Y)} \xrightarrow{n \rightarrow \infty} B_{12} = \frac{p_1(Y)}{p_2(Y)} \quad p.s.$$

avec

$$p^n(Y) = p_1^n(y_1) \prod_{t=1}^{T-1} p_{t+1}^n(y_{t+1} | y_{1:t}), \quad i = 1, 2$$

$$\text{et } p_{t+1}^n(y_{t+1} | y_{1:t}) = \frac{1}{n} \sum_{i=1}^n K_{h_n}^y(y_{t+1} - \tilde{y}_{t+1}^i)$$

# Application : Comparaison Baranyi / Rosso

$$B_{B/R}^n(t) = \frac{p_{Baranyi}^n(y_{0:t})}{p_{Rosso}^n(y_{0:t})}$$

Pour  $n = 10^4$

$t = 1, 72, 120, 168, 240, 264, 288, 336, 408, 504$  :

$t$	1	72	120	168	240
$B_{B/R}^n(t)$	0.934	0.938	0.941	0.944	0.944

$t$	264	288	336	408	504
$B_{B/R}^n(t)$	0.925	0.889	0.922	0.591	0.817



# Plan d'échantillonnage optimal pour le filtrage

# Plan d'échantillonnage optimal pour le filtrage

**Objectifs** : Favoriser l'estimation des paramètres par filtrage

⇒ choix des instants d'observations ?

**Principe** : Maximisation séquentielle de coefficients de sensibilité.

⇒ Accélération du rétrécissement des densités conditionnelles  $p_t(\theta|y_0, \dots, y_t)$  autour de leurs modes.

# Plan d'échantillonnage optimal pour le filtrage

Pour  $H$  temps d'observations futurs  $\in [0, T]$   
 et  $t_H \leq T$  :

- 1 Soit  $(t_1^*, t_2^*, \dots, t_H^*) =$   
 $\arg \max_{\{t < t_1 < t_2 < \dots < t_{H-1} < t_H < T\}} C(t_1, t_2, \dots, t_{H-1}, t_H)$

avec  $C(t_1, t_2, \dots, t_{H-1}, t_H) = \sum_{j=1}^{j=H} K(t_j)$ .

$$t \longleftarrow t_1^*$$

$$H \longleftarrow H - 1$$

- 2 Observation de  $y_t$   
 tant que  $H > 0$  aller à 1.

# Plan d'échantillonnage optimal pour le filtrage

*Choix de  $K$  :*

$$K(t) = Q(t)^T Q(t)$$

avec  $Q(t) = \left( S_1(t), S_2(t), \dots, S_j(t), \dots, S_p(t) \right)$ .

et  $S_j(t) = \frac{\text{Var}(\mathbb{E}[y_t | \theta_{j,t}])}{\text{Var}(y_t)}$

(coefficient de sensibilité de Sobol).

$S_j(t)$ ,  $j = 1, \dots, p$  : estimés par simulation particulière à partir de  $p_t^n(\theta_t | y_{t_1}, y_{t_2}, \dots, y_t)$ .

# Application à *Listeria Monocytogenes*

*Recherche d'un plan d'échantillonnage optimal selon le modèle de Baranyi*

- Domaine paramétrique :

$$[140; 260]_{x_0} ; [0.5 \times 10^8; 10 \times 10^8]_{x_{\max}} ; \\ [0.01; 0.09]_{\mu_{\max}} ; [10; 110]_{\lambda}$$

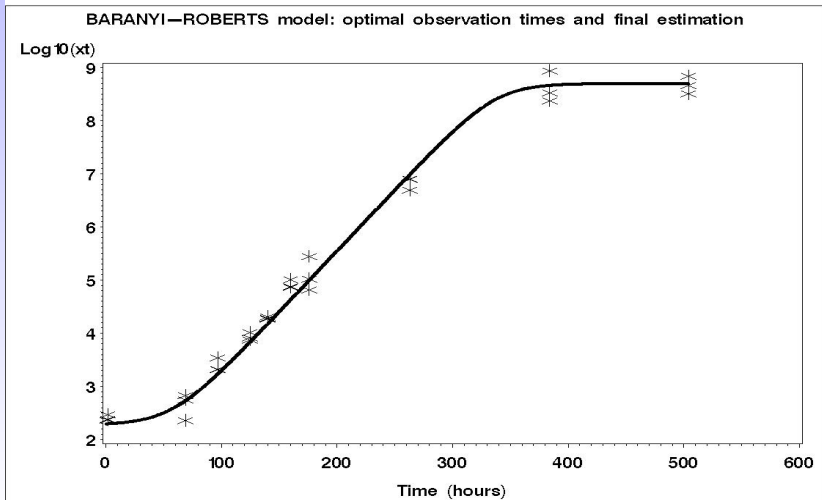
- Simulation de croissance avec :

$$x_0 = 200, x_{\max} = 5 \times 10^8, \mu_{\max} = 0.05, \lambda = 50.$$

$\implies$  Plan optimal pour le filtrage (pour  $H = 10$ ) :

$$t^* = 2, 69, 97, 125, 140, 160, 176, 263, 384, 504$$

# Application à *Listeria Monocytogenes*



# Application à *Listeria Monocytogenes*

Résultats :

(vraies valeurs :  $x_0 = 200$ ,  $x_{max} = 5 \times 10^8$ ,  $\mu_{max} = 0.050$ ,  $\lambda = 50$ )

paramètre	$x_0$	$x_{max}$	$\mu_{max}$	$\lambda$
<i>estimation</i>	198	$5.6 \times 10^8$	0.051	51.6
borne inférieure	193	$5.3 \times 10^8$	0.049	48.6
borne supérieure	204	$5.8 \times 10^8$	0.053	54.6

# Application à *Listeria Monocytogenes*

Résultats :

(vraies valeurs :  $x_0 = 200$ ,  $x_{max} = 5 \times 10^8$ ,  $\mu_{max} = 0.050$ ,  $\lambda = 50$ )

paramètre	$x_0$	$x_{max}$	$\mu_{max}$	$\lambda$
<i>estimation</i>	198	$5.6 \times 10^8$	0.051	51.6
borne inférieure	193	$5.3 \times 10^8$	0.049	48.6
borne supérieure	204	$5.8 \times 10^8$	0.053	54.6

A comparer à ceux obtenus par le protocole initial des microbiologistes :

paramètre	$x_0$	$x_{max}$	$\mu_{max}$	$\lambda$
<i>estimation</i>	196	$5.85 \times 10^8$	0.054	65.1
borne inférieure	191	$5.7 \times 10^8$	0.053	61.7
borne supérieure	200	$6.0 \times 10^8$	0.055	68.5



# Application à *Listeria Monocytogenes*

Résultats :

(vraies valeurs :  $x_0 = 200$ ,  $x_{max} = 5 \times 10^8$ ,  $\mu_{max} = 0.050$ ,  $\lambda = 50$ ) :

paramètre	$x_0$	$x_{max}$	$\mu_{max}$	$\lambda$
<i>estimation</i>	198	$5.6 \times 10^8$	0.051	51.6
borne inférieure	193	$5.3 \times 10^8$	0.049	48.6
borne supérieure	204	$5.8 \times 10^8$	0.053	54.6

# Application à *Listeria Monocytogenes*

Résultats :

(vraies valeurs :  $x_0 = 200$ ,  $x_{max} = 5 \times 10^8$ ,  $\mu_{max} = 0.050$ ,  $\lambda = 50$ ) :

paramètre	$x_0$	$x_{max}$	$\mu_{max}$	$\lambda$
<i>estimation</i>	198	$5.6 \times 10^8$	0.051	51.6
borne inférieure	193	$5.3 \times 10^8$	0.049	48.6
borne supérieure	204	$5.8 \times 10^8$	0.053	54.6

...et avec les estimations obtenues à partir d'un protocole naïf :  
10 temps équidistribués entre  $t = 0$  et  $t = 504$  :

paramètre	$x_0$	$x_{max}$	$\mu_{max}$	$\lambda$
<i>estimation</i>	195	$5.29 \times 10^8$	0.055	64.0
borne inférieure	191	$5.11 \times 10^8$	0.053	59.6
borne supérieure	199	$5.47 \times 10^8$	0.056	68.4

# Autres Extensions

## Inférence :

- ▶ Construction de *prédicteurs particuliers non paramétriques* à  $k$  pas en avant :

$$p_{t+k}^n(x|y_0, \dots, y_t) \text{ et } p_{t+k}^n(\theta|y_0, \dots, y_t)$$

- applications : *prédiction de prolifération bactérienne ; contrôle prédictif.*

## Inférence :

- ▶ *Détection de rupture de modèle : test CUSUM*
- application projetée :  
*variation inattendue des paramètres*

# Autres Extensions

## Inférence :

- ▶ Construction de *prédicteurs particuliers non paramétriques* à  $k$  pas en avant :

$$p_{t+k}^n(x|y_0, \dots, y_t) \quad \text{et} \quad p_{t+k}^n(\theta|y_0, \dots, y_t)$$

- applications : *prédiction de prolifération bactérienne ; contrôle prédictif.*

## Inférence :

- ▶ *Détection de rupture de modèle : test CUSUM*
- application projetée :  
*variation inattendue des paramètres*

# Autres Extensions

## Inférence :

- ▶ Construction de *prédicteurs particuliers non paramétriques* à  $k$  pas en avant :

$$p_{t+k}^n(x|y_0, \dots, y_t) \quad \text{et} \quad p_{t+k}^n(\theta|y_0, \dots, y_t)$$

- applications : *prédiction de prolifération bactérienne ; contrôle prédictif.*

## Inférence :

- ▶ *Détection de rupture de modèle : test CUSUM*  
→ application projetée :  
*variation inattendue des paramètres*

# Autres Extensions

## Inférence : Test CUSUM généralisé

Instant d'alarme  $\tau_C$  :

$$\tau_C = \inf\{t : \Gamma_t \geq \lambda\} \quad \text{avec} \quad \Gamma_t = \max_{1 < j \leq t} \sum_{i=j}^t Z_{i,j}$$

$$\text{et} \quad Z_{i,j} = \log \frac{p_{\theta_0, \theta_{1,j}}(y_i | y_{1:i-1})}{p_{\theta_0}(y_i | y_{1:i-1})},$$

$$\text{estimé par :} \quad Z_{i,j}^n = \log \frac{p_{\theta_0, \theta_{1,j}}^n(y_i | y_{1:i-1})}{p_{\theta_0}^n(y_i | y_{1:i-1})}$$

Propriétés : optimalité asymptotique ( $n \rightarrow \infty$ ).

# Autres Extensions

## Inférence : Test CUSUM généralisé

Instant d'alarme  $\tau_C$  :

$$\tau_C = \inf\{t : \Gamma_t \geq \lambda\} \quad \text{avec} \quad \Gamma_t = \max_{1 < j \leq t} \sum_{i=j}^t Z_{i,j}$$

$$\text{et} \quad Z_{i,j} = \log \frac{p_{\theta_0, \theta_{1,j}}(y_i | y_{1:i-1})}{p_{\theta_0}(y_i | y_{1:i-1})},$$

$$\text{estimé par : } Z_{i,j}^n = \log \frac{p_{\theta_0, \theta_{1,j}}^n(y_i | y_{1:i-1})}{p_{\theta_0}^n(y_i | y_{1:i-1})}$$

Propriétés : optimalité asymptotique ( $n \rightarrow \infty$ ).

# Travaux en cours :

## *contrôle prédictif par retour d'état filtré*

Critère : en  $t = j$ , pour un horizon  $H$

$$J_H(U_{j:j+H-1}) \\ = \sum_{t=j}^{j+H-1} \mathbb{E} \left[ c_{t+1}(x_{t+1}, U_t) \mid y_{1:j}, u_{0:j-1}, U_j, \dots, U_{t-1}, U_t \right]$$

Algorithme :

- 1 Soit  $u_j^*, \dots, u_{j+H-1}^* = \operatorname{argmin}_{U_j, \dots, U_{j+H-1}} J_H$ .
- 2 appliquer le contrôle  $u_j^*$  au système
- 3 enregistrer  $y_{j+1}$ .
- 4  $j = j + 1$ , retour à 1.



# Travaux en cours :

## *contrôle prédictif par retour d'état filtré*

Critère : en  $t = j$ , pour un horizon  $H$

$$J_H(U_{j:j+H-1}) \\ = \sum_{t=j}^{j+H-1} \mathbb{E} \left[ c_{t+1}(x_{t+1}, U_t) \mid y_{1:j}, u_{0:j-1}, U_j, \dots, U_{t-1}, U_t \right]$$

Algorithme :

- 1 Soit  $u_j^*, \dots, u_{j+H-1}^* = \operatorname{argmin}_{U_j, \dots, U_{j+H-1}} J_H$ .
- 2 appliquer le contrôle  $u_j^*$  au système
- 3 enregistrer  $y_{j+1}$ .
- 4  $j = j + 1$ , retour à 1.

# Travaux en cours :

## *contrôle prédictif par retour d'état filtré*

Critère : en  $t = j$ , pour un horizon  $H$

$$J_H(U_{j:j+H-1}) \\ = \sum_{t=j}^{j+H-1} \mathbb{E} \left[ c_{t+1}(x_{t+1}, U_t) \mid y_{1:j}, u_{0:j-1}, U_j, \dots, U_{t-1}, U_t \right]$$

Algorithme :

- 1 Soit  $u_j^*, \dots, u_{j+H-1}^* = \operatorname{argmin}_{U_j, \dots, U_{j+H-1}} J_H$ .
- 2 appliquer le contrôle  $u_j^*$  au système
- 3 enregistrer  $y_{j+1}$ .
- 4  $j = j + 1$ , retour à 1.

# Extensions microbiologiques

# Extensions microbiologiques

- ▶ Modélisation multi-espèce et inférence.
- ▶ Caractérisation de dysfonctionnements de dispositifs de conservation.
- ▶ Caractérisation et simulation d'évènements rares.
- ▶ Contrôle d'évolutions bactériennes critiques.

# Publications

- Bidot, C., Gauchi, J.P., Vila, J.P. (2009). Identification de systèmes dynamiques microbiologiques complexes par filtrage non linéaire. *41ième Journées de Statistique de la SFdS*, Bordeaux.
- Gauchi, J.P., Bidot, C., Augustin, J.C., Vila, J.P. (2009). Identification of complex microbiological dynamic system by nonlinear filtering. *6th Int. Conference on Predictive Modelling in Foods, Washington DC*.
- Gauchi, J.P., Vila, J.P., Coroller, L. (2010). New prediction confidence intervals and bands in the nonlinear regression model : Application to the predictive modelling in food. *Communications in Statistics, Simulation and Computation*, 39, 322-334.
- Gauchi, J.P., Vila, J.P. (2012). Nonparametric particle filtering approaches for identification and inference in nonlinear state-space dynamic systems. *Statistics and Computing, DOI 10.1007/s11222-012-9327-7*.
- Rossi, V. and Vila, J.P. (2005). Approche non paramétrique du filtrage de système non linéaire à temps discret et à paramètres inconnus, *C.R. Acad. Sci. Paris, I*, 340, 759-764.

# Publications - suite

- Rossi, V. and Vila, J.P. 2006. Nonlinear filtering in discrete time: A particle convolution approach. *Pub. Inst. Stat. Univ. Paris, L, 3, 71-102.*
- Verdier, G., Hilgert, N., Vila, J.P. (2008). optimality of CUSUM rule approximations in change-point problems : Application to nonlinear state-space systems. *IEEE Trans. on Inf. Theory, 54, 11, 5102-5112.*
- Vila, J.P., Saley, I. (2009). Estimation de facteurs de Bayes entre modèles dynamiques non linéaires à espace d'état. *C.R. Acad. Sci. Paris, Ser. I, 347, 429-434.*
- Vila, J.P., (2011). Nonparametric multi-step prediction in nonlinear state space dynamic systems. *Statistics and Probability Letters, 81, 71-76.*
- Vila, J.P, Gauchi, J.P. (2011). Predictive control of stochastic nonlinear state space dynamic systems : a particle nonparametric approach. *Technical Report, UMR MISTEA.*
- Vila, J.P., (2012). Enhanced consistency of the Resampled Convolution Filter. *Statistics and Probability Letters, 82, 786-797.*

**MERCI POUR VOTRE ATTENTION**