

Rubrique

## Analyse bayésienne de modèles markoviens d'évolution de ressources naturelles

Fabien Campillo\* — Rivo Rakotozafy\*\* — Vivien Rossi\*\*\*

\* INRIA/IRISA  
Campus de Beaulieu  
35042 Rennes  
FRANCE  
Fabien.Campillo@inria.fr

\*\* Faculté des Sciences  
Université de Fianarantsoa  
301 Fianarantsoa  
MADAGASCAR  
rrakotoz@mail.univ-fianar.mg

\*\*\* INRIA/IRISA  
Campus de Beaulieu  
35042 Rennes  
FRANCE  
Vivien.Rossi@inria.fr

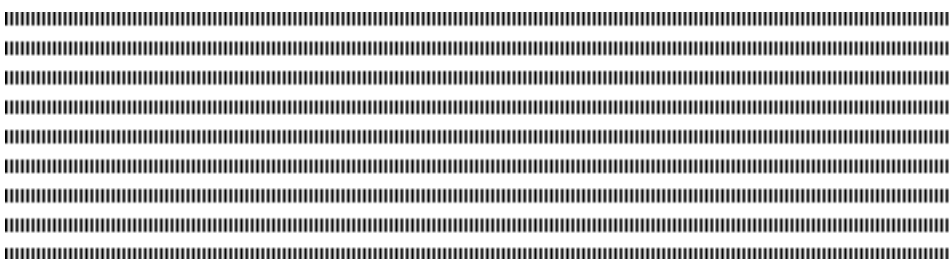


**RÉSUMÉ.** On applique des méthodes de Monte Carlo à des modèles à espace d'état avec paramètres inconnus. La première est un algorithme de Monte Carlo par chaîne de Markov. La seconde est le filtrage particulaire. Nous comparons ces méthodes sur un modèle d'évolution de la biomasse dans une pêcherie..

**ABSTRACT.** One applies Monte Carlo methods to state space models with unknown parameters. The first one is a Monte Carlo Markov Chain algorithm. The second one is the particle filtering. We compare these methods applied to a biomass evolution model for fisheries.

**MOTS-CLÉS :** Inférence bayésienne, Monte Carlo par chaînes de Markov, filtrage particulaire

**KEYWORDS :** Bayesian inference, Markov chain Monte Carlo, particle filtering



Considérons un modèle de Markov caché  $(x_{0:T}, y_{1:T})$  à valeurs dans  $\mathbb{R}^n \times \mathbb{R}^d$  et dépendant d'un paramètre  $\theta$  inconnu. Le processus d'état non observé  $x$  est une chaîne de Markov homogène de loi initiale

$$\mu^\vartheta(dx) = \mathbb{P}(x_0 \in dx | \theta = \vartheta)$$

et de noyau de transition

$$Q^\vartheta(x, dx') = \mathbb{P}(x_t \in dx' | \theta = \vartheta, x_{t-1} = x).$$

Le processus d'observation  $y$  est une suite de variables indépendantes conditionnellement à l'état  $x$  et au paramètre  $\theta$ , et la loi conditionnelle de  $y_t$  ne dépend que de  $x_t$  et  $\theta$ , i.e.

$$\mathbb{P}(y_{1:T} \in dy_{1:T} | \theta = \vartheta, x_{0:T} = x_{0:T}) = \prod_{t=1}^T \mathbb{P}(y_t \in dy_t | \theta = \vartheta, x_t = x_t).$$

On suppose de plus que la loi conditionnelle de  $y_t$  sachant  $x_t$  est absolument continue par rapport à la mesure de Lebesgue et on pose :

$$\mathbb{P}_\theta(y_t \in dy | \theta = \vartheta, x_t = x) = \psi_t^\vartheta(x, y) dy.$$

Les éléments du modèle, à savoir  $\mu^\vartheta$ ,  $Q^\vartheta$  et  $\psi_t^\vartheta$ , dépendent d'un paramètre inconnu  $\theta$  à valeurs dans  $\mathbb{R}^p$ . Ce paramètre est supposé indépendant de tous les autres termes et on se donne une loi a priori

$$\nu(d\theta) = \mathbb{P}(\theta \in d\theta).$$

Une des particularités des modèles d'évolution de ressources naturelles est que la taille  $T$  des échantillons est relativement faible, le plus souvent de l'ordre de quelques dizaines.

## 1. Méthodes de Monte Carlo par chaînes de Markov

Le but ici est de proposer, à partir de  $\mu^\vartheta$ ,  $Q^\vartheta$ ,  $\psi_{1:T}^\vartheta$  et  $\nu$ , un algorithme de construction d'une chaîne de Markov ergodique

$$(x_{0:T}^{(k)}, \theta^{(k)})_{k \in \mathbb{N}}$$

dont la loi invariante est la loi a posteriori

$$\pi_T(dx_{0:T}, d\theta | y_{1:T}) = \mathbb{P}(x_{0:T} \in dx_{0:T}, \theta \in d\theta | y_{1:T} = y_{1:T}). \quad (1)$$

Ainsi le théorème ergodique implique que

$$\frac{1}{K} \sum_{k=1}^K \phi(x_{0:T}^{(k)}, \theta^{(k)}) \xrightarrow[K \uparrow \infty]{\text{p.s.}} \mathbb{E}[\phi(x_{0:T}, \theta) | y_{1:T} = y_{1:T}]$$

On obtient ainsi des estimations de statistiques sur les variables latentes (cf. [4] pour plus de détails).

---

## 2. Méthodes de Monte Carlo séquentielles

Le contexte est ici légèrement différent, le but est de calculer séquentiellement une approximation particulière de

$$\rho_t(dx, d\vartheta|y_{1:t}) = \mathbb{P}(x_t \in dx, \theta_t \in d\vartheta|y_{1:t} = y_{1:t}). \quad (2)$$

Il suffit pour cela d'introduire un processus  $\theta_t$  dont la dynamique est  $\theta_t = \theta_{t-1}$  et la loi initiale est  $\nu$ . Le couple  $(x_t, \theta_t)_{t=0:T}$  est Markovien. On peut écrire le filtre bayésien optimal et séquentiel qui donne de façon récursive l'expression exacte de

$$\rho_t(dx, d\vartheta|y_{1:t}) = \mathbb{P}(x_t \in dx, \theta_t \in d\vartheta|y_{1:t} = y_{1:t}).$$

Cette expression n'est pas directement accessible, il est nécessaire de faire appel à une technique d'approximation. La technique rencontrant le plus de succès à l'heure actuelle est le filtrage particulaire (cf. [3] pour plus de détail). Elle permet d'échantillonner séquentiellement des particules  $(x_t^i, \theta_t^i)_{i=1:N}$  de telle façon que

$$\rho_t(dx, d\vartheta|y_{1:t}) \simeq \rho_t^N(dx, d\vartheta|y_{1:t}) = \frac{1}{N} \sum_{i=1}^N \delta_{(x_t^i, \theta_t^i)}(dx, d\vartheta)$$

---

## 3. Exemple : un modèle de pêche

On présente ici un exemple proposé dans [1].

Les techniques d'estimation des stocks de pêche visent à évaluer l'impact de différents scénarios d'exploitation et de capture sur l'évolution de l'abondance d'un stock de poissons. Cette abondance s'exprime en poids total de poissons vulnérables à la pêche (biomasse). L'accroissement de cette biomasse intègre l'augmentation du poids moyen de chaque individu ainsi que le recrutement, c'est à dire l'arrivée chaque année de nouvelles générations de poissons. À l'inverse, cette biomasse décroît du fait de la capture (mortalité par pêche) ou d'autres causes (mortalité naturelle : prédation, maladies, etc.).

Pour chaque année  $t = 1 \cdots T$  d'une période de  $T$  années,  $B_t$  désigne la biomasse en début d'année,  $C_t$  la biomasse capturée en cours d'année et  $I_t$  un indice d'abondance du stock. Cet indice, mesuré chaque année, est la capture par unité d'effort (CPUE). Les séries  $C_{1:T}$  et  $I_{1:T}$  sont données. Le but est de déterminer, au sein d'une classe de modèles, lequel s'ajuste le mieux aux observations.

À la suite des travaux de Meyer et Millar [7], on adopte une formulation à espace d'états non linéaire d'un modèle aux différences avec retard (delay difference model) de la dynamique de la biomasse et de l'observation de l'indice d'abondance de cette biomasse. Ce modèle a été introduit par Deriso [2] et généralisé par Schnute [8] (voir Hilborn et Walters [5] pour une présentation générale). Dans la représentation à espace d'états, les indices  $I_{1:T}$  seront considérées comme les observations,  $B_{1:T}$  comme les états (non observés) et  $C_{1:T}$  comme les entrées.

Les paramètres du modèle sont ajustés aux données par une procédure de Metropolis–Hastings “composante à composante”.

À l'année  $t$  la biomasse  $B_t$  est :

$$B_t = \sum_{a \geq k} w_a N_{a,t} \quad (3)$$

où  $N_{a,t}$  est la taille de la population d'âge  $a$  à l'année  $t$ ,  $w_a$  est le poids moyen d'un individu à l'âge  $a$ ,  $k$  est l'âge de recrutement (âge auquel on estime que les poissons viennent s'ajouter à la biomasse). Le premier terme  $R_t = w_k N_{k,t}$  dans la somme (3) est le recrutement de l'année  $t$ .

On suppose les individus de la biomasse uniformément vulnérables à la mortalité naturelle et à la capture. L'évolution annuelle de chaque cohorte (population d'individus de la même classe d'âge) est donc décrite par :

$$N_{a,t} = s_{t-1} N_{a-1,t-1}, \quad (4)$$

où  $s_{t-1}$  est le taux de survie à la mortalité naturelle et à la pêche. L'évolution de  $w_a$  est donnée selon la relation classique de croissance du poids selon l'âge :  $w_a = w_\infty (1 - e^{-\kappa(a-a_0)})$  pour  $a \geq a_0$  où  $a_0$  est l'âge auquel  $w_{a_0} = 0$  et  $\kappa > 0$ . Cela conduit à :

$$w_a = (1 + \rho) w_{a-1} - \rho w_{a-2} \quad (5)$$

où  $\rho = e^{-\kappa} \in (0, 1)$  est le taux de croissance. Les expressions (4) et (5) dans (3) conduisent à la récurrence :

$$B_t = (1 + \rho) s_{t-1} B_{t-1} - \rho s_{t-1} s_{t-2} B_{t-2} + R_t - \rho s_{t-1} w R_{t-1} \quad (6)$$

avec  $w = w_{k-1}/w_k \in (0, 1)$ .

On suppose que les causes de mortalité naturelle et de mortalité par capture agissent de façon indépendante sur l'ensemble des poissons, ainsi  $s_t = s_t^M s_t^F$  où le taux de survie à la mortalité naturelle  $s_t^M$  est supposé constant ( $s^M = e^{-M}$  avec  $M > 0$ ), et où le taux de survie à la capture est  $s_t^F = \frac{B_t - C_t}{B_t}$ . Le recrutement est ensuite supposé constant, i.e.  $R = R_t$ . L'équation (6) devient donc :

$$B_t = (1 + \rho) e^{-M} \frac{B_{t-1} - C_{t-1}}{B_{t-1}} B_{t-1} - \rho e^{-2M} \frac{B_{t-1} - C_{t-1}}{B_{t-1}} \frac{B_{t-2} - C_{t-2}}{B_{t-2}} B_{t-2} \\ + R \left( 1 - \rho e^{-M} w \frac{B_{t-1} - C_{t-1}}{B_{t-1}} \right).$$

Avant la première année  $t = 1$ , les captures sont supposées nulles  $C_{t < 0} = 0$ . La biomasse est supposée être à son équilibre jusqu'à l'année 1, i.e.  $B_{t \leq 1} = K$  où  $K$  est un paramètre inconnu (la biomasse vierge). La biomasse à l'année  $t = 2$  est donc :

$$B_2 = (1 + \rho - \rho e^{-M}) e^{-M} (B_1 - C_1) + R (1 - \rho e^{-M} w \frac{B_1 - C_1}{B_1}).$$

L'équation d'observation non bruitée est de la forme  $I_t = q B_t$  pour  $t = 1 \dots T$ , où  $I_t$  est un indice de biomasse relative et  $q$  est un coefficient de "capturabilité". Les processus de bruit d'état et d'observation sont multiplicatifs et lognormaux :

$$B_t = F_t(R, K, B_{t-1}, B_{t-2}) \times e^{\sigma_W W_t}, \quad (7)$$

$$I_t = q B_t \times e^{\sigma_V V_t}, \quad (8)$$

pour  $t = 1 \dots T$ ,  $W_{1:T}$  et  $V_{1:T}$  sont des bruits blancs gaussiens  $N(0, 1)$ . L'expression exacte de ces fonctions et ce modèle sont détaillés dans [1].

Pour revenir au modèle initial il suffit de poser

$$\theta \stackrel{\text{def}}{=} (K, R, q, \sigma_W^2, \sigma_V^2)$$

et

$$x_t \stackrel{\text{def}}{=} \begin{pmatrix} B_t \\ B_{t-1} \end{pmatrix}.$$

---

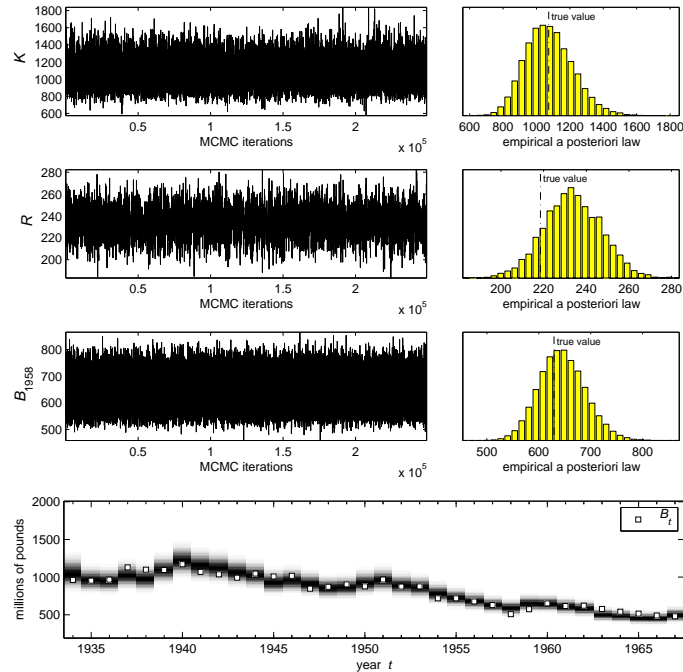
## 4. Résultats préliminaires

Nous appliquons les algorithmes précédents sur des données simulées comparables à celles utilisées dans Meyer et Millar [7] (eux mêmes font référence aux données de Pella–Tomlinson sur le thon albacore 1934–1967). Dans ces premiers essais, les variances  $\sigma_W^2$  et  $\sigma_V^2$  sont supposées connues.

La méthode MCMC retenue en pratique est un échantillonneur de Gibbs couplé avec un algorithme de Metropolis–Hastings à chaque itération. Les résultats présentés dans la Figure 1 sont bons. Il faut noter que l'algorithme converge très lentement : la chaîne tend très lentement vers son régime stationnaire.

Dans la Figure 1, nous comparons le filtre particulaire et la méthode MCMC. Les résultats sont comparables, on note toutefois que le filtre particulaire présente un peu plus de variance. Ceci est cohérent avec le fait que le filtre particulaire donne une approximation de  $\text{loi}(x_t, \theta | y_{1:t})$  alors que la méthode MCMC donne une approximation de  $\text{loi}(x_t, \theta | y_{1:T})$ .

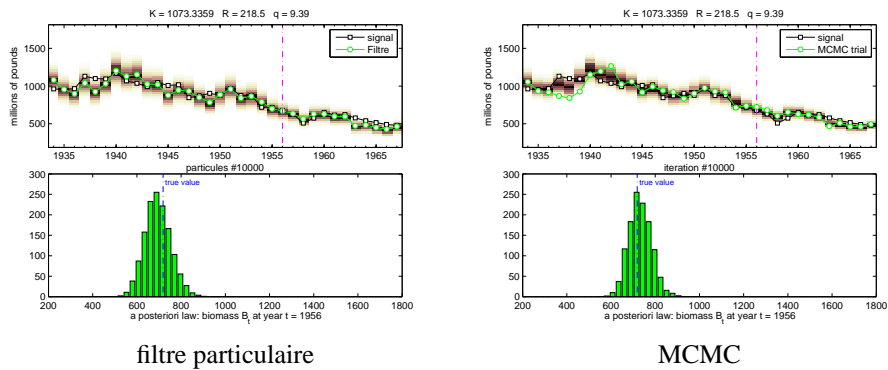
Dans la version finale de ce travail, nous présenterons des résultats de simulation plus détaillés. Il est également nécessaire d'améliorer les propriétés de convergence de la méthode MCMC, notamment sa vitesse de convergence.



**Figure 1.** Méthode MCMC : • Les figures du haut à gauche présente les trajectoires des chaînes  $k \mapsto (K^{(k)}, R^{(k)}, B_{1958}^{(k)})$ , elles permettent de constater que l'algorithme MCMC a atteint son régime d'équilibre. • Les figures du haut à droite présentent les lois a posteriori empiriques correspondant respectivement à  $K$ ,  $R$ ,  $B_{1958}$ . • La figure du bas présente l'évolution de la biomasse  $B_{1:T}$  ainsi, que pour chaque année  $t$ , la représentation en niveaux de gris de la loi a posteriori empirique associée.

## 5. Bibliographie

- [1] F. Campillo and R. Rakotozafy. MCMC for nonlinear/non-Gaussian state-space models, Application to fishery stock assessment. In *CARI'04*, Hammamet, Tunisia, 2004.
- [2] R. B. Deriso. Harvesting strategies and parameter estimation for an age-structured model. *Canadian Journal of Fisheries and Aquatic Sciences*, 37 :268–282, 1980.
- [3] Arnaud Doucet, Nando de Freitas, and Neil J. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. Springer-Verlag, New York, 2001.
- [4] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors. *Markov Chain Monte Carlo in practice*. Chapman & Hall, London, 1995.



**Figure 2.** Comparaison filtrage particulaire (gauche) et MCMC (droite). En haut l'évolution de la biomasse  $B_{1:T}$  ainsi, que pour chaque année  $t$ , la représentation en niveaux de gris de la loi a posteriori empirique associée. En bas, la densité empirique à l'année  $t = 1956$ .

- [5] R. Hilborn and C. Walters. *Quantitative Fisheries Stock Assessment : Choice, Dynamics and Uncertainty*. Chapman and Hall, 1992.
- [6] D.K. Kimura, J.W. Balsiger, and D.H. Ito. Kalman filtering the delay difference equation : practical approaches and simulations. *Fishery Bulletin*, 94(4) :678–691, 1996.
- [7] R. Meyer and R.B. Millar. Bayesian stock assessment using a state-space implementation of the delay difference model. *Canadian Journal of Fisheries and Aquatic Sciences*, 56 :37–52, 1999.
- [8] J. Schnute. A general theory for the analysis of catch and effort data. *Canadian Journal of Fisheries and Aquatic Sciences*, 42 :414–429, 1985.
- [9] H. Tanizaki. Non linear and non-Gaussian state–space modeling with Monte Carlo techniques : A survey and comparative study. In C.R. Rao and D.N. Shanbhag, editors, *Handbook of Statistics 21 : Stochastic Processes : Modeling and Simulation*, pages 871–929. North–Holland, 2003.