MCMC for non linear/non Gaussian state-space models: Application to fishery stock assessment

Fabien Campillo, Rivo Rakotozafy

Rubrique

MCMC for non linear/non gaussian state–space models

Application to fishery stock assessment

Fabien Campillo* — Rivo Rakotozafy**

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

** Faculté des Sciences
Université de Fianarantsoa
301 Fianarantsoa
MADAGASCAR
rrakotoz@syfed.refer.mg

ABSTRACT. We consider a Monte Carlo Markov chain (MCMC) algorithm for fisheries stock assessment. The biomass of this stock at a given year could be modeled as a nonlinear function of the biomass and catch for the two previous years, of different parameters (recruitment, growth rate, natural mortality rate). Given a time series of annual catch and effort data, we would like to achieve the best fitting between the data and a class of non linear/non Gaussian state–space models.

RÉSUMÉ. On applique un algorithme de Monte Carlo par Chaîne de Markov (MCMC) à l'évaluation d'un stock de pêcherie. La biomasse de ce stock à une année donnée peut être modélisée comme une fonction non linéaire des biomasses et des prises aux deux années précédentes, ainsi que de différents paramètres (recrutement, taux de croissance, taux de mortalité naturelle). Etant donnée une série temporelle de captures et d'efforts, on souhaite parvenir au meilleur accord entre ces données et une classe de modèles à espace d'état non linéaire/non gaussien.

KEYWORDS : Bayesian inference, Monte Carlo Markov Chain, Gibbs sampler, Metropolis–Hastings algorithm, nonlinear state-space models

MOTS-CLÉS : Inférence bayésienne, Monte Carlo Chaîne de Markov, échantillonneur de Gibbs, algorithme de Metropolis–Hastings, modèles non linéaire à espace d'état

The purpose of fishery stock assessment is to evaluate the status of a fish stock and to predict how the stock will respond to various exploitation or harvest scenarios. Fish abundance is expressed as the total fish weight (or "biomass"). Increases in the biomass are determined by body growth of individual fish in the population and the addition (or "recruitment") of the new generations of young fishes (or "recruits"). Those gains must then be balanced against the proportion of the population removed by harvesting (called fishing mortality, F) and other losses due to predation, starvation, or disease (called natural mortality, M).

For each year $t = 1 \cdots T$ of a period of T years, the basic data for stock assessments are C_t, the amount of fish caught, and I_t a measure of the abundance of the stock vulnerable to fishing during this year. The index I_t is a relative measure of the abundance of the stock, here we use the *catch per unit of effort* (CPUE) (i.e. ratio of fish caught to the time spent fishing). Using $C_{1:T}$ and $I_{1:T}$ we want to fit a model on these data in order to estimate the biomass B_t for all t, but also to evaluate short–term projections for the stock under a number of different fishing and management scenarios.

Following the work of Meyer and Millar [6], we propose to adopt a state–space representation of a *delay difference* model of the biomass dynamic. The delay difference model was originally introduced by Deriso [2] and generalized by Schnute [7] (for a general presentation see Hilborn and Walters [4]). The indexes $I_{1:T}$ will be considered as the observation process and $C_{1:T}$ as given parameters.

One adjusts the parameters of the model to the data $C_{1:T}$ and $I_{1:T}$ by a MCMC (Monte Carlo–Markov Chain) method, namely a Metropolis–Hastings within Gibbs algorithm. For a general presentation of MCMC methods one can consult Gilks et al [3], for a survey of MCMC methods for non linear/non–Gaussian state–space modeling one can consult Tanizaki [8].

This paper is organized as follows. In the fist section we describe the delay difference state–space model. The second section introduces the proposed MCMC method : the Metropolis–Hasting method within Gibbs sampling.The results are presented in the last section.

1. A state–space model for fishery stock assessment

We are interested in modeling the evolution of the total mass of a fish population vulnerable to fishery (the biomass) along a given series $t = 1 \cdots T$ of years. At the beginning of year t, the biomass B_t is :

$$B_t \stackrel{\triangle}{=} \sum_{a \geq k} w_a \, N_{a,t} \tag{1}$$

where $N_{a,t}$ is the *population size* at age a in year t, w_a is the *average weight* at age a, and k is the *age of recruitment* (i.e. the age from which an individual is entered in the biomass). The first term $R_t = w_k\, N_{k,t}$ in (1) represents the *recruitment* that is the part of the biomass that has just become susceptible to the fishery.

We assume that all recruited fishes are equally vulnerable to natural mortality and to the fishery. The annual evolution of each cohort is described by :

$$N_{a,t} = s_{t-1}\, N_{a-1,t-1}\,, \tag{2}$$

s_{t-1} denotes the *total survival probability* in year $t-1$. The evolution of w_a is given by the following weight–at–age growth curve $w_a = w_\infty\,(1 - e^{-\kappa\,(a-a_0)})$ for $a \geq a_0$, where a_0 is the age for which $w_{a_0} = 0$, and $\kappa > 0$. This gives :

$$w_a = (1+\rho)\, w_{a-1} - \rho\, w_{a-2} \tag{3}$$

where $\rho = e^{-\kappa} \in (0,1)$ is the *growth rate*. One can easily check that (2) and (3) in (1) gives the recurrence :

$$B_t = (1+\rho)\, s_{t-1}\, B_{t-1} - \rho\, s_{t-1}\, s_{t-2}\, B_{t-2} + R_t - \rho\, s_{t-1}\, w\, R_{t-1} \tag{4}$$

with $w = w_{k-1}/w_k \in (0,1)$. Under the assumption of independence of fishing and natural mortality, we have $s_t = s_t^M\, s_t^F$ where s_t^M is the survival rate to natural mortality, it is supposed to be a constant $s^M = e^{-M}$ with $M > 0$, and s_t^F is the survival rate to fishery : $s_t^F = \frac{B_t - C_t}{B_t}$ where C_t is the catch during year t. Then we suppose that the recruitment is constant, i.e. $R = R_t$. Equation (4) becomes :

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

We suppose that the catch is null for $t < 0$, and that B_t is at equilibrium before time $t = 1$, i.e. $B_1 = K$ (the virgin biomass), and :

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

The (deterministic) observation equation is $I_t = q\, B_t$ for $t = 1 \cdots T$, where I_t is a relative biomass index and q is the "catchability" coefficient.

State and observation error processes are modelized as a multiplicative lognormal i.i.d. processes (Meyer and Millar [6] propose a conditional additive Gaussian noises). We end with the following model :

$$B_1 = K \times e^{\sqrt{\sigma_W^2}\, W_1}\,, \tag{5}$$

$$B_2 = F_2(R, B_1) \times e^{\sqrt{\sigma_W^2}\, W_2}\,, \tag{6}$$

$$B_t = F_t(R, B_{t-1}, B_{t-2}) \times e^{\sqrt{\sigma_W^2}\, W_t}\,, \qquad t = 3 \cdots T\,, \tag{7}$$

$$I_t = \left[q\, B_t\right] \times e^{\sqrt{\sigma_V^2}\, V_t}\,, \qquad t = 1 \cdots T\,, \tag{8}$$

$W_{1:T}$ and $V_{1:T}$ are $N(0,1)$ Gaussian noises, and

$$F_2(R, B_1) = (1 + \rho - \rho\, e^{-M})\, e^{-M}\, (B_1 - C_1) + R\left(1 - \rho\, e^{-M}\, w\, \tfrac{B_1 - C_1}{B_1}\right),$$

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

In this application the parameters ρ, M, w are supposed to be given. The parameters K, R, q, σ_W^2, σ_V^2 are supposed to be random variables. These parameters, and the noise processes $W_{1:T}$ and $V_{1:T}$ are supposed to be mutually independent. The *a priori* laws of K, R, q, σ_W^2, σ_V^2 are specified later.

This model is a good tradeoff between surplus production models and age–structured models : from one hand surplus production models are too simplistic (in these models all the components of the evolution of the biomass are aggregated into a unique one), and from the over hand age–structured models are more sophisticated but rely on age–structured catch data which are not available is most applications.

2. MCMC method

The catches C_t are given. The parameters ρ, ω, and M are supposed to be given. In this first work, we also suppose that the variance parameters σ_W^2, σ_V^2 are known, even if it is a restrictive hypothesis. Hence we want to infer K, R, q, $B_{1:T}$ from the observations $I_{1:T}$. For the sake of notational simplicity let :

$$X_{1:P} = [K, R, q, B_{1:T}] \text{ with } P = T + 3\,.$$

Our goal is to sample from the *a posteriori* target density[1] :

$$p(X_{1:P}|I_{1:T})\,. \tag{9}$$

1. For convenience, X will denote a random variable and its realizations, $p(X)$ will denote the density of this *rv* (it will also correspond to either a density or a probability measure), $p(X|Y)$ will denote the

Because sampling from this joint conditional density is not straightforward we propose to use a MCMC algorithm, i.e. to build a Markov chain $\{X_{1:P}^{(k)}\}_{k\geq 0}$ whose ergodic invariant measure admits the target density (9) so that simulating this chain will give samples from the target density.

Require: for each component p: the proposal sampler $\nu_p(\cdot\,|X_{\setminus p}, I_{1:T})$ and the likelihood $\psi_p(X_p, X_{\setminus p}, I_{1:T})$ (cf. (11)).
Ensure: sampling from the target density $p(X_{1:P}|I_{1:T})$ (cf. (9))

$k \leftarrow 0$
choose initial guess $X_{1:P}$
loop
  for $p = 1, \ldots, P$ **do**
    $X'_p \leftarrow \nu_p(\cdot\,|X_{\setminus p})$ {sample the new candidate}
    $\alpha = \psi_p(X'_p|X_{\setminus p}, I_{1:T})/\psi_p(X_p|X_{\setminus p}, I_{1:T})$ {cf. (12)}
    $U \sim U[0,1]$
    if $U \leq \alpha$ **then**
      $X_p \leftarrow X'_p$
    end if
  end for
  $k \leftarrow k+1$
  if $k \geq k_{\text{burn–in}}$ **then**
    return $X_{1:P}$ {after the burn–in period}
  end if
end loop

Algorithm 1: Metropolis–Hastings within Gibbs sampler

We apply the Gibbs sampler: we iteratively sample from each of the marginal conditional densities of X_p given other components $X_{\{1,\ldots,P\}\setminus\{p\}}$ (and $I_{1:T}$) i.e. from:

$$p(X_p|X_{\setminus p}, I_{1:T}) \text{ with } \setminus p \stackrel{\triangle}{=} \{1,\ldots,P\} \setminus \{p\} \tag{10}$$

for $p = 1, \ldots, P$. Even if these densities admit explicit analytical representations, we cannot sample directly from them. For this reason, we use a Metropolis–Hasting algorithm within the Gibbs sampling loop. Each of the previous densities will be decomposed as the product of a density we know how to sample from, and of a likelihood that we can compute explicitly (up to a multiplicative constant):

$$p(X_p|X_{\setminus p}, I_{1:T}) \propto \underbrace{\nu_p(X_p|X_{\setminus p}, I_{1:T})}_{\text{proposal density}} \times \underbrace{\psi_p(X_p, X_{\setminus p}, I_{1:T})}_{\text{likelihood}} . \tag{11}$$

conditional density law of X given Y (we should write $x \mapsto p_{X|Y}(x|y)$ for the conditional density of X given $Y = y$). The meaning will be clear from the context.

Starting from a given configuration $X_{1:P}^{(k)}$ we sample the next configuration $X_{1:P}^{(k+1)}$ as follow: let $X_{1:P}^{(k+1)} \leftarrow X_{1:P}^{(k)}$ then for each component $p \in \{1:P\}$ we sample a new pth configuration

$$X'_p \sim \nu_p(\cdot | X_{\setminus p}^{(k+1)}, I_{1:T}),$$

then with probability $1 \wedge \alpha$ where:

$$\alpha \stackrel{\triangle}{=} \frac{\psi_p(X'_p, X_{\setminus p}^{(k+1)}, I_{1:T})}{\psi_p(X_p^{(k)}, X_{\setminus p}^{(k+1)}, I_{1:T})}, \tag{12}$$

we accept X'_p, i.e. $X_p^{(k+1)} = X'_p$ (with probability $1 - 1 \wedge \alpha$ the pth component is not changed). All the components are updated in the same way. After a certain period (burn-in) the Gibbs sampler reaches its stationary behavior and produce samples from the posterior density. The detailed procedure is summarized in Algorithm 1. The marginal conditional densities and their decomposition between proposal densities and likelihood functions are partly detailed in the appendix.

3. Results and comments

We performed this algorithm on simulated data which match the data used in Meyer and Millar [6] (themselves taken from Kimura et al [5]). The stationarity of the sampler is not reached on the component q in less than millions of iterations, so we chose to froze this parameter to its true value (i.e. supposed that it is known). Then 250000 iterations of the Gibbs sampler were performed (with a burn-in period of 1000 iterations).

For the simulation and the MCMC algorithm we used the following values: $\rho = 1$, $\omega = 0$, $M = 0.6$, $\sigma_W^2 = \sigma_V^2 = 0.01$ (the parameter q is also supposed to be known $q = 9.39$). Moreover for the simulation we use the values $R = 218.5$, and $K = R\,\frac{1-\rho\,e^{-M}\,\omega}{1-(1+\rho)\,e^{-M}+\rho\,e^{-2M}} \simeq 1073.3$ (which corresponds to the virgin biomass, i.e. the asymptotic solution of (4) without catch), the biomass time series $B_{t=1934\cdots1967}$ and the CPUE one $I_{t=1934\cdots1967}$ are simulated through the system (5–8).

As illustrated by Fig. 1, the match between the parameters used in the simulation and the a posteriori law inferred by the Gibbs sampler (cf. Figure 1) is fairly good. The correspondence between the "true" (simulated) biomass time series and the corresponding a posteriori law derived from the Gibbs sampler is also good.

The main drawback of this approach is that it is computationally too slow and cumbersome, so it seams necessary to use more elaborated MCMC techniques like interacting Metropolis procedures (Del Moral and Doucet [1]). Moreover, it is necessary to compare this approach to a non Bayesian one (e.g. to approximate the maximum likelihood estimator with an EM algorithm).

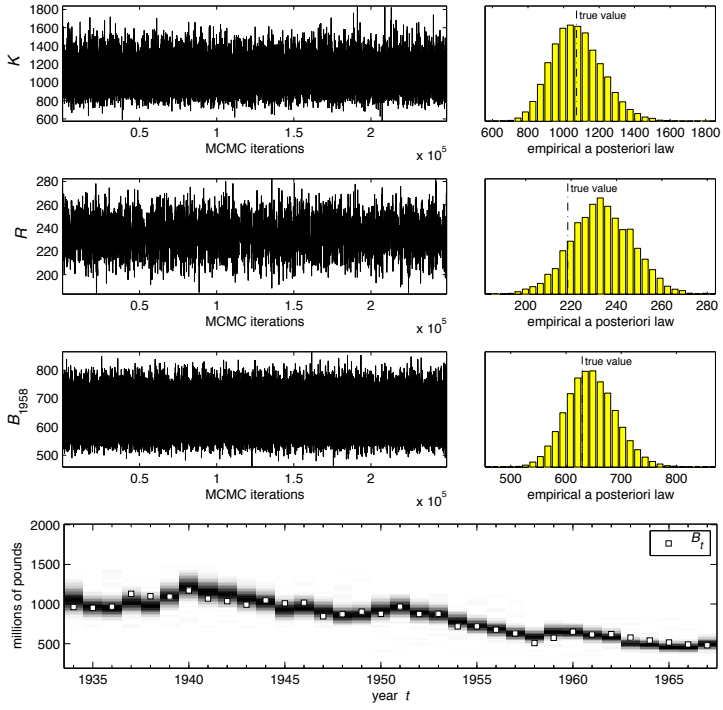

Figure 1. *For the components K, R, B_{1958}, the iterations of the MCMC procedure are displayed on the left and the resulting empirical a posteriori laws (together with the true values of the parameters) on the right. Below is displayed the biomass time series B_t with, for each year t, the corresponding empirical a posteriori law represented in grey levels.*

Acknowledgements

M. Rakotozafy's work has been partially supported by the French Ministry of Foreign Affairs (French embassy in Madagascar), part of this work was carried out during a stay at IRISA.

4. References

P. Del Moral and A. Doucet. On a class of genealogical and interacting metropolis models. In J.-M. Morel, F. Takens, and B. Teissier, editors, *Séminaire de Probabilités XXXVII*, volume 1832 of *Lecture Notes in Mathematics*, pages 415 – 446. Springer–Verlag, 2003.

R. B. Deriso. Harvesting strategies and parameter estimation for an age-structured model. *Canadian Journal of Fisheries and Aquatic Sciences*, 37:268–282, 1980.

W. Gilks, S. Richardson, and D. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*. Interdisciplinary Statistics. Chapman and Hall, 1996.

R. Hilborn and C. Walters. *Quantitative Fisheries Stock Assessment : Choice, Dynamics and Uncertainty*. Chapman and Hall, 1992.

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.

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.

J. Schnute. A general theory for the analysis of catch and effort data. *Canadian Journal of Fisheries and Aquatic Sciences*, 42:414–429, 1985.

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.

Appendix : the marginal conditional densities

The marginal conditional densities (10) are :

$$
\begin{array}{ll}
① & p(K|R,q,B_{1:T},I_{1:T}) \\
② & p(R|K,q,B_{1:T},I_{1:T}) \\
③ & p(q|K,R,B_{1:T},I_{1:T}) \\
④_t & p(B_t|K,R,q,B_{\setminus t},I_{1:T})\ ,\ t=1\cdots T
\end{array}
$$

For each of these marginal densities we specify the decomposition (11) between the proposal density and the likelihood function :

$$
\begin{array}{lcll}
① & \propto & \pi(K) & \times\ p(B_1|K) \\
② & \propto & \pi(R) & \times\ p(B_2|R,B_1)\ \prod_{t=3}^{T} p(B_t|R,B_{t-1},B_{t-2}) \\
③ & \propto & \pi(q) & \times\ \prod_{t=1}^{T} p(I_t|q,B_t) \\
④_1 & \propto & p(B_1|K) & \times\ p(B_2|R,B_1)\,p(B_3|R,B_2,B_1)\,p(I_1|q,B_1) \\
④_2 & \propto & p(B_2|R,B_1) & \times\ p(B_3|R,B_2,B_1)\,p(B_4|R,B_3,B_2)\,p(I_2|q,B_2) \\
④_t & \propto & p(B_t|R,B_{t-1},B_{t-2}) & \times\ p(B_{t+1}|R,B_t,B_{t-1})\,p(B_{t+2}|R,B_{t+1},B_t) \\
 & & & \quad p(I_t|q,B_t) \\
④_{T-1} & \propto & p(B_{T-1}|R,B_{T-2},B_{T-3}) & \times\ p(B_T|R,B_{T-1},B_{T-2})\,p(I_{T-1}|q,B_{T-1}) \\
④_T & \propto & p(B_T|R,B_{T-1},B_{T-2}) & \times\ p(I_T|q,B_T) \\
 & & \underbrace{\qquad\qquad\qquad} & \underbrace{\qquad\qquad\qquad\qquad\qquad\qquad} \\
 & & \text{proposal density} & \text{likelihood}
\end{array}
$$

and the prior densities on the parameters (whose choice will be detailed elsewhere) are : $\pi(K) \sim U[400, 4000], \pi(R) \sim U[0, 2000], \pi(q) \sim U[0, 100]$ (uniform laws).