

2007 International Conference in Honor of Claude Lobry

## Computational probability modeling and Bayesian inference

Fabien Campillo<sup>1\*</sup> — Rivo Rakotozafy<sup>2</sup> — Vivien Rossi<sup>3</sup>

<sup>1</sup> INRIA/INRA  
UMR ASB - 2 place Viala, 34060 Montpellier cedex 06, France  
Fabien.Campillo@inria.fr

<sup>2</sup> Université de Fianarantsoa  
BP 1264, Andrainjato, 301 Fianarantsoa, Madagascar  
rrakotozafy@uni-fianar.mg

<sup>3</sup> CIRAD  
Campus International de Baillarguet, 34398 Montpellier cedex 5, France  
vivien.rossi@cirad.fr

\* Corresponding author



**RÉSUMÉ.** La modélisation probabiliste et l'inférence bayésienne computationnelles rencontrent un très grand succès depuis une quinzaine d'années grâce au développement des méthodes de Monte Carlo et aux performances toujours croissantes des moyens de calcul. Au travers d'outils comme les méthodes de Monte Carlo par chaîne de Markov et les méthodes de Monte Carlo séquentielles, l'inférence bayésienne se combine efficacement à la modélisation markovienne. Cette approche est également très répandue dans le domaine de l'écologie et l'agronomie. Nous faisons le point sur les développements de cette approche appliquée à quelques exemples de gestion de ressources naturelles.

**ABSTRACT.** Computational probabilistic modeling and Bayesian inference has met a great success over the past fifteen years through the development of Monte Carlo methods and the ever increasing performance of computers. Through methods such as Monte Carlo Markov chain and sequential Monte Carlo Bayesian inference effectively combines with Markovian modelling. This approach has been very successful in ecology and agronomy. We analyze the development of this approach applied to a few examples of natural resources management.

**MOTS-CLÉS :** Modélisation computationnelle markovienne, inférence bayésienne computationnelle, modélisation bayésienne hiérarchique, méthode de Monte Carlo par chaîne de Markov, méthode de Monte Carlo séquentielle, écologie numérique

**KEYWORDS :** Computational Markovian modeling, computational Bayesian inference, hierarchical Bayesian modeling, Monte Carlo Markov chain, sequential Monte Carlo, computational ecology



---

## 1. Introduction

The past fifteen years have seen considerable activity in Bayesian inference in many areas, including ecology and agronomy [36, 39, 37, 16, 10, 11, 19, 23]. In the article [10], entitled “Why environmental scientists are becoming Bayesians”, James S. Clark tries to explain this trend. The boom can also be seen in areas other than ecology [30, 2].

In statistical inference, Bayesian and frequentist approaches have given rise to much debate, notably in the 1930s. These discussions, sometimes intense, can not be reduced to a simple opposition between a “frequentist clan” embodied by Ronald Fisher and a “Bayesian clan” embodied by Harold Jeffreys [17, 27]. Our goal is not to address the epistemological relevance of these approaches but rather to consider these approaches from a practical point of view. It is important to note that the success of one approach over the other is also due to its ability to solve practical problems. Indeed, more than the criticism of subjectivity made by supporters of the frequentist approach about the Bayesian approach, it is the impossibility of applying Bayesian methods to non-academic problems which has promoted the development of frequentist methods to the detriment of Bayesian methods. Over the past fifteen years this impossibility has steadily declined, on the one hand with the appearance of new methods, on the other hand thanks to the development of computers. A computational approach allows us to propose effective approximations techniques in the Bayesian framework. In many areas the situation is now reversed : Bayesian methods offer numerical tools that are more affordable than those of frequentist methods.

These approximation techniques are based on Monte Carlo simulations. They can be classified into two categories : *Monte Carlo Markov chains methods* (MCMC) [22] and *sequential Monte Carlo methods* (SMC) [15]. The development of toolboxes such as BUGS and WinBUGS contributes greatly to the dissemination and success of MCMC methods. Static problems (non-temporal) are generally treated with MCMC methods and dynamic problems (temporal) wh SMC methods.

For these Monte Carlo methods, most often it is the process that underlies the phenomenon in question which is simulated. In other words, it is necessary to model the phenomenon studied before processing the data. This approach is consistent with the Bayesian approach, which requires some a priori knowledge. For this purpose and in addition to the MCMC and SMC methods, a modeling methodology called *hierarchical Bayesian modeling* has been developed for a decade [1]. These models are not specifically new, they are Markovian with a hierarchical structure, but first, they allow for the use of the numerical methods mentioned above, and second, they are particularly suited to applied modeling and engineering.

Ecology and agricultural areas are among the main fields of application of these approaches [4]. These areas often involve dynamical systems where the frequency of observations is low and where time series are short, typically one observation per year over a few tens of years. For this reason the application of frequentists methods, based on large samples, are difficult in this area.

That is one reason that makes frequentists methods difficult to apply in this area.

In this area we can either make use of SMC methods as well as MCMC methods. Indeed, for dynamical models where the frequency of observations is around a year, there is no real-time constraint and it possible to use both non-sequential (batch) and sequential

techniques [8, 9, 20]. Non-sequential approaches have been used for example in fisheries [35, 38]. In areas where the real-time constraint is higher (robotics, target tracking, image processing, speech processing etc.) MCMC methods presented in this article are not feasible.

Moreover, ecology poses specific modeling problems, including the consideration of phenomena at different scales. The hierarchical approaches are therefore natural [41], which explains the success of hierarchical Bayes models in this area [12]. In this article we limit ourselves to the problem of estimation.

In Section 2 we set the general framework, we present problems static and dynamic time cases. In Section 3 we present MCMC and SMC methods, with several examples. We explain the specifics of each approach and how they adapt to the application field considered here.

---

## 2. Hierarchical Bayes models

We focus here on the estimation problem : observations are given and the goal is to “determine” unknown parameters and state from these observations. Modeling will make the link between observations and unknown states and parameters. We consider the context of continuous space (e.g.  $\mathbb{R}^n$ ); finite and countable cases require specific treatments. We place ourselves in the Bayesian context where the unknown parameters will be treated as random variables.

We will use a notational convention which is widely accepted in application. It is not mathematically rigorous but it is often more descriptive and intuitive. If  $Y$  denotes the observation, its probability density function (pdf) will be denoted by  $p(Y)$  (and all probability distributions are supposed to admit densities), hence :

$$\mathbb{P}(Y \in B) = \int_B p(Y) dY, \quad \mathbb{E}\phi(Y) = \int \phi(Y) p(Y) dY.$$

This notation cannot be used in a mathematical framework, it generates inaccuracies (it is indeed difficult to define a function from its argument). If  $\theta$  is an unknown parameter, the joint pdf of  $(\theta, Y)$ , is denoted by  $p(\theta, Y)$ . The conditional pdf of  $\theta$  given  $Y$  is denoted by  $p(\theta|Y)$ . From the definition of the conditional pdf, we get :

$$p(\theta|Y) \stackrel{\text{def}}{=} \frac{p(\theta, Y)}{p(Y)} = \frac{p(\theta, Y)}{\int p(\theta, Y) d\theta}.$$

From the observation of  $Y$ , we want to estimate the unknown parameter  $\theta$ . The Bayesian approach is to calculate the *posterior pdf*  $\theta \mapsto p(\theta|Y)$  to obtain estimators like the mean :

$$\hat{\theta} \stackrel{\text{def}}{=} \int \theta p(\theta|Y) d\theta \tag{1}$$

In fact,  $p(\theta|Y)$  represents all information on  $\theta$  contained in the observation  $Y$ , in a priori knowledge on  $\theta$  and in the model. Indeed, the Bayes formula writes :

$$p(\theta|Y) = \frac{p(Y|\theta) p(\theta)}{p(Y)} = \frac{p(Y|\theta) p(\theta)}{\int p(Y|\theta) p(\theta) d\theta}$$

which is usually presented as :

$$p(\theta|Y) \propto p(Y|\theta) p(\theta) \quad (2)$$

that is, for fixed  $Y$  (i.e. as a function of  $\theta$  only),  $p(\theta|Y)$  is proportional to the product of the prior pdf  $p(\theta)$  by the likelihood function  $\theta \mapsto p(Y|\theta)$ .  $p(\theta)$  is the a priori knowledge on  $\theta$  and  $p(Y|\theta)$  is the observation model.

The Bayes formula may appear as a mere tautology and one may wonder how in (2), the right-hand term — the product of the prior pdf by the likelihood function — is easier to determine than the left-hand term — the posterior pdf. In fact the prior pdf  $p(\theta)$  is given and the likelihood function can often be expressed in a natural way, it represents the density of observation when the value of the unknown parameter is fixed.

**Example 2.1** Consider the following observation model :

$$Y = h(\theta) + V$$

where  $V \sim \mathcal{N}(0, 1)$  and where  $\theta$  and  $V$  are independants (denoted  $\theta \perp V$ ). As  $p(Y|\theta) \propto \exp[-\frac{1}{2}(Y - h(\theta))^2]$  (suppose  $p(\theta)$  given), one can easily deduce the posterior pdf  $p(\theta|Y)$ .

An important aspect of the Bayes formula is its sequential nature :

$$\text{posterior} \leftarrow \text{likelihood} \times \text{prior} .$$

It allows us to take data into account as it becomes available and also to integrating heterogeneous data.

## 2.1. The static case

We consider a mortality model for a population of  $N$  trees indexed by  $i$ . The variable  $Y_i$  corresponds to the observed state of the tree  $i$  at the measurement campaign :  $Y_i = 1$  or 0 depending on whether the tree is dead or alive. To study the factors favoring mortality we usually use a logistic model, which in the Bayesian framework is expressed in a hierarchical structure.

The state of the tree is binomial random variable :

$$Y_i \sim \text{Ber}(p_i). \quad (3)$$

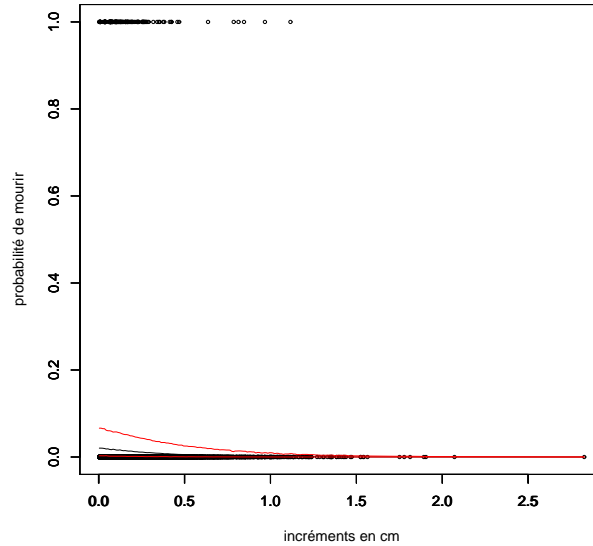
with parameter  $p_i$ . We suppose this parameter connected to covariates  $C_i = (\delta_i, \Delta_i)$ . For each individual  $i$ ,  $\delta_i$  is the population density within a 30 meter perimeter and  $\Delta_i$  is the diameter increase between the last two campaigns measures. It is assumed that  $p_i$  is related to  $C_i$  through a logit function :

$$\log \frac{p_i}{1 - p_i} = \theta_0 + \theta_1 \delta_i + \theta_2 \Delta_i .$$

One can express  $p_i$  as a sigmoid function of the covariates  $C_i$  :

$$p_i = f(\theta, C_i) \stackrel{\text{def}}{=} \frac{1}{1 + \exp\{-\theta_1 + \theta_2 \delta_i + \theta_3 \Delta_i\}}$$

with  $\theta = (\theta_1, \theta_2, \theta_3)$ .



**Figure 1.** Estimation of the mortality probability by a maximum likelihood approach from model (4). This estimate is plunged into the model (4), then by sampling the model one can calculate a confidence interval for each value of  $\Delta$ . This interval is represented by 3 curves (average in the center and 95 % bounds above and below). The measurements are represented by points. Although the death increases significantly whenever the trees have grown little or not at all, the model does not account for the variability of the observations. This problem would also appear with a Bayesian estimator.

Assuming that trees are mutually independent, the likelihood function associated with this model is :

$$p(Y|C, \theta) = \prod_{i=1}^N p(Y_i|\theta, C_i) \quad (4a)$$

with

$$p(Y_i|\theta, C_i) = \text{Ber}(Y_i|f(\theta, C_i)) \stackrel{\text{def}}{=} f(\theta, C_i)^{Y_i} (1 - f(\theta, C_i))^{1-Y_i}. \quad (4b)$$

One can compute the maximum likelihood estimate of  $\theta$  and plunge this estimated value in this model : from Monte Carlo samples one can deduce empirical confidence intervals. We see that the value of the parameters does not account for the variability of observations (Figure 1). Although the mortality increases whenever the trees have grown little or not at all, the model does not account for the variability of the measurements. This problem would have also raised with Bayesian estimators.

It is indeed necessary to improve the model. As in natural forests several tree species coexist, it is realistic to assume that the values of the parameter  $\theta$  vary depending on the species. It is admitted that the mortality of trees depends on their shade tolerance. To simplify we consider two groups : a shade-tolerant group “s” and a heliophilous group (“h”). It is assumed that mortality parameters differ depending on whether the tree is shade-tolerant or not. The model thus becomes a mixture  $Y_i \sim \rho_i \text{Ber}(p_i^h) + (1 - \rho_i) \text{Ber}(p_i^s)$

where  $p_i^s$  and  $p_i^h$  are expressed as sigmoid functions of the covariates  $C_i$   $p_i^\Delta = f(\theta^\Delta, C_i)$  for  $\Delta = "s"$  or  $"h"$ ,  $\theta^h$  and  $\theta^s$  are the parameters of the logistic model for heliophilous species and shade-tolerant species respectively. Thus the pdf of observations  $Y$  is :

$$p(Y_i|C_i, \rho_i, \theta) = \rho_i \text{Ber}(Y_i|f(\theta^h, C_i)) + (1 - \rho_i) \text{Ber}(Y_i|f(\theta^s, C_i)). \quad (5a)$$

For the mixture parameter  $\rho_i$ , we choose the following prior pdf :

$$p(\rho_i) = U[0, 1] \quad (5b)$$

which is also a beta distribution with parameters  $(1, 1)$ .

As each of these two groups contains several species, we will therefore assume that these parameters  $\theta_j^h$  and  $\theta_j^s$  are random. It is assumed that for  $j = 1, 2, 3$  and  $\Delta = "s", "h"$ , the parameters  $\theta_j^\Delta$  are independent and, for simplicity, normal :

$$p(\theta_j^\Delta | \mu_j^\Delta, \sigma_j^{\Delta,2}) = \mathcal{N}(\theta_j^\Delta | \mu_j^\Delta, \sigma_j^{\Delta,2}). \quad (5c)$$

It remains now to define the prior pdf's for the hyper-parameters of the hierarchical model. For parameters of variance, the usual approach is to choose Inverse-Gamma distributions because they are conjugate to the normal distribution :

$$p(\sigma_j^{\Delta,2}) = \text{InvGamma}(\alpha, \beta), \quad p(\mu_j^\Delta) = \mathcal{N}(0, \gamma) \quad (5d)$$

The values of  $\alpha$ ,  $\beta$  and  $\gamma$  are chosen to obtain noninformative associated priors.

The model can be represented graphically as a directed acyclic graph (DAG) (see Figure 2).

The pdf of the model is :

$$p(Y, C, \rho, \theta, \mu, \sigma^2) \propto p(Y, C, \rho | \theta, \mu, \sigma^2) p(\theta, \mu, \sigma^2)$$

One the one hand :

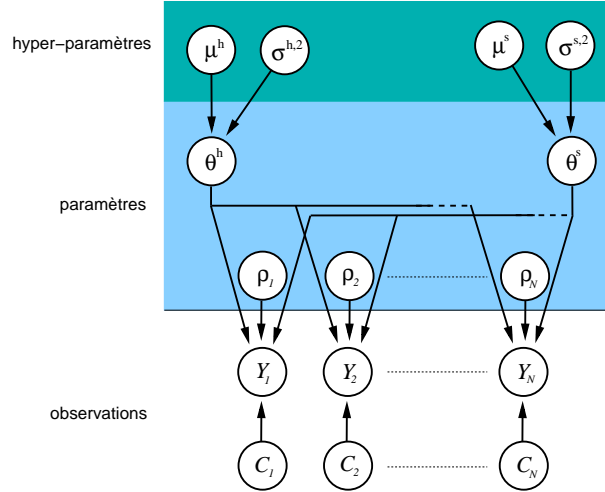
$$\begin{aligned} p(\theta, \mu, \sigma^2) &= \prod_{\Delta=s,h} p(\theta^\Delta, \mu^\Delta, \sigma^{\Delta,2}) = \prod_{\Delta=s,h} \prod_{j=1}^3 p(\theta_j^\Delta, \mu_j^\Delta, \sigma_j^{\Delta,2}) \\ &= \prod_{\Delta=s,h} \prod_{j=1}^3 [p(\theta_j^\Delta | \mu_j^\Delta, \sigma_j^{\Delta,2}) p(\mu_j^\Delta) p(\sigma_j^{\Delta,2})] \end{aligned}$$

and secondly

$$p(Y, C, \rho | \theta, \mu, \sigma^2) = \prod_{i=1}^N p(Y_i, C_i, \rho_i | \theta, \mu, \sigma^2) = \prod_{i=1}^N [p(Y_i | C_i, \rho_i, \theta) p(C_i) p(\rho_i)],$$

finally the pdf of the model reads :

$$\begin{aligned} p(Y, C, \rho, \theta, \mu, \sigma^2) &= \prod_{i=1}^N [p(Y_i | \theta, \mu, \sigma^2) p(C_i) p(\rho_i)] \\ &\quad \times \prod_{\Delta=s,h} \prod_{j=1}^3 [p(\theta_j^\Delta | \mu_j^\Delta, \sigma_j^{\Delta,2}) p(\mu_j^\Delta) p(\sigma_j^{\Delta,2})] \quad (6) \end{aligned}$$



**Figure 2.** Graphical representation of the model (6) in the form of a directed acyclic graph (DAG). This diagram represents the dependencies of each variable on others. For example we can see that  $Y_1$  depends on others variables only through  $C_1$ ,  $\rho_1$ ,  $\theta^h$  and  $\theta^s$ ; we can also see that conditionally on  $Y_2$ ,  $C_2$  is independent of all other variables etc.

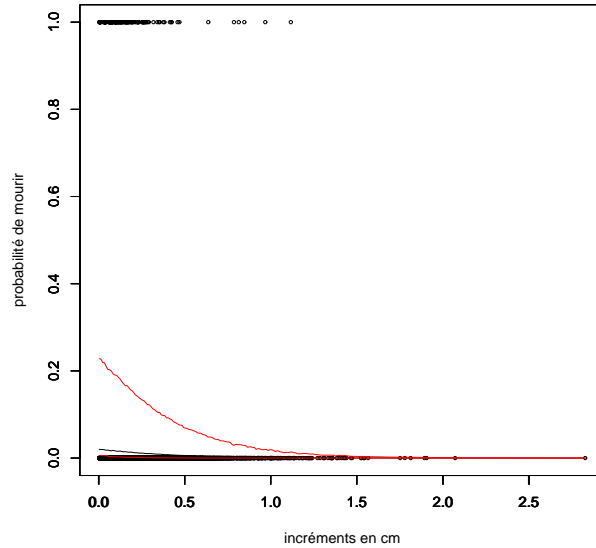
where different terms are given by the equations (5) (except that of  $p(C_i)$  which is irrelevant here). Thus the posterior pdf of the parameters is :

$$p(\rho, \theta, \mu, \sigma^2 | Y, C) \propto \prod_{i=1}^N [p(Y_i | \theta, \mu, \sigma^2) p(\rho_i)] \\ \times \prod_{\Delta=s,h} \prod_{j=1}^3 [p(\theta_j^\Delta | \mu_j^\Delta, \sigma_j^{\Delta,2}) p(\mu_j^\Delta) p(\sigma_j^{\Delta,2})] \quad (7)$$

Although explicit, this expression is not very useful because it is not possible to integrate it in order to calculate estimators like (1). However, the hierarchical nature of the model, represented graphically in Figure 2, will be effectively used by the methods presented in the next section.

The inference of this model better accounts for the variability of measurements particularly when the diameter increases between field measurements have been low (see Figure 3).

It would be interesting to perform such an analysis without specifying in advance the number of groups. Such a framework would be much more relevant, indeed for a given site ecologists are not always able to provide a number of groups of species. This approach is more difficult because the number of unknown parameters could change depending on the number of groups. The flexibility of Bayesian hierarchical framework can make this approach feasible in particular through reversible jump Markov chains algorithms [25].



**Figure 3.** Estimation of the mortality probability by maximum likelihood estimation from model (6). This example of hierarchical Bayes model (see Figure 2) is more elaborate than the first model (4). It gives a better account of the measurements. Indeed, one can estimate the hyper-parameters of this model and plunge the values obtained in the model (6), which can then be simulated to determine confidence intervals for each value increment  $\Delta$ . This confidence interval is represented by 3 curves (mean in the center and 95% bounds above and below). The observations are represented by the dots. Compared to the Figure 1, this model better accounts for the variability of the measurements.

## 2.2. The dynamic case

We consider the discrete-time case :  $t = 1, 2, 3, \dots$ . At time  $t$  we have the following measurements :

$$Y_{1:t} = (Y_1, Y_2, \dots, Y_t)$$

we want to estimate the following hidden state variables :

$$X_{1:t} = (X_1, X_2, \dots, X_t)$$

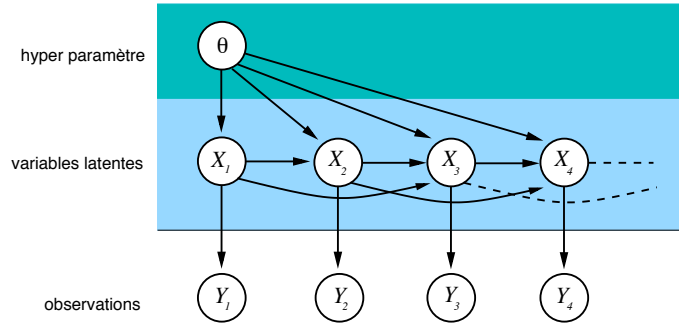
and possibly to estimate unknown parameters  $\theta$ .

We first present some examples of such models.

**Example 2.2 (Fishery, Deriso–Schnute model)** We are interested in modeling the evolution of the total mass  $X_t$  of a fish population vulnerable to fishery (the biomass) along a given series  $t = 1, \dots, T$  of years. This biomass evolves according to a delay difference model :

$$X_t = \left[ (1 + \rho) e^{-M} \frac{X_{t-1} - C_{t-1}}{X_{t-1}} X_{t-1} - \rho e^{-2M} \frac{X_{t-1} - C_{t-1}}{X_{t-1}} \frac{X_{t-2} - C_{t-2}}{X_{t-2}} X_{t-2} + R \left( 1 - \rho e^{-M} \omega \frac{X_{t-1} - C_{t-1}}{X_{t-1}} \right) \right] \times e^{wt} \quad (8a)$$





**Figure 4.** Graphical representation associated with Equations (8) of the Example 2.2 of the Deriso–Schnute model. It is an hidden Markov model of order 2.

where  $C_t$  denotes the catch during year  $t$ ,  $\rho$  the growth rate (that we assume is estimated elsewhere),  $K = X_{-1} = X_0$  the virgin biomass,  $R$  the recruitment (constant).

The measurement process is of the form :

$$Y_t = q X_t e^{v_t} \tag{8b}$$

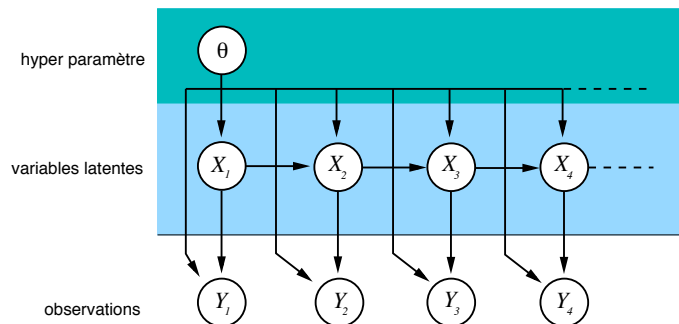
it is a relative biomass index,  $q$  is a catchability coefficient (see [7] for details).

Here  $w_t$  and  $v_t$  are independent white Gaussian noise processes with variance  $\sigma_w^2$  and  $\sigma_v^2$  respectively. This model is Markovian of order 2.

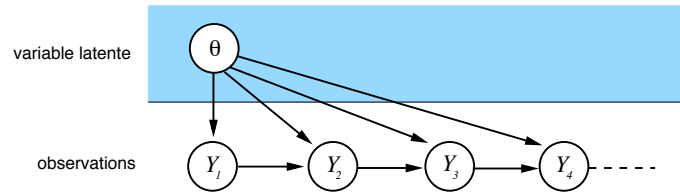
The unknown parameter is  $\theta = (K, R, q, \sigma_w^2, \sigma_v^2)$ . This model is illustrated in Figure 4.

**Example 2.3 (Fishery, Ricker model)** We suppose that the biomass  $X_t$  available in a fishery at year  $t$  evolves according to a Ricker growth model :

$$X_{t+1} = (X_t - C_t) e^{a-b(X_t-C_t)} e^{w_t} \tag{9a}$$



**Figure 5.** Graphical representation associated with Equations (9) of Example 2.3 of the Ricker model. It is a first order hidden Markov model.



**Figure 6.** Graphical representation associated with Example 2.4 of model for forest dynamics. It is a first order Markov model with a hidden parameter.

where  $a$  is the growth rate of the population,  $a/b$  the carrying capacity,  $C_t$  the catch at year  $t$ . Like in the previous example, the measurement process is :

$$Y_t = q X_t e^{v_t} \quad (9b)$$

where  $q$  is a catchability coefficient. Here  $w_t$  et  $v_t$  are independent white Gaussian noise processes with variance  $\sigma_w^2$  and  $\sigma_v^2$  respectively. The unknown parameter is  $\theta = (r, b, q, \sigma_w^2, \sigma_v^2)$ . This model is illustrated in Figure 5.

**Example 2.4 (A forest dynamic model)** In this example the dynamic of the forest is directly observed (i.e. no hidden process  $X_t$ ). The observation state vector :

$$Y_t = \begin{pmatrix} Y_t^1 \\ Y_t^2 \\ Y_t^3 \\ Y_t^4 \end{pmatrix} \begin{array}{l} \# \text{ individuals } 10\text{cm} < \varnothing < 20\text{cm} \\ \# \text{ individuals } 20\text{cm} < \varnothing < 30\text{cm} \\ \# \text{ individuals } 30\text{cm} < \varnothing < 40\text{cm} \\ \# \text{ individuals } 40\text{cm} > \varnothing \end{array} \quad (10)$$

is the size of population in 4 diameter size classes (10-20 cm, 20-30 cm, 30-40 cm, >40 cm diameter at breast height). It is possible to write a detailed dynamic for this system as a first order Markov process with transition kernel  $p(Y_{t+1}|\theta, Y_t)$ . It is also possible to consider a Gaussian approximation of such a kernel [18]. This transition operator will depend on a parameter  $\theta$  that needs to be estimated. This type of model corresponds to the Figure 6.

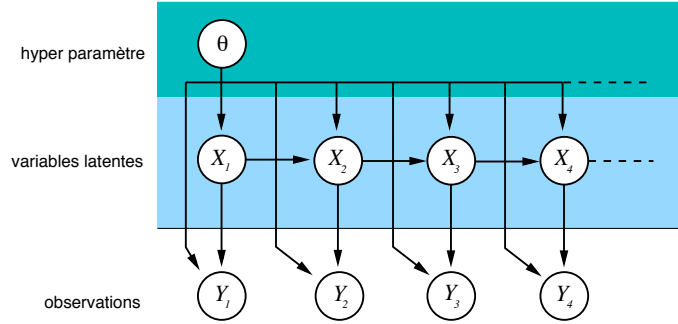
In all these examples we are trying to determine the joint conditional pdf of the parameters  $\theta$  and of the hidden states  $X_t$  given the measurement  $Y_t$ . For this we must exploit certain features of the examples. Indeed, the pdf of the model can be factorized in the form :

$$p(Y_{1:T}, X_{1:T}, \theta) = p(Y_{1:T}|X_{1:T}, \theta) p(X_{1:T}|\theta) p(\theta) \quad (11)$$

but this expression is too broad to be exploited in practice. It is necessary to particularize the model. The first assumption is that, conditionally on the parameter  $\theta$ , the hidden process  $X_t$  is Markovien, i.e. :

$$p(X_{1:T}|\theta) = \prod_{t=1}^T p(X_t|X_{1:t-1}, \theta)$$

where by convention  $p(X_1|X_0, \theta) = p(X_1|\theta)$ .



**Figure 7.** Graphical representation of the hidden Markov model (12) with an unknown parameter.

The second hypothesis is that, conditionally on  $\theta$  and  $X_{1:T}$ , the observations  $Y_t$  are independent, and that  $Y_t$  depends on  $(\theta, X_{1:T})$  only through  $(\theta, X_t)$ . This hypothesis, usually referred to as the memoryless channel in signal processing, reads :

$$p(Y_{1:T}|X_{1:T}, \theta) = \prod_{t=1}^T p(Y_t|X_t, \theta)$$

In conclusion, we have supposed that the pdf (11) of the model is of the form :

$$p(Y_{1:T}, X_{1:T}, \theta) = p(\theta) \prod_{t=1}^T [p(Y_t|X_t, \theta) p(X_t|X_{t-1}, \theta)]. \quad (12)$$

These models are usually referred to as *hidden Markov models* (with parameters), their basic components are :

- $p(\theta)$  priori density of the parameter,
- $p(X_1|\theta)$  initial pdf of the Markov chain  $X_t$ ,
- $p(X_t|X_{t-1}, \theta)$  transition pdf of the Markov chain  $X_t$ ,
- $p(Y_t|X_t, \theta)$  emission pdf.

These models are in fact equivalent to state-space models of the form :

$$X_t = f(\theta, X_{t-1}, w_t) \quad (13a)$$

$$Y_t = h(\theta, X_t, v_t) \quad (13b)$$

where  $w_t$  and  $v_t$  are Gaussian white noise (i.i.d. variables with zero mean);  $w_t, v_t, X_1, \theta$  are independents (see Figure 7).

The sequential processing of this problem is to treat the measurement data  $Y_t$  one after another in chronological order. In non-sequential methods (batch methods), data are processed globally. It is necessary to use sequential methods when the real-time constraints are strong or when there is a lot of data to be processed (data mining). Sequential methods are not necessary when working on a finished horizon or when there is a lot of time between two observations. This is precisely the case for applications of interest to us here, we can therefore appeal equally to sequential or non-sequential methods.

### Sequential approach

The sequential approach most often uses filtering, which determines the conditional pdf at time  $t$  sequentially from the conditional pdf at time  $t - 1$  and the new measurement  $Y_t$  :

$$p(X_{t-1}, \theta | Y_{1:t-1}) \xrightarrow{Y_t \downarrow} p(X_t, \theta | Y_{1:t})$$

For the sake of simplicity suppose that the parameter  $\theta$  is known and the focus is on the identification of the hidden process  $X_t$ . It is possible to determine the conditional pdf  $p(X_t | Y_{1:t})$  recursively with the help of the *sequential Bayes formula*. It consists in coupling the a time iteration of the Chapman-Kolmogorov equation with the Bayes formula. The iteration  $p(X_{t-1} | Y_{1:t-1}) \rightarrow p(X_t | Y_{1:t})$  is achieved in two classic steps :

– Prediction (Chapman-Kolmogorov) :

$$p(X_t | Y_{1:t-1}) = \int p(X_t | X_{t-1}) p(X_{t-1} | Y_{1:t-1}) dX_{t-1}. \quad (14a)$$

– Correction (Bayes) :

$$p(X_t | Y_{1:t}) = \frac{p(Y_t | X_t) p(X_t | Y_{1:t-1})}{\int p(Y_t | X_t) p(X_t | Y_{1:t-1}) dX_t} \propto p(Y_t | X_t) p(X_t | Y_{1:t-1}). \quad (14b)$$

These equations can be explicitly solved in the linear/Gaussian case (and in a few other very special cases), which then leads to the Kalman filter. In other cases, numerical approximation procedures should be used.

In the Bayesian context, a way to take into account the unknown parameter  $\theta$  is the following augmented state space method :

$$X'_t \stackrel{\text{def}}{=} \begin{pmatrix} \theta_t \\ X_t \end{pmatrix} \quad (15)$$

where  $\theta_t$  has a constant dynamic  $\theta_t = \theta_{t-1}$  with the prior on the parameter  $\theta$  as an initial law.

### Non-sequential approach

Our aim is to determine the posterior pdf  $p(X_{1:T}, \theta | Y_{1:T})$ . From Equation (12), we get :

$$p(X_{1:T}, \theta | Y_{1:T}) \propto \left[ \prod_{t=1}^T p(Y_t | X_t, \theta) \right] \times \left[ \prod_{t=1}^T p(X_t | X_{t-1}, \theta) \right] \times p(\theta) \quad (16)$$

that again features the product :

$$[\text{posterior pdf}] \propto [\text{likelihood}] \times [\text{system pdf (prior)}] \times [\text{hyper-parameter pdf (prior)}].$$

As in the static case, the expression 16 for the posterior pdf is not usable because it cannot be explicitly integrated. Nevertheless, its particular structure, graphically represented in Figure 7, will be adapted for MCMC methods. The fact that the expression (16) is known up to a multiplicative constant will not be a problem.

### 3. Computational Bayesian inference

#### 3.1. Introduction

The renewal of interest in Bayesian methods mainly originates in the new developments of Monte Carlo methods [29]. The integration of the posterior pdf (7) in the static case, as the integration of the expressions (14) or (16) in the dynamic case cannot be made explicitly. Monte Carlo methods are specifically adapted for approximation of such expressions. This efficiency is also due to the development of pseudo-random number generators and the ever increasing performance of computers. Monte Carlo methods go far beyond the question of computational statistics.

The aim of Monte Carlo methods is to approximate deterministic quantities by means of random simulations. In order to obtain an empirical approximation of a posterior pdf  $p(\theta|Y)$ , one generate a sample of size  $N$  of that pdf :

$$\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(N)} \stackrel{\text{iid}}{\sim} p(\theta|Y) \quad (17)$$

then, according to the law of large numbers,

$$\mathbb{P}[\theta \in B|Y] \simeq \frac{1}{N} \sum_{i=1}^N 1_B(\theta^{(i)}) \quad \mathbb{E}[\phi(\theta)|Y] \simeq \frac{1}{N} \sum_{i=1}^N \phi(\theta^{(i)})$$

that is :

$$p(\theta|Y) \simeq p^N(\theta|Y) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^N \delta_{\theta^{(i)}}.$$

Hence  $p^N(\theta|Y)$  appears to be an *empirical* approximation of the pdf  $p(\theta|Y)$ . This is an empirical approximation insofar as it is based on *in silico* (computer) experiments of the underlying phenomenon. Monte Carlo methods can therefore be seen as in silico experimental methods.

Contemporary Monte Carlo methods are among the algorithms adapted to computers. They were indeed originally developed to be used on the first computer at Los Alamos Laboratory during World War II. Around John von Neumann, scientists like Nicholas Metropolis and Stanislaw Ulam are behind the Monte Carlo method in its contemporary version (the origins of the method are much older) [34, 26, 32]. These scientists also initiated Monte Carlo Markov chain methods [33].

Here, we do not develop the general aspects of Monte Carlo methods (see [29] for such a presentation), but will present the Monte Carlo methods that are behind the success of computational Bayesian methods.

#### 3.2. Monte Carlo Markov chains methods (MCMC)

It is almost always not possible to sample according to a given target density like in (17). Let  $\pi(z)$  denote this target density. For some probability distributions – like the uniform distribution, the Gaussian distributions etc. – there are specific algorithms for generating pseudo-random numbers. Suppose that we did not know easily how to sample according to the target pdf  $\pi(z)$ , but that the analytical expression for this density is known, up to a multiplicative constant (generally this constant is not known and cannot be easily computed). The aim of the method is to (numerically) build a Markov chain  $(Z^{(k)})_{k \geq 0}$  whose limit density is precisely  $\pi(z)$ . By simulating *sufficiently many iterations* of this

chain we will obtain the desired sample. The convergence of this approach is based on ergodic like theorems [40, 22, 31].

### 3.2.1. Metropolis-Hastings sampler

The basic algorithm, called Metropolis-Hastings sampler, answered the following question : how to build a Markov chain (therefore a transition pdf) whose limit density is  $\pi(z)$  ? At first glance it seems impossible. The idea is to use (almost) any transition pdf  $q^{\text{prop}}(z'|z)$ , called proposal transition, and to astutely perturb it in order that it admits  $\pi(z)$  as an invariant density.

Starting from a point  $Z^{(0)}$ , the Metropolis-Hastings iteration  $Z^{(k)} \rightarrow Z^{(k+1)}$  proceeds in two steps :

- generate a new candidate :

$$\tilde{Z} \sim q^{\text{prop}}(\cdot | Z^{(k)}) \quad (18a)$$

- acceptance/rejet

$$Z^{(k+1)} \leftarrow \begin{cases} \tilde{Z} & \text{with probability } \alpha \quad (\text{acceptance}) \\ Z^{(k)} & \text{with probability } 1 - \alpha \quad (\text{rejection}) \end{cases} \quad (18b)$$

The probability  $\alpha$  is chosen in such a way that  $\pi$  is invariante for the Metropolis-Hastings transition. One can easily check that :

$$\alpha = \frac{\pi(\tilde{Z}) q^{\text{prop}}(\tilde{Z} | Z^{(k)})}{\pi(Z^{(k)}) q^{\text{prop}}(Z^{(k)} | \tilde{Z})} \wedge 1 \quad (18c)$$

This method is presented in Algorithm 1.

```

initial configuration  $z$ 
for  $k = 1, 2, \dots$  do
   $\tilde{z} \sim q^{\text{prop}}(\cdot | z)$  {sampling a new candidate}
   $\alpha \leftarrow \frac{\pi(\tilde{z}) q^{\text{prop}}(\tilde{z} | z)}{\pi(z) q^{\text{prop}}(z | \tilde{z})}$ 
  if  $\alpha > \text{rand}()$  then
     $z \leftarrow \tilde{z}$  {acceptance}
  end if
end for

```

**Algorithm 1:** *Metropolis–Hastings sampler* ( $\text{rand}()$  is the uniform distribution generator  $U[0, 1]$ ).

This algorithm can be applied under relatively broad conditions, it is simply necessary that the Metropolis-Hastings ratio (18c) is defined. However, conditions that ensure a fast enough convergence in practice are very difficult to establish on a theoretical point of view as well as a practical point of view [3, 13, 28].

### 3.2.2. Gibbs sampler

The Gibbs sampler can be seen as a variant of the Metropolis-Hastings algorithm. Suppose that the vector state  $Z$  either vector form : Suppose that the state vector is :

$$Z = (Z_1, \dots, Z_T)$$

and suppose that we know how to sample from the conditional marginal pdf's :

$$q_t^{\text{prop}}(z'_t | z_{-t}) = p(Z_t = z'_t | Z_{-t} = z_{-t}) \quad (19)$$

where  $Z_{-t} = \{Z_{-s}; s = 1 \dots T; s \neq t\}$ .

Starting from an initial configuration  $Z_{1:T}^{(0)}$ , the method propose to select  $t$  at random (or sequentially) and to update the component  $t$  by letting :

$$Z_t^{(k+1)} \sim q_t^{\text{prop}}(\cdot | Z_{-t}^{(k)})$$

where other components remain unchanged, i.e.  $Z_s^{(k+1)} = Z_s^{(k)}$  for  $s \neq t$ . This method is presented in Algorithm 2.

```

choose an initial configuration  $z_{1:T}$ 
for  $k = 1, 2 \dots$  do
  choose  $t$  at random in  $\{1, \dots, T\}$ 
   $z_t \sim p(Z_t | Z_{-t} = z_{-t})$ 
end for

```

**Algorithm 2: Gibbs sampler.**

### 3.2.3. Hybrid Metropolis-Hastings sampler

Suppose now that we do not know how to sample according to the conditional marginal pdf's (19). One can use at each iteration of the Gibbs sampler, a Metropolis-Hastings technique : instead of  $q_t^{\text{prop}}(z'_t | z_{-t})$  defined by (19), we consider a proposition kernel and use an acceptance/rejection technique like in (18b). This leads to the hybrid Metropolis-Hastings method (also called "Metropolis within Gibbs" sampler). It is necessary to decompose the marginal conditional pdf's as follows :

$$p(Z_t | Z_{-t}) \propto \underbrace{q_t^{\text{prop}}(Z_t | Z_{-t})}_{\text{proposition kernel}} \times \underbrace{\lambda_t(Z_t, Z_{-t})}_{\text{likelihood}} \quad (20)$$

This method leads to Algorithm 3.

```

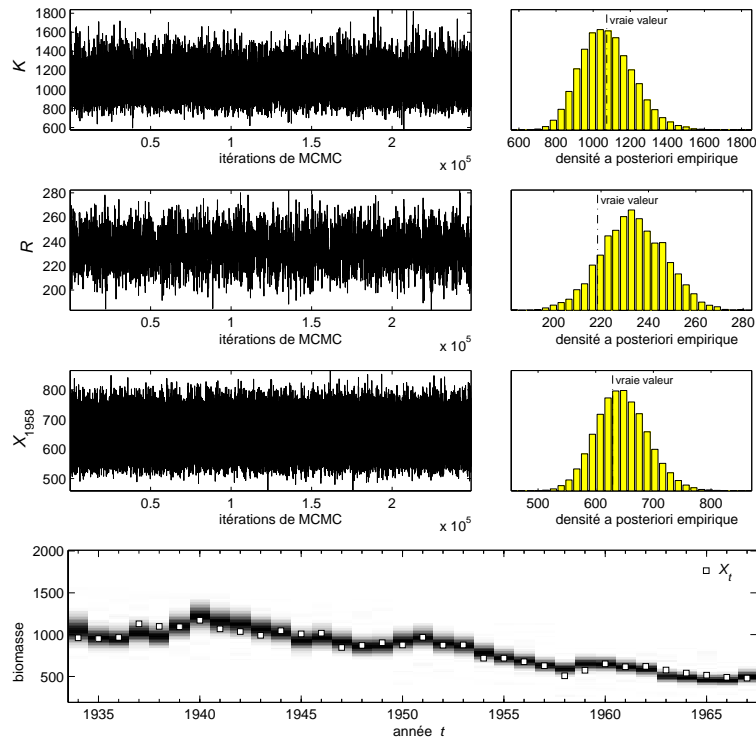
choose an initial configuration  $z_{1:T}$ 
for  $k = 1, 2 \dots$  do
  for  $t = 1 : T$  do
     $z'_t \sim q_t^{\text{prop}}(\cdot | z_{-t})$  {generate a candidate}
     $\alpha \leftarrow \lambda_t(z'_t, z_{-t}) / \lambda_t(z_t, z_{-t})$  {cf. Equation (20)}
    if  $\alpha > \text{rand}()$  then
       $z_t \leftarrow z'_t$  {the new configuration is accepted}
    end if
  end for
end for

```

**Algorithm 3: Hybrid Metropolis-Hastings sampler** ( $\text{rand}()$  is the uniform law generator  $U[0, 1]$ ).

### 3.2.4. Application

The Hybrid Metropolis-Hastings sampler can potentially be applied to all hierarchical Bayes models presented in Section 2 :



**Figure 8.** Application of the hybrid Metropolis-Hastings sampler (cf. Section 3.2.3) to the Example 2.2 of fishery with a Deriso–Schnute model. Top/left : For the components  $K$ ,  $R$ ,  $X_{1958}$ , the iterations of the MCMC procedure are displayed on the left and the resulting empirical a posteriori pdf's (together with the true values of the parameters) on the right. Bottom : the biomass time series  $X_t$  with, for each year  $t$ , the corresponding empirical a posteriori pdf's represented in grey levels.

– For the static case (6) corresponding to Figure 2, the goal is to sample from the conditional pdf of :

$$Z = (\rho, \theta, \mu, \sigma^2)$$

given observations  $(Y, X)$ .

– For the dynamic case (12) corresponding to Figure 7, the goal is to sample from the conditional pdf of :

$$Z = (X_1, \dots, X_T, \theta)$$

given observations  $(Y_1, \dots, Y_T)$ .

For instance, in this last case and for the Example 2.2 (see [7]), typical results are depicted in Figure 8.

This method applies in a natural way to the hierarchical Bayes models. Take the expression (16) for the posterior pdf, and consider for example the conditional pdf of  $X_2$  given  $(X_{-2}, Y_{1:T}, \theta)$ . It is clear that :

$$p(X_2 | X_{-2}, Y_{1:T}, \theta) \propto p(Y_2 | X_2, \theta) p(X_2 | X_1, \theta) p(X_3 | X_2, \theta) .$$



hence, the decomposition (20) could be :

$$q^{\text{prop}}(X_2|X_{-2}) \stackrel{\text{def}}{=} p(X_2|X_1, \theta),$$

$$\lambda(X_2|X_{-2}) \stackrel{\text{def}}{=} p(Y_2|X_2, \theta) p(X_3|X_2, \theta)$$

(The other components of  $X_{1:T}$  and  $\theta$  are treated the same way). This choice for the proposal kernel  $q^{\text{prop}}(X_2|X_{-2})$  is perhaps not the most efficient, but it applies to all hierarchical Bayes models and all hidden Markov models.

MCMC methods are extremely successful. Indeed, they are simple to set, they allow many variants and can be applied to many problems. They can be interconnected with other Monte Carlo methods and can be applied to hidden Markov models and to hierarchical Bayes models. This success is also due to software like WinBUGS or OpenBUGS (that also can be called from inside the R statistical package).

However, these methods may feature poor mixing properties and can be very slow. This is particularly the case with nonlinear systems in high dimension like the systems analyzed here by the hybrid Metropolis-Hastings method.

This latter approach is used extensively because in many situations it is the only method that can be applied. Current effort is focused on interacting parallel versions of such methods [5] and on the comparisons with other methods [19].

### 3.3. Sequential Monte Carlo methods (SMC)

Equations (14) for the nonlinear optimal filter can be used in practice. The purpose of the Monte Carlo sequential methods, also called *particle filters*, is to propose a Monte Carlo approximation of the optimal filter. These methods are now widely developed in practice and have a mathematical framework [, 14]. They were proposed in their present form in the early 1990's [24].

For the sake of simplicity, we consider Equations (14), i.e. without unknown parameter. The goal of the SMC method is to propose an empirical approximation of  $p(X_t|Y_{1:t})$  :

$$p(X_t|Y_{1:t}) \simeq p^N(X_t|Y_{1:t}) = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_t^{(i)}}(X_t).$$

here the question is to determine the positions  $\xi_t^{(i)}$  of the  $N$  particles. It is also possible to seek an importance sampling approximation :

$$p(X_t|Y_{1:t}) \simeq p^N(X_t|Y_{1:t}) = \sum_{i=1}^N \omega_t^{(i)} \delta_{\xi_t^{(i)}}(X_t).$$

and the question here is to determine the positions  $\xi_t^{(i)}$  and weights  $\omega_t^{(i)}$  of the  $N$  particles.

For the bootstrap approximation, which is the simplest implementation of SMC methods, the iteration  $p^N(X_{t-1}|Y_{1:t-1}) \rightarrow p^N(X_t|Y_{1:t})$  is in two steps :

**Prediction (mutation).** For each  $i = 1, \dots, N$ , one compute the predicted particle positions with the help of the transition kernel  $p(X_t|X_{t-1})$  of the Markov chain :

$$\xi_t^{(i)} \sim p(X_t|X_{t-1} = \xi_{t-1}^{(i)}) \quad (21a)$$

independently of one another.

Correction (selection). With the new measurement  $Y_t$  one computes the weight (likelihood) of each particle :

$$\omega_{t-}^{(i)} \propto p(Y_t | X_t = \xi_{t-}^{(i)}) \quad (21b)$$

then one resample the particles :

$$\xi_t^{(1)}, \xi_t^{(2)}, \dots, \xi_t^{(N)} \sim \sum_{i=1}^N \omega_{t-}^{(i)} \delta_{\xi_{t-}^{(i)}} \quad (21c)$$

this step consists in duplicate particles with high weight (the ones that most likely match the observation  $Y_t$ ) and in suppressing particles with low weight (the ones which most likely do not match the observation  $Y_t$ ).

For this method to be applied we need to :

- sample from the state system, i.e. to simulate  $X_t$  for a given  $X_{t-1}$ , cf. (21a).
- compute the likelihood associated with each particle, i.e. compute  $\xi \mapsto p(Y_t | X_t = \xi)$  for  $N$  values of  $\xi$ , cf. (21b).
- resample the particles cf. (21c).

It is therefore not necessary to know the analytical expression of the transition kernel pdf for the Markov chain  $X_t$ , it is only necessary to “mimic” the evolution of the system.

SMC methods have the same qualities as the MCMC methods : there feature many variants, they are relatively simple to implement (only the resampling procedure requires some attention).

Among the related methods, it should be noted that the ensemble Kalman filter (EnKF) method is increasingly used in sequential data assimilation. SMC methods have been very successful in challenging applications where the real-time constraints is very strong (movement or object tracking in video sequences, robots tracking, cellphone geo-location tracking etc.).

```

 $\xi_{1-}^{(i)} \sim p(X_1)$  for  $i = 1 : N$ 
 $\omega_{1-}^{(i)} \propto p(Y_1 | X_1 = \xi_{1-}^{(i)})$  for  $i = 1 : N$ 
 $\xi_{1-}^{(i)} \sim \sum_{j=1}^N \omega_{1-}^{(j)} \delta_{\xi_{1-}^{(j)}}$  for  $i = 1 : N$ 
for  $t = 2 : N$  do
   $\xi_{t-}^{(i)} \sim p(X_t | X_{t-1} = \xi_{t-1}^{(i)})$  for  $i = 1 : N$ 
   $\omega_{t-}^{(i)} \propto p(Y_t | X_t = \xi_{t-}^{(i)})$  for  $i = 1 : N$ 
   $\xi_{t-}^{(i)} \sim \sum_{j=1}^N \omega_{t-}^{(j)} \delta_{\xi_{t-}^{(j)}}$  for  $i = 1 : N$ 
end for

```

**Algorithm 4:** An example of SMC method : the bootstrap filter allows for the simulation of an empirical approximation  $p^N(X_t | Y_{1:t})$  of the optimal filter (14) for the system (12).

### 3.4. Comparison

MCMC and SMC differs on 3 essential points :

- MCMC methods are iterative methods and SMC are not. SMC methods sequentially process the observations  $Y_t$  for  $t = 1$  to  $T$ . MCMC methods are iterative and questions of diagnosing convergence is quite difficult.
- SMC methods (in their basic versions) approximate  $p(X_t|Y_{1:t})$  while MCMC methods approximate  $p(X_t|Y_{1:T})$ . The first is a filtering problem, the second is a smoothing problem. Smoothing methods for SMC are not yet well developed. The difference is important : for  $t = 1$  in the SMC case one considers  $p(X_1|Y_1)$  whereas in the MCMC case one considers  $p(X_1|Y_{1:T})$ . The variance of the latter expression is lower.
- Taking into account the identification of unknown parameters is very different in the two approaches. MCMC methods take these into account in a natural way. For SMC methods, the state augmentation method (15) has its limits. Alternative SMC methods using kernel techniques could be applied [6].

An important point is that all SMC methods have been developed with expressed aim of being applied to real-time applications. This constraint does not exist in the applications considered here. It is therefore possible to use cumbersome methods. One natural idea is to use MCMC methods to propagate particles (see for example the “resample-move” algorithm proposed by Gilks and Berzuini [21]). Instead of propagating particles according to the transition kernel of the Markov chain (cf. (21a)), one can proposed sampling from a more relevant target pdf with a MCMC technique.

---

## 4. Conclusions

During the past fifteen years, Monte Carlo methods have developed considerably. They provide a computational framework for Bayesian inference methods. Compared with frequentist approaches, Bayesian approaches are best suited to applications in ecology where one usually has limited amount of data. There is now percolation between application, probability modeling and computational inference. One reason for this success is the availability of efficient software accessible from widespread platforms like R. Statistical inference was often wrongly opposed to modeling, notably to deterministic modeling. With the development of Markov modeling, inference and modeling have formed a fruitful dual relationship. The couple Markov modeling and Bayesian inference now fits in a computational framework which makes it a powerful tool for in silico experiment and analysis.

---

## 5. Bibliographie

- [1] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer–Verlag, 1985. 2nd Edition.
- [2] S. P. Brooks. Bayesian computation : A statistical revolution. *Transactions of the Royal Society, Series A*, 361 :2681–2697, 2003.
- [3] S. P. Brooks and G. O. Roberts. Convergence assessment techniques for Markov chain Monte Carlo. *Statistics and Computing*, 8(4) :319–335, 1998.

- [4] S.T. Buckland, K.B. Newman, C. Fernández, L. Thomas, and J. Harwood. Embedding population dynamics models in inference. *Statistical Science*, 22(1) :44–58, 2007.
- [5] F. Campillo, P. Cantet, R. Rakotozafy, and V. Rossi. Méthodes mcmc en interaction pour l'évaluation de ressources naturelles. *Revue Africaine de la Recherche en Informatique et Mathématiques Appliquées (ARIMA)*, 2007. To appear.
- [6] F. Campillo and V. Rossi. Convolution filter based methods for parameter estimation in general state–space models. *IEEE Transactions on Aerospace and Electronic Systems*, 2008. To appear.
- [7] F. Campillo and R. Rakotozafy. MCMC for nonlinear/non–Gaussian state–space models, Application to fishery stock assessment. In *CARI'04*, Hammamet, Tunisia, 2004.
- [8] S. Chib and E. Greenberg. Markov Chain Monte Carlo simulation methods in econometrics. *Econometric Theory*, 12 :409–431, 1996.
- [9] S. Chib, F. Nardari, and N. Shephard. Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, 108(2) :281–316, 2002.
- [10] J. S. Clark. Why environmental scientists are becoming Bayesians. *Ecology Letters*, 8(1) :2–14, 2005.
- [11] J. S. Clark. *Models for Ecological Data*. Princeton University Press, 2007.
- [12] J. S. Clark and A. E. Gelfand. A future for models and data in environmental science. *Trends in Ecology & Evolution*, 21(7) :375–380, July 2006.
- [13] M. K. Cowles and B. P. Carlin. Markov chain Monte Carlo convergence diagnostics : a comparative review. *Journal of the American Statistical Association*, 91(434) :883–904, 1996.
- [14] P. Del Moral. *Feynman-Kac formulae – Genealogical and interacting particle approximations*. Springer–Verlag, New York, 2004.
- [15] A. Doucet, N. de Freitas, and N. J. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer–Verlag, New York, 2001.
- [ ] A. Doucet, A. Logothetis, and V. Krishnamurthy. Stochastic sampling algorithms for state estimation of jump Markov linear systems. *IEEE Transactions on Automatic Control*, AC–45(2) :188–202, February 2000.
- [16] A. M. Ellison. Bayesian inference in ecology. *Ecology Letters*, 7(6) :509–520, 2004.
- [17] S. E. Fienberg. When did Bayesian inference become “Bayesian”? *Bayesian Analysis*, 1(1) :1–40, 2006.
- [18] A. Franc, S. Gourlet-Fleury, and N. Picard. *Une introduction à la modélisation des forêts hétérogènes*. ENGREF, 2000.
- [19] C. Gaucherel, F. Campillo, L. Misson, J. Guiot, and J.-J. Boreux. Parameterization of a process-based tree-growth model : comparison of optimization, MCMC and particle filtering algorithms. *Environmental Modelling and Software*, 23(10-11) :1280–1288, 2008.
- [20] S. G. Giakoumatos, P. Dellaportas, and D. N. Politis. Bayesian analysis of the unobserved arch model. *Statistics and Computing*, 15(2) :103–111, 2005.
- [21] W. R. Gilks and C. Berzuini. Following a moving target — Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society, Series B*, 63(1) :127–146, 2001.
- [22] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov Chain Monte Carlo in practice*. Chapman & Hall, London, 1996.
- [23] O. Gimenez, S. Bonner, R. King, R. A. Parker, S.P. Brooks, L. E. Jamieson, V. Grosbois, B. J. T. Morgan, and L. Thomas. WinBUGS for population ecologists : Bayesian modeling using Markov Chain Monte Carlo methods. *Environmental and Ecological Statistics*, 2008. In press.
- [24] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non–Gaussian Bayesian state estimation. *IEE Proceedings, Part F*, 140(2) :107–113, April 1993.

- [25] P. J. Green. Reversible jump MCMC computation and Bayesian model determination. *Biometrika*, 82(4) :711–732, December 1995.
- [26] F. Harlow and N. Metropolis. Computing and computers – weapons simulation leads to the computer era. *Los Alamos Science*, 7 :132–141, 1983.
- [27] D. Howie. *Interpreting probability : Controversies and developments in the early twentieth century*. Cambridge University Press, 2002.
- [28] R. E. Kass, B. P. Carlin, A. Gelman, and R. M. Neal. Markov Chain Monte Carlo in practice : A roundtable discussion. *The American Statistician*, 52 :93–100, 1998.
- [29] J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer–Verlag, New York, 2001.
- [30] D. Malakoff. Bayes offers a “new” way to make sense of numbers. *Science*, 286 :1460–1464, 1999.
- [31] J.-M. Marin and C. P. Robert. *Bayesian Core : A Practical Approach to Computational Bayesian Statistics*. Springer-Verlag, 2007.
- [32] N. Metropolis. The beginning of the Monte Carlo method. *Los Alamos Science*, 15 :125–130, 1987.
- [33] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6) :1087–1091, 1953.
- [34] N. Metropolis and S. Ulam. The Monte Carlo method. *Journal of the American Statistical Association*, 44(247) :335–341, February 1949.
- [35] 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.
- [36] A. E. Punt and R. Hilborn. Fisheries stock assessment and decision analysis : the Bayesian approach. *Reviews in Fish Biology and Fisheries*, 7 :35–63, 1997.
- [37] K. H. Reckhow. Bayesian approaches in ecological analysis and modeling. In C. D. Canham, J. J. Cole, and W. K. Lauenroth, editors, *Models in ecosystem science*, pages 168–183. Princeton University Press, Princeton, New Jersey, USA., 2003.
- [38] E. Rivot, E. Prevost, E. Parent, and J.-L. Blaginière. A Bayesian state-space modelling framework for fitting a salmon stage-structured population dynamic model to multiple time series of field data. *Ecological Modelling*, 179 :463–485, 2004.
- [39] Carl J. Schwarz and George A. F. Seber. Estimating animal abundance : Review III. *Statistical Science*, 14(4) :427–456, 1999.
- [40] L. Tierney. Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics*, 22(4) :1701–1728, December 1994.
- [41] J. R. Webster. Hierarchy theory and ecosystem models. In E. Halfon, editor, *Theoretical systems ecology*, pages 119–129. Academic Press, 1978.