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 :

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

This notation cannot be used in a mathematical framework, it generates inaccuracies (it is indeed difficult to define a function from its argument). If θ is an unknown parameter, the joint pdf of (θ, Y), is denoted by $p(\theta, Y)$. The conditional pdf of θ 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 θ. 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 θ contained in the observation Y, in a priori knowledge on θ and in the model. Indeed, the Bayes formula writes :

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

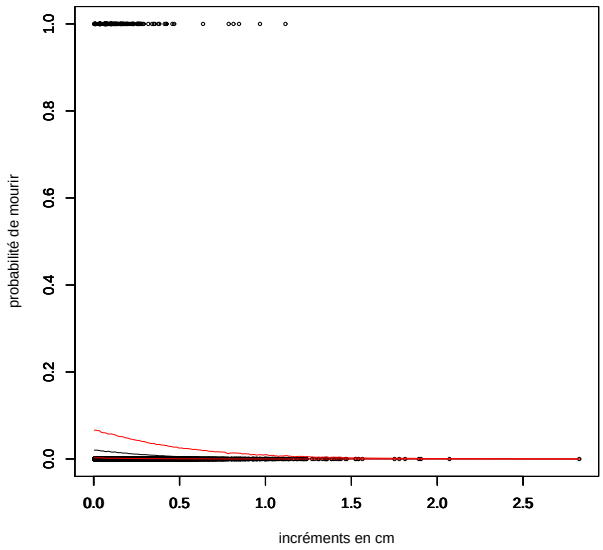

Figure 1. *Estimation of the mortality probability by a maximum likelihood approach from model (4). This estimate is plunged into the model (4), then by sampling the model one can calculate a confidence interval for each value of Δ. This interval is represented by 3 curves (average in the center and 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 :

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

with

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

One can compute the maximum likelihood estimate of θ 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 et 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 θ 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 $Y_i \sim \rho_i\,\text{Ber}(p_i^{\text{h}}) + (1-\rho_i)\,\text{Ber}(p_i^{\text{s}})$

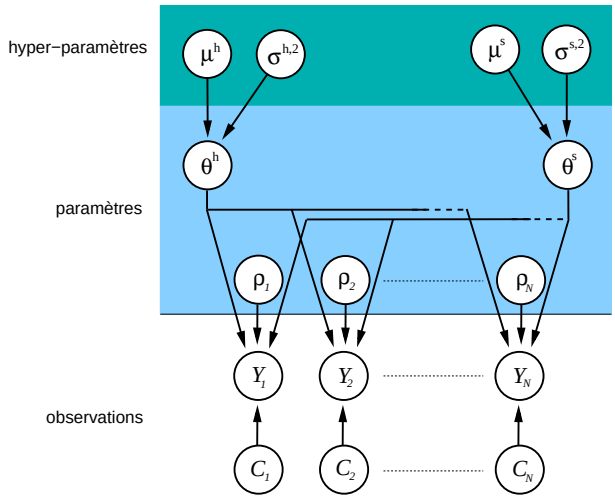

Figure 2. *Graphical representation of the model (6) in the form of a directed acyclic graph (DAG). This diagram represents the dependencies of each variable on others. For example we can see that Y_1 depends on others variables only through C_1, ρ_1, θ^h and θ^s ; we can also see that conditionally 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 :

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

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].

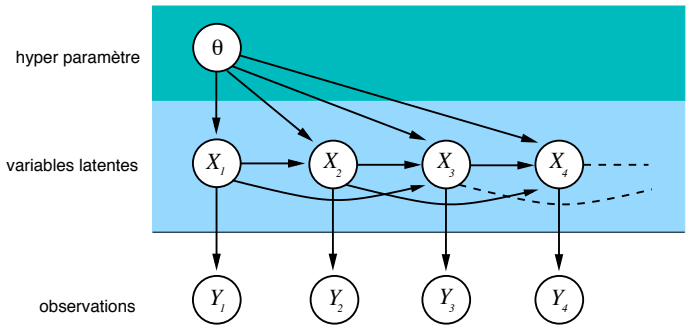

Figure 4. *Graphical representation associated with Equations (8) of the Example 2.2 of the Deriso–Schnute model. It is an hidden Markov model of order 2.*

where C_t denotes the catch during year t, ρ 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 σ_w^2 and σ_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 :*

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

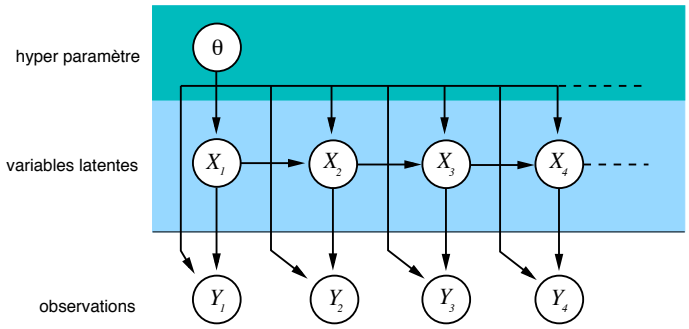

Figure 5. *Graphical representation associated with Equations (9) of Example 2.3 of the Ricker model. It is a first order hidden Markov model.*

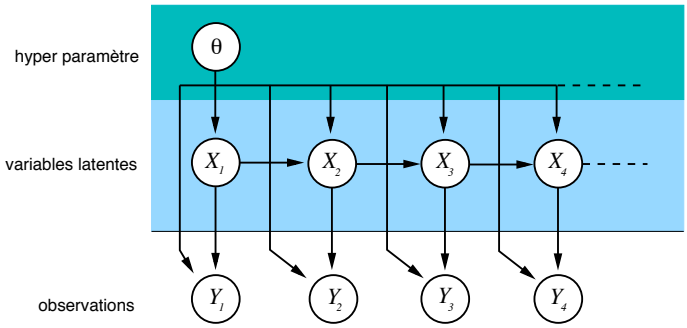

Figure 7. *Graphical representation of the hidden Markov model (12) with an unknown parameter.*

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

The second hypothesis is that, conditionally on θ and $X_{1:T}$, the observations Y_t are independent, and that Y_t depends on $(\theta, X_{1:T})$ only through (θ, 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 :

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

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 :

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

$$Y_t = h(\theta, X_t, v_t) \tag{13b}$$

where w_t and v_t are Gaussian white noise (i.i.d. variables with zero mean) ; w_t, v_t, X_1, θ 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.

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 :

$$\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(N)} \overset{\text{iid}}{\sim} p(\theta|Y) \tag{17}$$

then, according to the law of large numbers,

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

that is :

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

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

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

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

where $Z_{\neg t} = \{Z_{\neg s}\,;\, s = 1 \cdots T; s \neq t\}$.

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 :

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

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 $k = 1, 2 \ldots$ **do**
  choose t at random in $\{1, \ldots, T\}$
  $z_t \sim p(Z_t|Z_{\neg t} = z_{\neg t})$
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 $q_t^{\text{prop}}(z'_t|z_{\neg t})$ 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 :

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

This method leads to Algorithm 3.

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

Algorithm 3: *Hybrid Metropolis-Hastings sampler (*`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 :

hence, the decomposition (20) could be :

$$q^{\text{prop}}(X_2|X_{\neg 2}) \stackrel{\text{def}}{=} p(X_2|X_1,\theta)\,,$$
$$\lambda(X_2|X_{\neg 2}) \stackrel{\text{def}}{=} p(Y_2|X_2,\theta)\,p(X_3|X_2,\theta)$$

(The other components of $X_{1:T}$ and θ are treated the same way). This choice for the proposal kernel $q^{\text{prop}}(X_2|X_{\neg 2})$ 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 $p^N(X_{t-1}|Y_{1:t-1}) \to p^N(X_t|Y_{1:t})$ is in two steps :

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

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

independently of one another.

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.

[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.