

Stochastic virtual population analysis

M. Virtala, S. Kuikka, and E. Arjas



Virtala, M., Kuikka, S., and Arjas, E. 1998. Stochastic virtual population analysis. ICES Journal of Marine Science, 55: 892–904.

The Metropolis–Hastings algorithm was applied to a standard Virtual Population Analysis using a simulated data set containing catch data from 10 age groups over a 10-year period. Posterior distributions of fishing mortalities and population sizes are studied under different prior assumptions. The results demonstrate the usefulness of obtaining distributions of plausible parameter values instead of point estimates. Computational aspects of the algorithm are also discussed, as well as the estimation of the natural mortality.

© 1998 International Council for the Exploration of the Sea

Key words: Markov chain Monte Carlo estimation, catch-at-age, fisheries population model.

Received 21 April 1997; accepted 12 April 1998.

M. Virtala: Department of Mathematical Sciences, University of Oulu, FIN-90570 Oulu, Finland. S. Kuikka: Finnish Game and Fisheries Research Institute, PO Box 202, FIN-00151 Helsinki, Finland. E. Arjas: Rolf Nevanlinna Institute, PO Box 4, FIN-00014 University of Helsinki, Finland†. Correspondence to M. Virtala: tel: +358 8 5531834; fax: +358 8 5531848; e-mail: matti.virtala@oulu.fi*

Introduction

The Bayesian paradigm for modelling and analysing data has received increasing attention among fisheries biologists. A possible explanation is that it offers a flexible methodology for estimation, enabling one to express uncertainties regarding the true values of the unknown model parameters in terms of probability distributions. This is important from the point of view of quality assessment of the data, as well as for maintaining a realistic attitude towards risks in the decision making for fisheries management. The precautionary principle is an essential element of the new international fisheries agreements and management rules, reflecting a risk averse attitude in the management. In this, the assessment of the variances of different decision alternatives is an essential ingredient in the scientific analysis. For scientists, wide posterior distributions suggest that there is little information or inconsistencies in the data, or that an inappropriate model is being used. For managers, the shape of posterior distributions should be an important signal of the quality of information.

A simple example of the role of uncertainty in risk averse decision making is given in Figure 1 (Kuikka

and Varis, 1991). In this case, the size of the surviving wild salmon parent stock is estimated by a simple Monte Carlo simulation model, and the effect of TAC on the size of parent stock is calculated. The size of the needed parent size has been estimated to be 14 000 salmon. This is based on the genetical needs of the separate populations. The allowed TAC, based on a point estimate model, would be around 650 000 salmon (point A in Fig. 1). However, by using a risk averse attitude the allowed TAC would be lower. If one aims to have a 90% chance of getting the required number of spawning salmon to the rivers, the TAC should be less than 450 000 salmon (point B). High uncertainty (i.e. low quality of scientific knowledge) leads to a lower TAC. Managers and fishermen should have an interest in investing in better knowledge, because better knowledge would enable higher catches in the long run.

Both under- and overestimation of uncertainties should be avoided. For example, in a TAC based fisheries management underestimation of uncertainties will lead to decisions for higher quotas than are warranted. In the case of overestimation of uncertainties, the insurance fee paid in the form of lower quotas is too high for the fishermen. Due to these facts, the realistic assessment of probability distributions should be favoured by both scientists and managers.

*E-mail: sakari.kuikka@rktl.fi

†E-mail: elata@rni.helsinki.fi

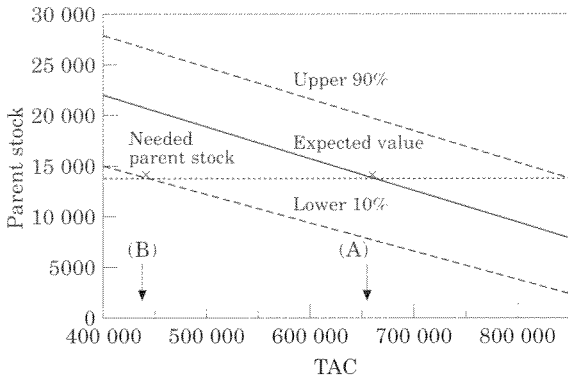


Figure 1. Example of the effect of uncertainty on the allowed TAC in a risk averse salmon fisheries management decision (from Kuikka and Varis, 1991). See text for explanations.

There are several recent papers applying Bayesian methodology to stock assessment (e.g. McAllister *et al.*, 1994; Walters and Ludwig, 1994; Walters and Punt, 1994; McAllister and Ianelli, 1997). The computational efficiency of the algorithms is not discussed in detail in the published papers. However, it is an essential consideration when the practical application of these algorithms is important, as for example in ICES working groups.

McAllister and Ianelli (1997) discuss the use of different Bayesian algorithms in fish stock assessment. The aim of this paper is to apply the Metropolis-Hastings algorithm (Hastings, 1970) to a standard Virtual Population Analysis (VPA, Pope, 1972), forming a particular version of the Markov chain Monte Carlo (MCMC) estimation techniques. First this methodology is described and then it is applied to a simulated catch-at-age data. The inclusion of other data sets and the computational efficiency of the method are also discussed briefly.

Principles of Bayesian inference

Suppose that the vector $\theta=(\theta_1, \dots, \theta_k)$ represents the unknown parameters in a model and let the vector $y=(y_1, \dots, y_n)$ contain the available data. Before any analysis of the data, we may have some prior beliefs about plausible values of the unknown parameters. This can for example be based on historical experience with other similar stocks (e.g. Punt and Hilborn, 1997). We can express these beliefs by defining a prior density $\pi(\theta)$ for the parameter vector θ . If we have no prior information about a given parameter, a "non-informative" prior can be used. If, for example, a parameter is restricted to the interval (0,1), one might use a prior density which has the constant value 1 over the whole interval. For a more detailed discussion of selecting a prior distribution in the context of fisheries management, see Walters and

Ludwig (1994). If the parameters can be considered independent, a prior distribution can be specified for each one separately, giving the product:

$$\pi(\theta)=\pi(\theta_1)\pi(\theta_2) \dots \pi(\theta_k).$$

The dependence of the data vector y on the parameter vector θ is given by the density function:

$$f(y|\theta).$$

If the model is deterministic, a given set of parameter values determines the data values uniquely. In a stochastic model $f(y|\theta)$ will be a genuine, non-degenerate probability density function. The joint probability density $f(y,\theta)$ of data and parameters is the product of the prior density $\pi(\theta)$ and the conditional density $f(y|\theta)$, i.e.

$$f(y,\theta)=\pi(\theta)f(y|\theta). \tag{1}$$

Given the available data, we can modify our prior beliefs about plausible parameter values by conditioning on the additional information provided by the data. The new beliefs about the parameter values are then expressed by the posterior density $\pi(\theta|y)$, which can be written as:

$$\pi(\theta|y)=\frac{f(y,\theta)}{f(y)}=\frac{\pi(\theta)f(y|\theta)}{\int \pi(\theta)f(y|\theta)d\theta}. \tag{2}$$

The quantity $\int \pi(\theta)f(y|\theta)d\theta$ in the denominator is essentially a proportionality constant needed to make the conditional density integrate to unity. We can thus write Equation (2) in the form:

$$\pi(\theta|y) \propto \pi(\theta)f(y|\theta). \tag{3}$$

The density $f(y|\theta)$, when considered as a function of θ , is called the likelihood and is denoted also by $L(y|\theta)=f(y|\theta)$.

Given the data, the prior density of the parameters, and the likelihood, Equation (2) shows how the posterior density of the parameters can be determined. When the set of parameters is large, calculating the posterior density analytically can be difficult or even impossible. Numerical methods must then be applied. The general idea behind Monte Carlo simulation is that, if it is possible to generate a large synthetic sample from the target distribution, the corresponding empirical distribution is used as the corresponding numerical approximation of the posterior distribution. For very high-dimensional distributions, however, drawing such a sample may be impossible in practice, even when the analytic form of the density is known. This is because for an exact simulation, the random vectors forming such a sample must be drawn one coordinate at a time, and, in

order to get started, one would have to “marginalize” all other dimensions except one from such a joint density. To make things even worse, in Bayesian estimation the starting point is almost always only the “up to proportionality” version (3) of Bayes’ formula, since calculating the proportionality constant $f(y)$ would involve an integration over the high-dimensional parameter space. For these reasons we have to look for conceptually somewhat more involved ways to generate a Monte Carlo sample. One of these ways is the Markov chain Monte Carlo Method (see e.g. Gelman *et al.*, 1995; Gilks *et al.*, 1996; Gamerman, 1997, and references therein). In that method consecutive sets of parameter values are generated sequentially. At each estimation cycle the current parameter values are updated to a new set of values. Having done this, say 1000 times and having stored the current parameter values from each cycle, one has produced empirical estimates of the posterior distributions of the model parameters.

The core of the MCMC method is the step from the current set of parameter values to the next set of values. The current parameter values describe, in a sense, the state of the model at that time, and the algorithm moves from that state to another. This movement is done in such a way that the resulting sequence of states constitutes a Markov chain, with the posterior as its limiting stationary distribution. The Monte Carlo aspect comes from the fact that the movement to the next state is not deterministic but depends on the prior distributions and the likelihood functions in the model. First we describe the methods used for parameter updating (see Besag *et al.*, 1995, for a more detailed account of the method), and then the method is used to analyse a stochastic version of Virtual Population Analysis. (For an account of deterministic Virtual Population Analysis see, for example, Hilborn and Walters, 1992.)

Markov chain Monte Carlo estimation

We can construct an ergodic Markov chain which has the desired posterior $\pi(\theta; y)$ as its stationary limiting distribution. Considering then a sequence of states $\theta^{(1)}$, $\theta^{(2)}$, ... from this chain, we find that sample path averages from a long sequence behave almost as if $\theta^{(1)}$, $\theta^{(2)}$, ... were drawn independently from $\pi(\theta; y)$. More exactly, if f is some bounded test function, the Ergodic Theorem implies that the averages $1/N \sum_{i=1}^N f(\theta^{(i)})$ tend to, with probability one, the limit $E_{\pi(\cdot|y)}(f(\theta))$ as $N \rightarrow \infty$.

To produce the sequence $\theta^{(1)}$, $\theta^{(2)}$, ... we need a method to move from state $\theta^{(t)}$ to state $\theta^{(t+1)}$. For simplicity, we shall in the following denote the current parameter vector by θ and the updated values by θ' . One possibility to construct θ' is to change only one component of the parameter vector at a time. When the i^{th} component is changed, the other components, denoted

collectively by $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k)$, remain constant and the parameter vector $\theta = (\theta_1, \dots, \theta_i, \dots, \theta_k)$ changes to $\theta' = (\theta_1, \dots, \theta'_i, \dots, \theta_k)$. The simplest method is to start from the first component θ_1 and sequentially update all components 2, ..., k. When this cycle is completed, we have obtained one new vector θ' .

A concrete possibility for such a simulation scheme is offered by the so called Metropolis–Hastings algorithm (Hastings, 1970). The procedure consists of two steps.

(1) First, a new proposal value θ'_i is generated from a transition kernel with density $R_i(\theta_i \rightarrow \theta'_i; \theta_{-i})$ which has the following properties:

- (i) $\theta'_{-i} = \theta_{-i}$
- (ii) $R_i(\theta_i \rightarrow \theta'_i; \theta_{-i}) > 0 \Leftrightarrow R_i(\theta'_i \rightarrow \theta_i; \theta_{-i}) > 0$.

Condition (i) states the fact that when moving from state θ to state θ' , beside θ_i , the other components θ_{-i} of the parameter vector are not changed. Condition (ii) requires that if the density $R_i(\cdot; \theta_{-i})$ for a move from θ_i to θ'_i is positive, then it must be positive also for the opposite move from θ'_i to θ_i (but not necessarily equal). The proposal density is essentially arbitrary and we may use any function we like, as long as condition (ii) is satisfied.

(2) The proposed new value θ'_i is not automatically accepted to replace the old value θ_i . Instead, in the second step the new vector $\theta' = (\theta_1, \dots, \theta'_i, \dots, \theta_k)$ is accepted as a new state with probability $A_i(\theta_i \rightarrow \theta'_i; \theta_{-i})$, where:

$$A_i(\theta_i \rightarrow \theta'_i; \theta_{-i}) = \min \left\{ 1, \frac{\pi(\theta'|y) R_i(\theta'_i \rightarrow \theta_i; \theta_{-i})}{\pi(\theta|y) R_i(\theta_i \rightarrow \theta'_i; \theta_{-i})} \right\}. \tag{4}$$

If θ' is not accepted, then the i^{th} component of θ is not changed.

If, in step 1, a proposal value for a parameter is generated using a symmetric proposal distribution R , then $R_i(\theta_i \rightarrow \theta'_i; \theta_{-i}) = R_i(\theta'_i \rightarrow \theta_i; \theta_{-i})$, and the acceptance probability (4) simplifies to:

$$A(\theta_i \rightarrow \theta'_i; \theta_{-i}) = \min \left\{ 1, \frac{\pi(\theta'|y)}{\pi(\theta|y)} \right\} = \min \left\{ 1, \frac{\pi(\theta') L(y|\theta')}{\pi(\theta) L(y|\theta)} \right\}. \tag{5}$$

This simpler version is called the Metropolis algorithm. As can be seen from (5), if a proposal θ' has a greater posterior density $\pi(\theta'; y)$ than is the density $\pi(\theta; y)$ of the current value θ , θ' is always accepted to replace θ .

When the sequence $\{\theta^{(1)}, \theta^{(2)}, \dots\}$ is constructed in the above way, the stationary distribution of the resulting Markov chain will be $\pi(\cdot; y)$ (for a proof see, for

example, Gamerman, 1997, pp. 171–172). However, usually the initial values of the sequence are not distributed according to $\pi(\cdot|y)$, for the very reason that we do not have a method for drawing observations directly from a high-dimensional posterior. Indeed, if it were possible to draw such values, we might as well have considered a simple random sample from the posterior, thereby avoiding completely the use of MCMC sampling methods. This complication is taken into account by running the chain $\text{bf}:\theta^{(1)}, \theta^{(2)}, \dots$ first through an initial “burn-in period”. It is hoped that after a sufficiently long period the values of $\text{bf}:\theta^{(t)}$ will have a distribution closer to that of $\pi(\cdot|y)$ (for some methods to monitor the convergence, see Gelman, 1996).

In the next Section we shall define a stochastic version of virtual population analysis and show how the Markov chain Monte Carlo method can be applied to calculate the posterior distributions of model parameters.

Stochastic Virtual Population Analysis

Let us begin by first summarizing the main features of deterministic virtual population analysis. Suppose that we have catch data C_{ay} from $a=1, \dots, A$ age classes and from $y=1, \dots, Y$ years. Our purpose is to estimate the size N_{ay} of the fish stock and the fishing mortality F_{ay} for $a=1, \dots, A, y=1, \dots, Y$. The natural mortality rate M is assumed known and constant over both time and age classes. Considering N as continuous and assuming that the fishing and natural mortality act on the population continuously throughout the year, the size of a cohort changes according to:

$$\frac{dN}{dt} = -N(M+F) \tag{6}$$

(see, for example, Hilborn and Walters, 1992, p. 354). If we let N_{ay} be the number of fish in the a^{th} age class alive at the beginning of the y^{th} year, then:

$$N_{a+1,y+1} = N_{ay} e^{-M - F_{ay}} \tag{7i}$$

will be alive at the beginning of the next year. Letting D_{ay} denote the number of fish dying from natural causes, we can write:

$$N_{ay} = N_{a+1,y+1} + D_{ay} + C_{ay}.$$

The values of D_{ay} and C_{ay} can be calculated as:

$$\begin{aligned} D_{ay} &= \frac{M}{M+F_{ay}}(N_{ay} - N_{a+1,y+1}) \\ &= \frac{M}{M+F_{ay}} N_{ay} (1 - e^{-M - F_{ay}}) \end{aligned} \tag{7ii}$$

$$\begin{aligned} C_{ay} &= \frac{F_{ay}}{M+F_{ay}}(N_{ay} - N_{a+1,y+1}) \\ &= \frac{F_{ay}}{M+F_{ay}} N_{ay} (1 - e^{-M - F_{ay}}). \end{aligned} \tag{7iii}$$

In virtual population analysis the values of N_{ay} and F_{ay} are estimated by applying backward induction, starting from the terminal class of a cohort and proceeding backwards in time to the start. Let (a,y) be the terminal class of a cohort. It contains two unknowns, N_{ay} and F_{ay} , with only the value of C_{ay} known and M assumed to be known. Hence either N_{ay} or F_{ay} must be given an estimated value. Usually an estimate of F_{ay} is provided (terminal F assumption). When F_{ay} is given, the value of N_{ay} follows from equation (7iii), or:

$$N_{ay} = \frac{M+F_{ay}}{F_{ay}} \cdot \frac{C_{ay}}{1 - e^{-M - F_{ay}}} \tag{8}$$

To estimate $N_{a-1,y-1}$ and $F_{a-1,y-1}$, we first solve for $F_{a-1,y-1}$ from (7i), getting:

$$F_{a-1,y-1} = -\log\left(\frac{N_{ay}}{N_{a-1,y-1}}\right) - M. \tag{9}$$

Substituting this into (7iii) results in the equation:

$$\begin{aligned} C_{a-1,y-1} &= \left[1 - \frac{M}{\log(N_{a-1,y-1}) - \log(N_{ay})} \right] \\ &\quad \times (N_{a-1,y-1} - N_{ay}). \end{aligned} \tag{10}$$

This can be solved for $N_{a-1,y-1}$ either numerically or, alternatively, using Pope’s (1972) approximation:

$$N_{a-1,y-1} = N_{ay} e^{M+C_{a-1,y-1}} e^{M/2}.$$

After the value of $N_{a-1,y-1}$ is determined, $F_{a-1,y-1}$ is calculated from (9).

A stochastic version of the above model can be defined as follows. Suppose that there are N_{ay} fish alive at the beginning of the y^{th} year in age class a , where $1 \leq a \leq A, 1 \leq y \leq Y$. According to Equation (7i), there will be $N_{a+1,y+1} = N_{ay} e^{-M - F_{ay}}$ fish alive at the beginning of the next year. However, instead of letting the value of $N_{a+1,y+1}$ to be a deterministic fraction $p = e^{-M - F_{ay}}$ of N_{ay} , we can consider each individual fish separately and interpret p as the probability that a fish avoids both natural and fishing mortality and thus survives to the next year. We have thus N_{ay} trials with a probability p of success, so that:

$$N_{a+1,y+1} \sim \text{Bin}(N_{ay}, e^{-M - F_{ay}}). \tag{11i}$$

From the $N_{ay} - N_{a+1,y+1}$ fish which do not survive to the next year, D_{ay} die from natural causes and C_{ay} are caught. Given that a fish of age a will not survive year y ,

the probability that it meets a natural death is $M/(M+F_{ay})$ and the probability that it contributes to the fishermen's catch is $F_{ay}/(M+F_{ay})$. Hence, given $N_{ay} - N_{a+1,y+1}$:

$$D_{ay} \sim \text{Bin}\left(N_{ay} - N_{a+1,y+1}, \frac{M}{M + F_{ay}}\right) \tag{11ii}$$

and:

$$C_{ay} \sim \text{Bin}\left(N_{ay} - N_{a+1,y+1}, \frac{F_{ay}}{M + F_{ay}}\right). \tag{11iii}$$

In fact, we can combine these formulae together, stating that the vector $(N_{a+1,y+1}, D_{ay}, C_{ay})$ has a multinomial distribution:

$$(N_{a+1,y+1}, D_{ay}, C_{ay}) \sim \text{Mult}(N_{ay}; p_1, p_2, p_3), \tag{12}$$

where:

$$p_1 = e^{-M - F_{ay}} \tag{13i}$$

$$p_2 = (1 - e^{-M - F_{ay}}) \frac{M}{M + F_{ay}} \tag{13ii}$$

$$p_3 = (1 - e^{-M - F_{ay}}) \frac{F_{ay}}{M + F_{ay}}. \tag{13iii}$$

Of course, $N_{a+1,y+1} + D_{ay} + C_{ay} = N_{ay}$ and $p_1 + p_2 + p_3 = 1$. Comparing Equations (11) to Equations (7) shows immediately that the deterministic model is obtained when the random variables $N_{a+1,y+1}$ and D_{ay} are replaced by their expected values. Since the catch data C_{ay} are known, we are in fact studying the conditional distributions of N and F given the data C .

In the deterministic model, estimation usually starts at the end of a cohort and proceeds towards the earlier age groups. If a value of either F or N is provided for the first age class, the equations can also be solved by proceeding from the start of a cohort towards the end. In the stochastic model the dynamics of the model and the likelihood functions are defined in a forward direction, but otherwise there is no explicit "direction", since a change in a certain parameter estimate affects the estimation of all those parameters which are related to it either in the forward or backward direction.

To be able to get posterior distributions for the parameters, we must for each unknown parameter either define a prior distribution or specify its conditional probability distribution given the values of the other quantities in the model. The number of required prior distributions might suggest that we need a lot of information to be able to define the model. This is not the case, however. If we have little prior information about a parameter, a non-informative prior can be used. On the other hand, we can easily take advantage of any

prior information about a parameter which we may have.

Consider now a birth cohort which is followed K years, from year 1 to year K , so that the indices (a,y) run from $(1,1)$ to (K,K) . The joint density of the data and parameters for that cohort is:

$$f(y, \theta) = f(C_{11}, \dots, C_{KK}, F_{11}, \dots, F_{KK}, N_{11}, \dots, N_{KK}, N_{K+1,K+1}).$$

Assuming prior independence of $N_{11}, F_{11}, F_{22}, \dots, F_{KK}$, and using the rules of conditional probability, this can be written as:

$$\begin{aligned} f(y, \theta) &= \pi(N_{11})\pi(F_{11}) \dots \pi(F_{KK}) \\ &\times p(C_{11}, \dots, C_{KK}, N_{22}, \dots, N_{K+1,K+1} | N_{11}, F_{11}, \dots, F_{KK}) \\ &= \pi(N_{11})\pi(F_{11}) \dots \pi(F_{KK}) \\ &\times p(C_{11}, N_{22} | N_{11}, F_{11})p(C_{22}, N_{33} | N_{22}, F_{22}) \\ &\dots \\ &\times p(C_{KK}, N_{K+1,K+1} | N_{KK}, F_{KK}). \end{aligned} \tag{14}$$

Consider, for example, the updating of F_{22} . Let us use a symmetric proposal distribution and let F'_{22} be the new proposal value for F_{22} . Using Equation (14) and cancelling common terms, the acceptance probability (5) reduces to:

$$\begin{aligned} A(F_{22} \rightarrow F'_{22}) &= \min \left\{ 1, \frac{\pi(F'_{22})p_1(F'_{22})^{N_{33}}p_2(F'_{22})^{N_{22} - N_{33} - C_{33}}p_3(F'_{22})^{C_{22}}}{\pi(F_{22})p_1(F_{22})^{N_{33}}p_2(F_{22})^{N_{22} - N_{33} - C_{33}}p_3(F_{22})^{C_{22}}} \right\}, \end{aligned} \tag{15}$$

where $p_1, p_2,$ and p_3 are defined by Equations (13). Similar reductions hold for other parameters as well, so that calculating the ratio (5) is in fact relatively easy.

A practical example of the method using simulated catch data is given next.

An analysis of simulated catch data

As an example, a data set consisting of catch data from 10 consecutive years, with 10 age groups was generated. The natural mortality rate was set to $M=0.10$. In each cohort the same structure of fishing mortalities was used, F taking the value $F=0.2$ in each age group. For a full-length cohort with 10 age classes, the number of fish in the first age class was set to $N=1\,000\,000$. The catch data and the sizes of other age classes were then calculated using expectations of the distributions (11).

Prior distributions need to be specified only for the first N in each cohort. A prior density constant on the interval $(0, 5 \times 10^6)$, was used. For the F s, a Gamma-distribution with a given mean and coefficient of

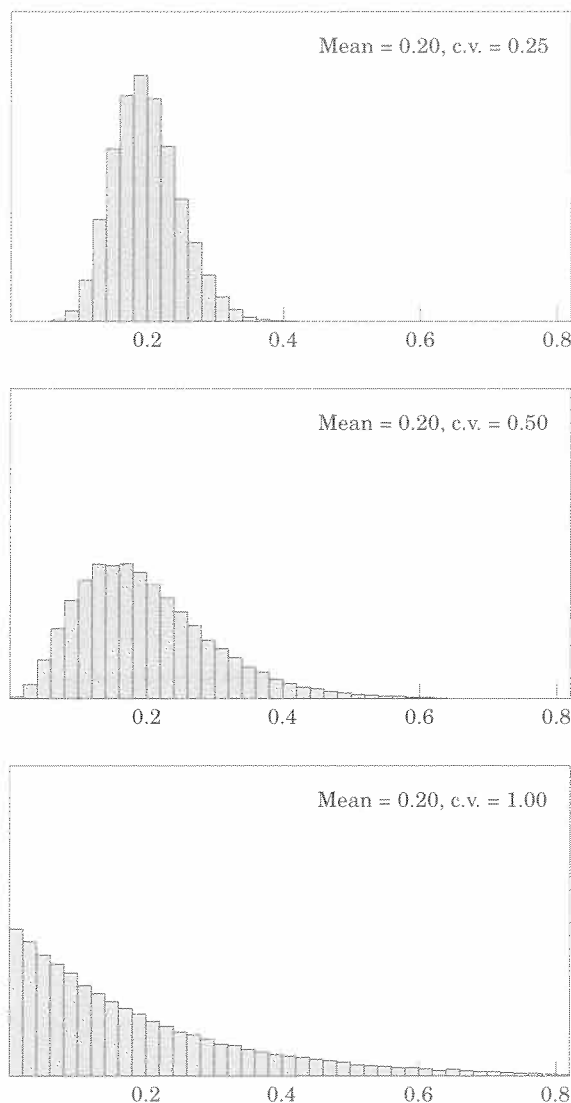


Figure 2. Gamma-distributions with a common mean 0.20 and with different coefficients of variation (c.v.).

variation (c.v.) was used. A pattern similar to deterministic VPA was chosen: the mean and the coefficient of variation of the terminal F s were set to mean=0.2 and c.v.=0.25 for all terminal age groups, whereas for the other age groups the values were set to mean=1.0 and c.v.=1.0. For comparison, Figure 2 shows three Gamma distributions with common mean 0.2 and coefficient of variation 0.25, 0.50, and 1.0, respectively.

In Section 2 we described a simple parameter-updating scheme in which a new value was proposed for each variable in turn, and the proposal value was either accepted or rejected depending on the values of the prior density and the likelihood for the proposed and the old parameter values. However, during the analysis of the

data it soon became evident that this method of updating parameters was not practical. The reason is that given M and the catch data C , the values of N and F are strongly correlated. If new values for N and F are proposed independently of each other, the likelihood of the proposals is usually very small and the algorithm's ability to explore the parameter space is poor. Instead of updating each parameter value in turn, the following procedure was thus used:

Step 1: For a given cohort, a new proposal value for N_1 was generated. Starting from this value, the values of N and F for the rest of the cohort, i.e. F_1 , $\{N_2, F_2\}$, ..., $\{N_K, F_K\}$, and N_{K+1} were then calculated as follows: When N_1 is given, it follows from the multinomial distribution (12), that the expected catch is:

$$E(C_1) = (1 - e^{-M - F_1}) \frac{F_1}{M + F_1}.$$

The value of F_1 was now adjusted so that the calculated expected catch was equal to the actual catch C_1 . The value of N_2 was then set by calculating the expected value of its conditional distribution, given N_1 , M , and F_1 [cf. Equation (11i)], and adding a small random perturbation to this value. The random perturbation was obtained by approximating the binomial distribution (11i) with a normal distribution with the same expected value and a variance 0.1 times the variance of the binomial distribution. A random value from this normal distribution was then added to N_2 . For this N_2 , the value of F_2 was then determined similarly as the value of F_1 before. The values of N_3 and F_3 were determined in a similar manner, etc., ending finally at N_{K+1} .

The likelihood of these new values of $\{N_1, F_1\}$, ..., $\{N_K, F_K\}$, and N_{K+1} was then compared to the likelihood of the original cohort and the new values were either accepted or rejected according to Equation (5).

Step 2: The values of F_1, \dots, F_K and N_1, \dots, N_{K+1} were updated separately, too. This was done to avoid the possibly unnecessarily tight coupling of values of N and F which might result from the use of Step 1 only. For example, the separate updating of the value of F_1 , say, was done as follows. Denote the current value of F_1 by F . A new proposed value F' was calculated as $F' = F + u$, where the value of u was sampled from a uniform distribution on the interval $(-s, s)$. The value of s was the current spread of the proposal distribution of F . If the resulting value of F' was negative, an absolute value of F' was taken as the proposal value. The acceptance or rejection of the proposed value was then done using Equation (5). This method was used to generate separate proposal values for the N s, too.

During the estimation the proportion of accepted proposals was automatically monitored as follows. For

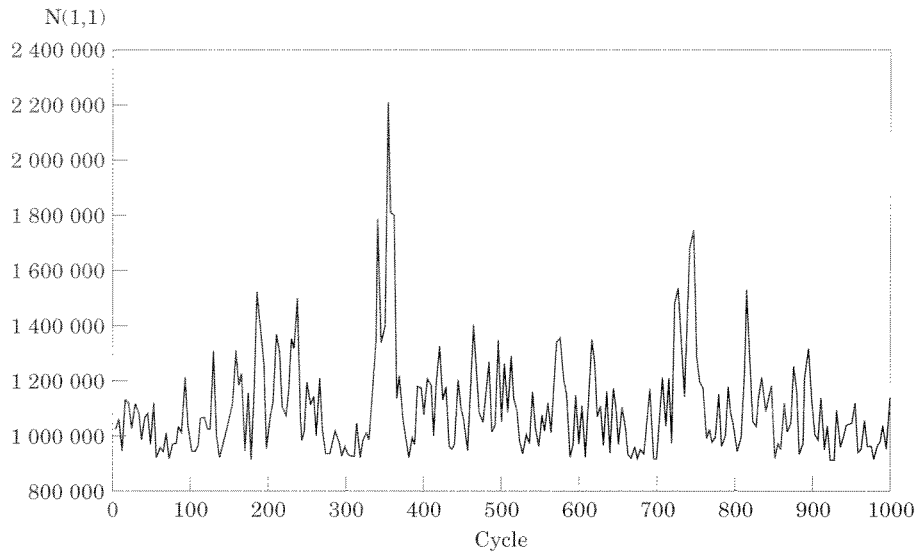


Figure 3. A sequence of values for $N(1,1)$ obtained from the first 1000 cycles of the estimation algorithm.

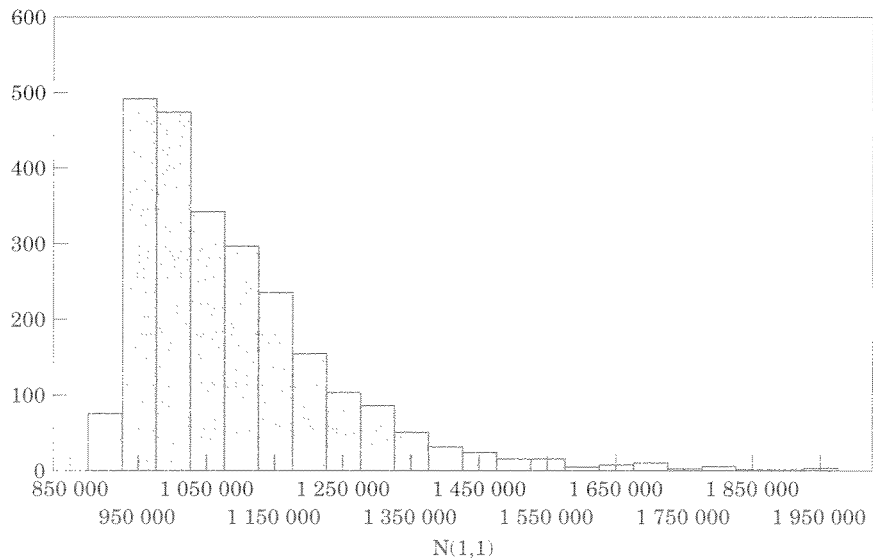


Figure 4. Histogram based on 2500 simulated values of $N(1,1)$, giving an approximate posterior distribution for the parameter.

each consecutive 100 proposed values for the parameters, the number of accepted proposals for each parameter was counted. If less than 20 proposed values for a parameter were accepted, the spread of the proposal distribution for that parameter was decreased. If more than 40 proposals were accepted, the spread of the proposal distribution was increased. This made the proportion of accepted proposals to remain in the range from 20 to 40% for all parameters.

The generated sequences of parameter values give as a result the approximate posterior distributions. Figure 3 shows a series of generated values for the size of the first

age group of the diagonal cohort $(1, 1), \dots, (10, 10)$. The resulting posterior distribution is shown in Figure 4. The length of the burn-in period was determined by graphically examining the sequences of generated parameter values. When the curves for different parameters seemed to be "regular" without any clear tendencies towards lower or higher values, the length of the burn-in period (10 000 cycles) was considered as sufficient. From the next 10 000 generated values, only every 4th was sampled. For a sample of this size, this resulted in somewhat smoother distributions than taking directly the first 2500 consecutive observations. In Figure 5 the values of

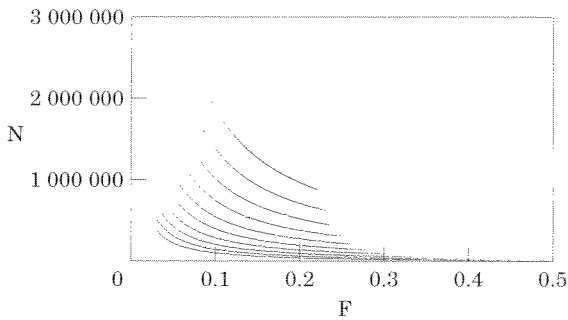


Figure 5. A scatter plot of 2500 simulated values for the pairs $[N(1,1), F(1,1)], \dots, [N(10,10), F(10,10)]$. Each curve displays the set of values for a particular pair $[N(i,i), F(i,i)]$. The figure shows clearly how tightly the values of $N(a,y)$ and $F(a,y)$ for a given age group depend on each other.

$(N_{11}, F_{11}), \dots, (N_{10,10}, F_{10,10})$ for the diagonal cohort are plotted simultaneously, showing the strong mutual dependence of N_{ii} and F_{ii} .

Generally, for a given cohort the estimated fishing mortality probability distributions are quite narrow in the younger age groups and become wider in the older age groups (Fig. 6a). This is the same phenomenon as convergence in point estimation VPA. The posterior distribution of F for the first age group of a cohort becomes narrower when the follow-up time increases (Fig. 6b). For a cohort for which there is catch data from only the first age group, the posterior distribution of F for that age group mainly reflects the prior distribution. Figure 6(c) shows the estimated distributions of $F_{1,10}, \dots, F_{10,10}$. As expected, the distribution of $F_{1,10}$ is similar to its prior distribution. For longer follow-up times, the terminal F s are slightly underestimated. This is probably due to the skewness of the distributions of the N s, as shown in Figure 4 for N_{11} . The skewness results from the fact that all values of N are restricted from below by the catch data, whereas the larger values are only constrained by the prior distribution. Furthermore, the variance of the posterior distribution of $F_{10,10}$ is larger than the variance of $F_{1,10}$. For longer follow-up times, the importance of the last age groups in the likelihood-ratio comparison $L(y|\theta')/L(y|\theta)$ in Equation (5) is smaller, so that more deviant proposals for F can be accepted.

To study the influence of terminal- F assumptions on the posterior distributions, the common mean 0.2 of the terminal- F 's prior distributions was multiplied by 0.5 and 2.0. As Figure 7(a) shows, the longer the follow-up time, the better the posterior distribution of F in the youngest age group approximates the distribution used in the generation of the data. This is not, however, true for the last age groups of the cohorts. According to Figure 7(b), the posterior distributions of the terminal- F s remain quite distinct in the three cases

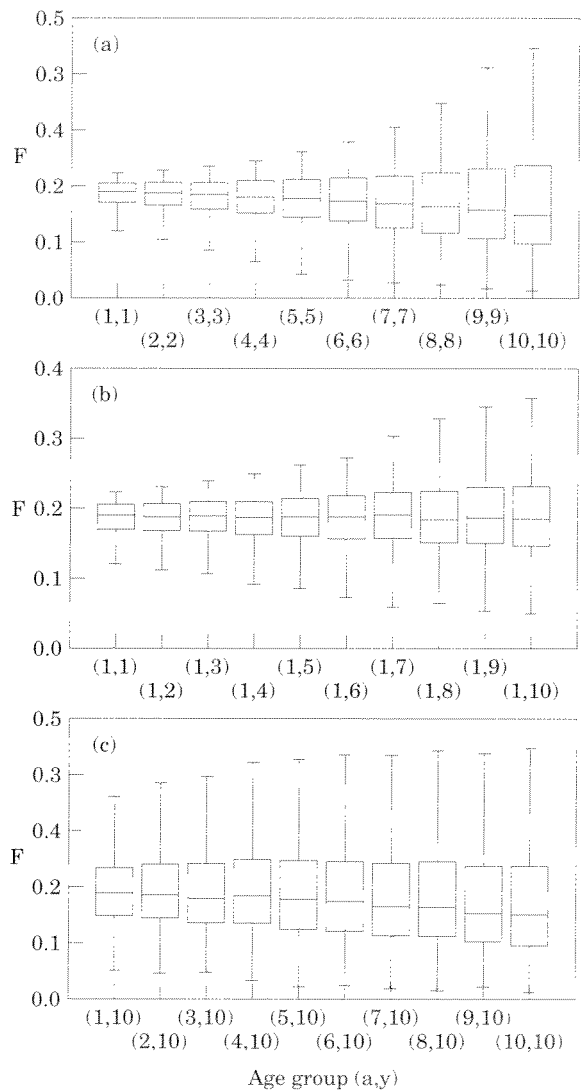


Figure 6. (a) Approximate posterior distributions for the fishing mortalities $F(1,1), \dots, F(10,10)$. The box contains 50% of the values and the horizontal line shows the median of the distribution. The vertical lines encompass all values which are not interpreted as outliers (a value is considered an outlier if its distance is more than 1.5 box-lengths from the 75th percentile upwards or 1.5 box-lengths from the 25th percentile downwards). (b) Posterior distributions of $F(1,1), \dots, F(1,10)$. $F(1,1)$ belongs to a cohort which has catch data for 10 age groups, whereas for $F(1,10)$ the value of $C(1,10)$ is the only catch data available. (c) Posterior distributions of the terminal fishing mortalities $F(1,10), \dots, F(10,10)$.

where the mean of the prior distribution has been multiplied by 0.5, 1.0, and 2.0, respectively.

The fact that the posterior distributions of F s for the older age groups of a cohort do not converge towards the true values affects also the predicted numbers of fish in the next year (year 11 in our model). Figure 8 displays

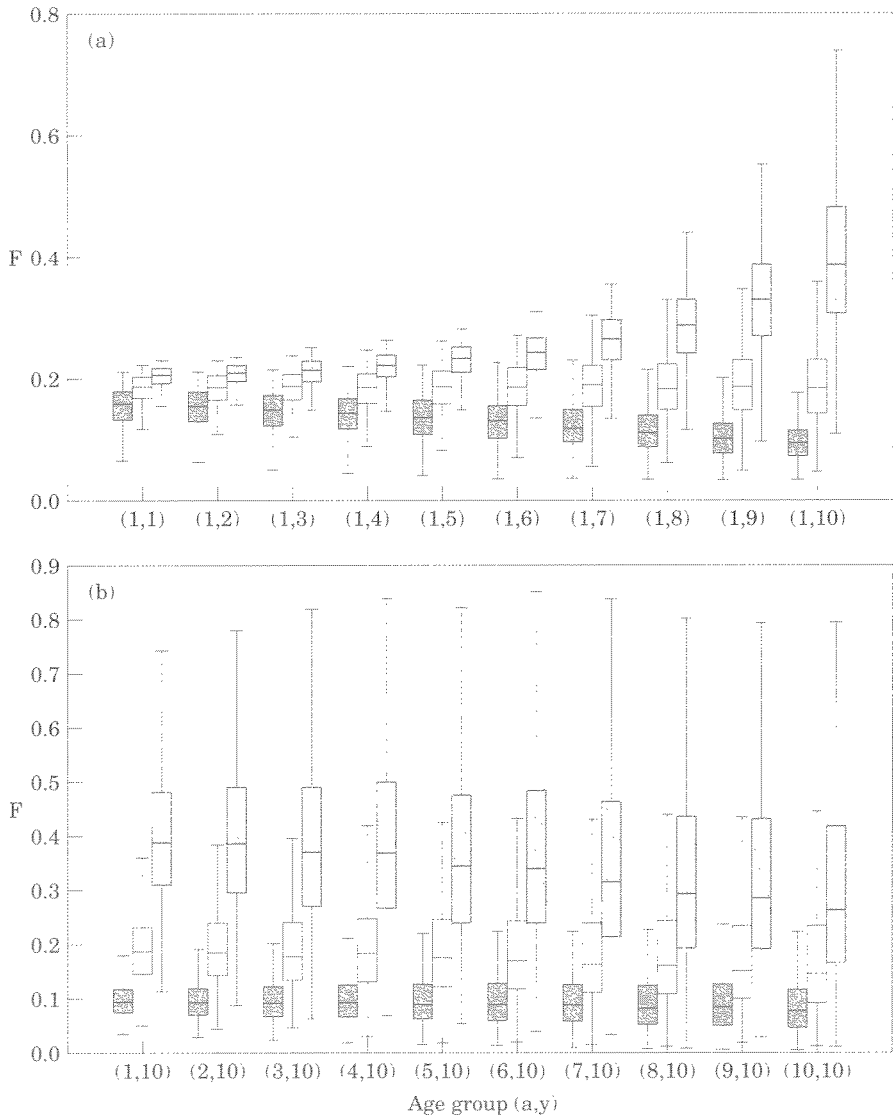


Figure 7. (a) Posterior distributions of $F(1,1)$, . . . , $F(1,10)$ for three different sets of terminal-F prior distributions. The set of boxes shaded \square correspond to a model in which the mean 0.2 of the terminal-F's prior distributions are multiplied by 0.5 (underestimated prior distributions of the terminal-Fs). The boxes shaded \square show the results when the mean is multiplied by 2.0 (overestimated terminal-F prior distributions). The set of unshaded boxes correspond to a model in which the means of the terminal-F's prior distributions have the correct value 0.2. (b) Posterior distributions of terminal-Fs for the same three models as in (a). 0.5 (\blacksquare); 1.0 (\square); 2.0 (\square).

the posterior distributions for $N(2, 11)$, . . . , $N(11, 11)$, showing how much incorrect assumptions regarding the values of terminal-Fs affect the predictions.

Discussion

The main benefit of the Bayesian modelling and the application of the Markov chain Monte Carlo method, as compared to deterministic virtual population analysis, is that instead of point estimates, posterior distribu-

tions of plausible parameter values are obtained. They are important signals about the quality of the scientific information. In addition to one-dimensional marginal distributions, we can also produce joint posterior distributions of any combination of variables we like (cf. Fig. 5). If the variables are correlated, their effects on the posterior distributions are also accounted for.

In Equation (15) the acceptance probability of a new proposal F'_{22} was calculated. The formula might suggest that the posterior distribution of F_{22} depends only on

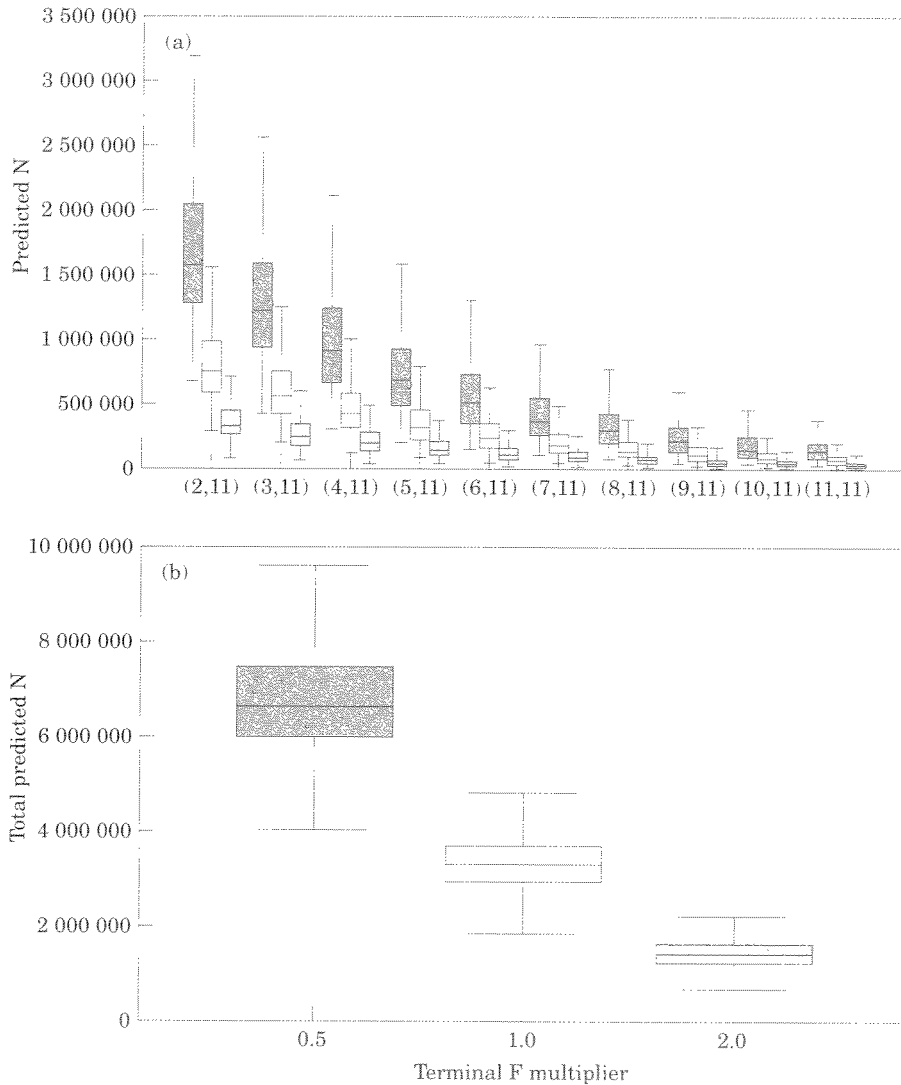


Figure 8. (a) Posterior distributions for the predicted values of $N(2,11), \dots, N(11,11)$ for three different terminal-F assumptions (see Figure 7(a) for an account of the models used). (b) Posterior distributions for the total predicted size of the population for three different terminal-F assumptions. 0.5 (□); 1.0 (▨); 2.0 (■).

those quantities which appear explicitly in (15). However, this was not the case; consider, for example, the prior distribution of N_{11} . It influences the acceptance probability of new proposals for N_{11} . A given value of N_{11} further influences the acceptance of new values of N_{22} , which directly influences the acceptance of the proposals for F_{22} . In fact, the posterior distribution of F_{22} depends on the prior distribution of $\pi(N_{11})$ of N_{11} and on all the prior distributions $\pi(F_{11}), \dots, \pi(F_{KK})$. This is because the values of N and F for a given cohort depend on each other as specified by the stochastic model dynamics (11). However, in our model separate cohorts do not interact with each other, which influences our ability to make correct predictions concerning future

population sizes. As Figure 8 shows, the predictions are quite sensitive to the assumed terminal-F distributions. One way to make the predictions more robust against erroneous terminal-F assumptions is to tie together the estimation of different cohorts. If, for example, there is some yearly effort data $E_y, y=1, \dots, Y$ available, the fishing mortalities F_{ya} can be modelled by:

$$F_{ya} = q_a E_y.$$

where q_a is an age-specific vulnerability. Estimation of the fishing mortalities F_{ya} is now replaced by the estimation of the coefficients q_a , depending for example on the selectivity values of the gear used and on the growth

rates of individuals in the population. Instead of specifying the prior distributions $\pi(F_{ay})$, we now need the distributions $\pi(q_a)$. Note that in this model the cohorts can not be estimated independently of each other but the data and prior distributions in one cohort influence the estimation of all other cohorts as well.

The simulated catch data were generated using expected values in Equations (11). This resulted in equal catches in each cohort. A random catch can be generated by sampling the values of $N_{a+1,y+1}$ and C_{ay} from the binomial distributions (11). However, the resulting variation in the catch was so small that estimates did not differ much from those obtained by using expected values. This is due to the small spread of the binomial distributions. Suppose, for example, that in a cohort $\{N_1, N_2, \dots, N_{K+1}\}$, a new value for N_2 is proposed. Since all other parameters are considered as constants, the distribution of N_2 is:

$$N_2 \sim \text{Bin}(N_1, p).$$

where

$$p = e^{-M - F_1}.$$

The standard deviation of this distribution, $\sqrt{N_1 p(1-p)}$, which for $N_1 = 1\,000\,000$, $M = 0.1$, and $F_1 = 0.2$, say, is 438. Given N_1 and F_1 , the conditional distribution of N_2 thus has a small standard deviation. Similarly, given N_1 and C_1 , the conditional distribution of F_1 has a small variance, too. One possibility to obtain more realistic variation in the catch data would be to add random variation to the natural mortality. Instead of the binomial model (11ii), the fishing process could also be modelled differently to allow more variation in the catch.

Extensive testing was in fact done with a larger model, containing an additional level of hierarchy and thereby "more randomness". In this extension the probabilities (p_1, p_2, p_3) of surviving, dying of a natural cause, or being caught by fishermen, were viewed as random variables drawn from a Dirichlet distribution with (conditional) expected values given by the right hand sides of (13i-iii). This was thought to make the likelihood surface smoother and thereby both reduce the near-singular behaviour of the joint posterior as illustrated in Figure 5 and, at the same time, stabilize the numerical simulation algorithm. However, the overall experience from this additional modelling exercise was that it was not really worth the extra effort.

The practical usefulness of the method depends somewhat on the computational efficiency of the algorithm, especially during meetings of international working groups, where simulations may have to be re-run several times under different assumptions. In this respect, the techniques based on Bayesian belief networks, suggested

by Varis and Kuikka (1997; based on discretized probability distributions and spreadsheet software) are computationally much more effective, allowing for interactive parameter estimation and selection of models. The mathematical background of the use of belief networks in parameter estimation is not analysed in complete details, but it seems to be a promising tool for resource management modelling (see, for example, Varis, 1995 for an environmental management example).

More detailed studies are needed to be able to compare these two methodologies.

McAllister and Iannelli (1997) use the Sampling/Importance Resampling Algorithm (Rubin, 1987, 1988) to calculate posterior distributions. In the SIR algorithm an independent sample of parameter vectors θ is taken from a distribution defined by an "importance function" $h(\theta)$. After calculating a certain "importance ratio" for each sampled parameter vector, an estimate of the posterior distribution $\pi(\theta|\text{data})$ is obtained. If posterior distributions of known functions of θ are needed, then a new random sample from the estimated posterior distribution of θ is drawn. In contrast, in the MCMC method applied in our study the posterior distributions of functions of θ are constructed at the same time as that for θ . In their study McAllister and Iannelli found the SIR algorithm to run several times faster than a corresponding MCMC algorithm (McAllister and Iannelli, 1996). We have not applied the SIR algorithm in our study but the 8 min required on a 120 MHz Pentium for running 10 000 iterations using a 10×10 matrix of catch data seems an acceptable time for practical calculations.

There are a number of ways to change the calculation procedure presented here. The preferred procedure depends on the stock and fishery considered. Additional biomass data (CPUE, echo sounding) would improve the estimates, as well as the use of F-structure (selectivity knowledge) of the stock. Such data is needed for the estimates of the terminal stock, which are used in predictions and are, from the point of view of fisheries management, the most important quantities to be estimated.

In this study we have assumed that the input data (total catches, age distribution in the catches) are known precisely, however, this is usually not the case. Misreporting of the catches is common in many fisheries managed by TAC. Especially in cases where TAC is strongly restrictive, the risk for misreporting and unreported discarding increases. The inclusion of input data uncertainties is an important step in the further improvement of this methodology.

In the analysis we have assumed that the natural mortality rate M is known. Schnute and Richards (1995) use state space models and describe circumstances in which it is possible to estimate the value of M . In our current application the effect of predation on the values of M could be studied in a similar way as is done in

traditional multi-species VPA. Another possibility would be to use only the biomass of the predator stocks, to see whether the catch-age data together with predator biomass information would allow for the estimation of natural mortality. Reliable effort and selectivity data of the gears used would further improve the possibilities of making reliable inferences on the natural mortality.

When essentially only the prior distributions of the terminal-Fs are somewhat informative, the value of M cannot be estimated from the catch data. To study the effect of uncertainty in M, we might also treat M as an unknown parameter and specify a prior distribution for it. The posterior distributions of the other parameters depend on the assumed prior distribution of M, thus giving a method for assessing the effect of uncertainty in M. This approach was not quite satisfactory, however. Even though the catch data give essentially no information on the value of M, the sampled values of M in the MCMC-simulation shifted towards zero, the shift being larger with non-informative priors. One possible explanation to this is the following. According to Equation (3), given the catch data C and the model parameters θ , the posterior density $\pi(\theta|C)$ of the parameters is:

$$\pi(\theta|C) \propto \pi(\theta)L(C|\theta).$$

When a new value M' for M is proposed using Equation (5), the likelihood $L(C|\theta')$ is a product of multinomial densities defined by Equations (11) and (12). For a given C_{ay} , the multinomial density is:

$$\binom{N_{a-1,y-1}}{N_{ay}D_{ay}C_{ay}} p_1^{N_{ay}} p_2^{D_{ay}} p_3^{C_{ay}} \quad (16)$$

where p_1 , p_2 and p_3 depend on M and F_{ay} as defined by Equations (13). Let us consider two different values M_1 and M_2 , where $M_1 < M_2$. When the other parameters F_{ay} , N_{ay} and $N_{a+1,y+1}$ are allowed to move towards values which maximize (16), it is found that the maximum value of (16) is greater for M_1 than it is for M_2 . This may cause the sampled values of M in the MCMC-simulation to shift towards zero. Unless stricter prior information is incorporated into the estimation, this tendency has too much influence on the results. One alternative is to replace the single value of M with age- and year-dependent quantities M_{ay} . This gives more weight to the prior density as compared to the likelihood, but some tendency towards small values of M_{ay} was still seen in the resulting posterior distributions.

In the multinomial model used in this study it was assumed that each individual fish was sampled independently from the entire population. This was done to keep the model as simple as possible, but in many cases it is an unrealistic assumption, due to age and size selective properties of the fishing gear and fishing activities, non-random distribution of individual fish in the ocean,

etc. If groups of same aged fishes are sampled independently rather than individual fishes of each age, the estimates of F and N can be biased. It is thus important in real applications to check the model assumptions and adjust the model to account for the differences, if necessary.

Any additional information can easily be incorporated into the model. All that is needed is a probability model specifying the relationship of the new variables to the other variables in the model, so that the values of the likelihood functions can be calculated. For example, we could include in the model relevant information concerning different age groups. Especially for the youngest age groups in the terminal stock environmental data and S/R models might be used as an extra source of information. The value of the new data can then be assessed by examining the resulting new posterior probability distributions.

References

- Besag, J., Green, P., Higdon, D., and Mengersen, K. 1995. Bayesian computation and stochastic systems. *Statistical Science*, 10: 3-66.
- Gamerman, D. 1997. Markov Chain Monte Carlo. Stochastic simulation for Bayesian inference. Chapman & Hall, London.
- Gelman, A. 1996. Inference and monitoring convergence. *In* Markov Chain Monte Carlo in practice (cited in Gilks *et al.*, 1996, pp. 131-143).
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 1995. Bayesian data analysis. Chapman & Hall, London.
- Gilks, W. R., Richardson, S., Spiegelhalter, D. J. (Eds) 1996. Markov Chain Monte Carlo in practice. Chapman & Hall, London.
- Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57: 97-109.
- Hilborn, R., and Walters, C. J. 1992. Quantitative fisheries stock assessment. Choice, dynamics and uncertainty. Chapman and Hall, New York.
- Kuikka, S., and Varis, O. 1991. Probabilistic assessment of TAC based fisheries management of Baltic salmon stocks. ICES C.M. 1991/M:30.
- McAllister, M. K., and Janelli, J. N. 1997. Bayesian stock assessment using catch-age data and the sampling/importance resampling algorithm. *Canadian Journal of Fisheries and Aquatic Science*, 54: 284-300.
- McAllister, M. K., Pikitch, E. K., Punt, A. E., and Hilborn, R. 1994. A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. *Canadian Journal of Fisheries and Aquatic Science*, 51: 2673-2687.
- Pope, J. G. 1972. An investigation of the accuracy of virtual population analysis using cohort analysis. *ICNAF Research Bulletin*, 9: 65-74.
- Punt, A. E., and Hilborn, R. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. *Reviews in Fish Biology and Fisheries*, 7: 1-29.
- Rubin, D. B. 1987. Comment on 'The calculation of posterior distributions by data augmentation'. *Journal of the American Statistical Association*, 82: 543-546.

- Rubin, D. B. 1988. Using the SIR algorithm to simulate posterior distributions. *In* Bayesian Statistics 3: Proceedings of the Third Valencia International Meeting, June 1-5, 1987. Ed. by J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. M. Smith. Clarendon Press, Oxford.
- Schnute, J. T., and Richards, L. J. 1995. The influence of error on population estimates from catch-age models. *Canadian Journal of Fisheries and Aquatic Science*, 52: 2063-2077.
- Varis, O., and Kuikka, S. 1997. Analysis of uncertainty in Resource Management: Joint Use of Multiple Environmental Assessment Models by a Bayesian Meta-Model. *Ecological Modelling*, 102: 341-351.
- Varis, O. 1995. A belief network approach to optimization and parameter estimation in resource and environmental management models. WP-95-11. International Institute for Applied Systems Analysis, Laxenburg, Austria. 38+3 pp.
- Walters, C., and Ludwig, D. 1994. Calculation of Bayes posterior probability distributions for key population parameters. *Canadian Journal of Fisheries and Aquatic Science*, 51: 713-722.
- Walters, C. J., and Punt, A. E. 1994. Placing odds on sustainable catch using virtual population analysis and survey data. *Canadian Journal of Fisheries and Aquatic Science*, 51: 946-958.