ICES Journal of Marine Science, 55: 1031–1060. 1998
Article No. jm980425

Bayesian stock assessment: a review and example application using the logistic model

M. K. McAllister and G. P. Kirkwood

McAllister, M. K. and Kirkwood, G. P. 1998. Bayesian stock assessment: a review and example application using the logistic model. – ICES Journal of Marine Science, 55: 1031–1060.

Bayesian statistical methods have recently been combined with conventional methods for fisheries stock assessment (e.g. catch-age analysis) to provide a conceptually elegant approach for providing fishery management advice under uncertainty. Uncertainties in the advice provided can be conveyed using posterior probability distributions (or "posteriors") for the potential outcomes of each policy option. Posteriors can be estimated using data (e.g. catch-age data and relative abundance indices) for the fish population of interest and prior probability distributions for population model parameters (e.g. stock-recruit function parameters) based on data from similar fish populations. Despite growing interest, Bayesian methods remain accessible to relatively few. To increase the accessibility of these methods, the conceptual basis for Bayesian statistical estimation is reviewed and set in the context of fisheries stock assessment. The use of Bayesian methods is illustrated by fitting a logistic model to relative abundance indices for Namibian hake (*Merluccius capensis* and *M. paradoxus*) and presenting a decision analysis of alternative harvest policy options. Some alternative approaches are outlined for constructing prior and posterior probability distributions and some recent applications in fisheries. Some of the problems that can be encountered while implementing Bayesian methods are also discussed.

Key words: Bayesian, decision analysis, policy evaluation, prior distribution, posterior distribution, uncertainty, stock assessment, fishery management.

Received 19 February 1997; accepted 3 June 1998.

M. K. McAllister and G. P. Kirkwood: Renewable Resources Assessment Group, Centre for Environmental Technology, T. H. Huxley School of Environment, Earth Science & Engineering, Imperial College of Science, Technology, and Medicine, 8 Prince's Gardens, London, SW7 1NA, UK. Correspondence to M. K. McAllister: tel: +44 171 594 9330; fax: +44 171 589 5319; email: m.mcallister@ic.ac.uk

Introduction

Until recently, policy advice to fishery managers from stock assessments has frequently been based on single point estimates of biological quantities; uncertainties in fisheries stock assessments over stock size and productivity and the potential biological and economic consequences of alternative policy (e.g. total allowable catch) options were seldom conveyed and taken seriously. Recognition of the potential disadvantages of ignoring uncertainty in policy advice, for example, increased risks of resource collapse and underutilization, has led to the development of a variety of approaches to accounting for uncertainties in stock assessment and in providing policy advice (Walters and Hilborn, 1976; Bergh and Butterworth, 1987; Francis, 1992; Restrepo *et al.*, 1992; Hilborn and Walters, 1992; Smith *et al.*, 1993; Walters and Ludwig, 1994). Among these, the Bayesian alternative offers a simple and conceptually elegant approach to fisheries stock assessment. Bayesian methods have recently been developed to rigorously incorporate expert judgment and inferences into conventional stock assessment methods using data from similar fish populations; these methods have also been used to convey uncertainties in policy advice to decision makers.

In this paper the Bayesian approach to fisheries stock assessment is reviewed and a detailed example application provided in order to increase the accessibility of Bayesian methods to a wider audience. In the introductory sections that follow, the basic problems of decision making under uncertainty faced by fishery managers are briefly outlined. Three commonly used approaches to

1054–3139/98/061031+30 $30.00/0

dealing with uncertainties in stock assessment and some potential problems with their application are then discussed. Some of the potential advantages of the Bayesian approach to fisheries stock assessment are then outlined.

Providing quantitative advice for decision making in fishery management: the problem

In the making of fisheries policy decisions, fishery managers want at least two things: high biological yields and high rates of employment. They also want to avoid bad things happening such as a fishery closure due to stock collapse. In fishery management, risk is often defined as the probability of an undesirable event (Francis, 1992; McAllister *et al.*, 1994). A common difficulty is that policy options with the highest yields also have the highest risks of stock depletion. The problem for fishery managers then is to choose a policy that has the highest expected benefits but also acceptably low risks.

The solution to this policy problem is easy if decision makers are certain about the outcome of each policy option, or if they know that one policy is best across all possible states of nature. In such cases, the decision makers can easily choose the policy that best meets their goals. However, this situation is almost never the case in fishery management. Most often, the solution is difficult because there are normally several competing alternative hypotheses for the true state of nature, the best policy option depends strongly on which state of nature is the true one; and there is commonly large uncertainty about which hypothesis best describes the true state of nature.

To address this problem, stock assessment scientists should provide fishery managers with quantitative predictions about the potential biological and economic consequences of alternative policy options. However, to many, stock assessment simply means the estimation of parameters in mathematical models, something quite different from the provision of management advice. In contrast, the authors are following the viewpoint of Hilborn and Walters (1992): the whole point of stock assessment is to provide advice to fishery managers and the provision of quantitative management advice should be an integral part of fisheries stock assessment. Stock assessment can thus be divided into two interrelated quantitative activities: estimation and policy evaluation. Estimation entails the construction of mathematical models for policy evaluation and the analysis of fishery, research survey, and other data to estimate model parameters. Policy evaluation entails the quantitative evaluation of the potential consequences of alternative management actions. The authors believe that Bayesian statistical methods and decision theory (Berger, 1985; Gelman *et al.*, 1995) provide a particularly suitable approach to providing such advice for decision making under uncertainty in fishery management. To indicate why this is so the authors begin by discussing below three different non-Bayesian approaches to dealing with uncertainty in the provision of quantitative advice for fishery management.

Alternative approaches to dealing with uncertainty in stock assessment

Before proceeding the authors would like to emphasize that despite their advocacy of the Bayesian approach and subsequent discussion of the potential problems with the following three non-Bayesian approaches, the latter are still found to be worthwhile and the authors continue to apply them. The authors also acknowledge that Bayesian methods are not without their own troublesome problems and these are discussed later on. The various problems with non-Bayesian approaches that are mentioned below were part of the motivation for the introduction and development of Bayesian methods for stock assessment.

One approach to uncertainty in stock assessment is to conduct a "sensitivity analysis" (e.g. Goodyear, 1995). In almost all stock assessments, some input parameters (e.g. the rate of natural mortality) are each fixed at some value and treated as known when they are actually uncertain. The effect of uncertainty in such parameters is evaluated by re-running the entire stock assessment using alternative values for such input parameters. This approach may fail to reveal disastrous outcomes that could result under combinations of parameters that have not been tested. Furthermore, if stock assessment results were highly sensitive to a number of different parameters, presentation of such results to managers can be very cumbersome without some form of integration of results.

A second approach is to estimate confidence bounds or intervals for quantities of interest (e.g. Mohn, 1993). Confidence bounds can be calculated using maximum likelihood estimates of parameters and asymptotic assumptions or by using bootstrapping methods (Mohn, 1993). These intervals can be difficult to interpret and could be misused by decision makers if they acted upon either the most pessimistic or optimistic side of the interval (Hilborn and Walters, 1992).

A third approach is to estimate the sampling distributions of estimated quantities of interest (e.g. stock biomass) using jackknife or bootstrap procedures, and to use these distributions as proxies for uncertainty (Francis, 1992; Restrepo *et al.*, 1992; Smith *et al.*, 1993). There exists a large variety of such procedures (Smith *et al.*, 1993). The simplest of these (e.g. the unconditional non-parametric bootstrap) randomly resamples the data with replacement and re-estimates the quantities of interest from each resampled data set. These steps are repeated many times to produce a sampling distribution of estimates of the quantities of interest.

However, the interpretations of such distributions and their use in the evaluation of alternative harvest policy options has provoked controversy (Francis, 1992; Hilborn *et al.*, 1993b; Restrepo *et al.*, 1992; Walters, 1993; Cordue and Francis, 1994). For example, decision analysis methods (Raiffa, 1968; Berger, 1985), which can be used to evaluate the consequences of alternative policy options, typically require probability distributions for alternative hypotheses. The sampling distributions for estimated quantities that are produced by methods such as bootstrapping are not the same thing (Gelman *et al.*, 1995, p. 110). However, such sampling distributions for estimated management quantities have often been applied and interpreted as if they were probability distributions for alternative hypotheses.

A Bayesian approach to stock assessment

An alternative to these three approaches is the use of Bayesian statistical methods (Walters and Hilborn, 1976; Thompson, 1992; McAllister *et al.*, 1994; Walters and Ludwig, 1994; Raftery *et al.*, 1995a; Kinas, 1996; Punt and Hilborn, 1997; McAllister and Ianelli, 1997). Bayesian methods have been advocated for stock assessment for several reasons. Firstly, data for many fish stocks often contain little information about certain key input parameters to population dynamics models, e.g. parameters relating to the slope of a stock-recruit relationship. In conventional approaches to stock assessment, uncertainties in such parameters are often ignored and point estimates or assumed values are used instead. However, values for such parameters may be similar among ecologically and taxonomically similar populations and could be incorporated into Bayesian stock assessment in the form of prior probability distributions (priors) (Gelman *et al.*, 1995).

Secondly, data on certain key aspects of population biology and the fishery such as on stock structure and bycatch in other fisheries are often missing, i.e. there may be "gaps" in the data. Under conventional approaches to stock assessment, fixed assumptions are often made about the missing data that lead to simplifications in the analysis and the uncertainties resulting from data gaps are ignored (Goodyear, 1995). Under the Bayesian approach, these uncertainties can be incorporated in the evaluation as alternative hypotheses that may consist of alternative values for key parameters such as bycatch rates or structurally different models for fish population dynamics (Punt, 1993b).

Thirdly, Bayesian methods combine fishery data and prior information to calculate posterior probabilities for alternative hypotheses, for example, about structurally different models for population dynamics and management quantities such as the current biomass of a fish population (Sainsbury, 1988). This is particularly useful when the identification of the best policy depends strongly on the hypothesis that is assumed to be true and there is uncertainty about the alternative hypotheses. The calculation of posterior probabilities for alternative hypotheses provides a simple, effective means to indicate the empirical support for each hypothesis and to convey uncertainties in estimated quantities (Hilborn *et al.*, 1993a). The presentation of results is also simplified because Bayesian methods typically integrate over model parameters to produce marginal probability distributions for management quantities. For example, a marginal probability distribution for the maximum sustainable yield can be computed by integrating over all combinations of parameter values in the joint posterior distribution of model input parameters (e.g. r and K in the logistic model; see Equation (3) and the example below). However, evaluations of the sensitivity of model results to alternative combinations of the data and prior distributions for model parameters are usually appropriate.

Fourthly, the Bayesian approach provides a conceptually straightforward statistical procedure for providing management advice under uncertainty that is commonly referred to as "decision analysis" (Raiffa, 1968; Walters and Hilborn, 1976; Berger, 1985). Decision analysis procedures allow the computation of the consequences of alternative policy options, and explicitly account for uncertainty about alternative hypotheses on model structure (e.g. population dynamics). Bayesian estimation provides the posterior probabilities for the alternative hypotheses to indicate the support given to each by the best available information and data. The consequences of each policy option are calculated by weighing the consequences under each hypothesis by the posterior probability for the hypothesis. The results of the decision analysis (e.g. the expected consequences for each policy option and indications of the uncertainties) are then each presented to decision makers in the form of probability distributions and decision tables (see example below). They can thereby account for the expected consequences of each policy option under some of the alternative hypotheses, the empirical support for each hypothesis, and the expected consequences of each policy integrated over all hypotheses (Hilborn *et al.*, 1993a).

Aims of this Review

Hilborn *et al.* (1994) and Walters and Ludwig (1994) introduce the rudimentary concepts of Bayesian estimation for simple applications in fisheries stock assessment. Punt and Hilborn (1997) provide a review of the Bayesian approach to stock assessment and decision analysis and present some sophisticated age-structured applications to exemplify the approach. However, applications of Bayesian methods to fisheries problems remain relatively scarce. This is partly because fishery scientists require simple, clear examples that they can implement on their own in order to learn new methods; yet there exists a paucity of such examples in the fisheries literature. Most of the examples provided are complicated and not easily implemented because they use complicated models (e.g. age-structured rather than simpler biomass dynamic models) and sophisticated estimation algorithms (e.g. Bergh and Butterworth, 1987; McAllister *et al.*, 1994; Kinas, 1996; Punt and Hilborn, 1997); other examples are overly simplistic, using equilibrium-based models or assuming that the values of parameters that would normally be considered to be uncertain are instead known without error (Hilborn *et al.*, 1994; Hoenig *et al.*, 1994; Walters and Ludwig, 1994; Walters and Punt, 1994).

In order to stimulate further interest and increase the accessibility of Bayesian methods, this paper provides an overview of the Bayesian approach to fisheries stock assessment and a detailed example application using a non-equilibrium, logistic model. The review and example are intended as an introduction to the implementation of Bayesian methods for fisheries scientists unfamiliar with the methods. The paper also includes sections reviewing the main ideas and methods in Bayesian stock assessment for those seriously interested in applying the methods. Despite intending this paper primarily for those not familiar with Bayesian methods, these latter sections require at least an intermediate knowledge of Bayesian theory. To make the reading of this paper easier, at the end of the next paragraph the more challenging sections are indicated so that readers not so familiar with Bayesian methods may, if they choose, avoid these sections and perhaps come back to them later on.

This paper continues with a reformulation of the basic steps in a Bayesian stock assessment in which the consequences of alternative fishery management policy options are evaluated. This is to provide added clarity and completeness to the steps outlined in earlier works (Bergh and Butterworth, 1987; McAllister *et al.*, 1994; Punt and Hilborn, 1997; McAllister and Ianelli, 1997; McAllister and Pikitch, 1997). The Bayesian concept of probability is then briefly outlined and how it differs from the conventional frequentist concept is indicated. Thereby the authors indicate why Bayesian probability is well-suited to account for uncertainty in stock assessment. The calculation of Bayesian posterior probabilities and their application in stock assessment is illustrated by fitting a logistic model to relative abundance data for Namibian hake (*Merlucius capensis* and *M. paradoxus*). This example is used to illustrate in detail how decision tables can help fisheries managers to account for uncertainty in decision making. The authors also provide recommendations about how to construct and partition parameter space in decision tables. The example is also presented so that those interested can repeat the calculations. The authors discuss some recent applications of Bayesian methods in fisheries stock assessment, and summarize some guidelines for model selection and parameterization and the construction of prior probability distributions for Bayesian fisheries stock assessment. Finally, some problems are outlined that may be encountered in implementing Bayesian methods. Sections that are technically advanced include those on methods for numerical integration, applications of Bayesian methods, and Appendices 1–3. All other sections should be easily understood by people not familiar with Bayesian methods. Understanding of the example application requires a numerical proficiency common to most stock assessment biologists.

A framework for Bayesian stock assessment

One of the main goals of fisheries stock assessment is to evaluate the potential consequences of alternative management options (Hilborn and Walters, 1992). Punt and Hilborn (1997) provide steps to indicate how this is done using the Bayesian statistical approach. Below the authors reformulate these steps by reordering them, adding an additional step (step 2), and providing clearer justifications for some of the steps. This is to improve upon their logical sequence, consistency, and completeness so that they are more easily understood by readers unfamiliar with Bayesian theory. Note that some of the steps are not uniquely Bayesian (e.g. steps 1 and 2) but are still part of Bayesian decision analysis (Berger, 1985).

(1) Identify each alternative management action that could be taken

The set of alternative actions that could be taken is usually identified through discussions among fishery managers, scientists, and industry members. The actions considered are often relatively simple to evaluate and implement, for example, setting alternative series of fixed catch quotas for the next few years. In some situations, for example, the International Whaling Commission, more sophisticated management policy options are considered that involve implementing decision rules for setting catch quotas. These rules have sometimes used as inputs, annual estimates of stock biomass and the level of uncertainty in population dynamic model parameters. This latter alternative can be considerably more complicated to evaluate because it can require the simulation of stock assessment data, the stock assessment procedure, and the application of a decision rule to the resulting estimate of, e.g. stock biomass, in each future year. A shortcut stock assessment procedure is typically simulated to make these computations tractable (Punt, 1993a; McAllister *et al.*, 1994; McAllister and Pikitch, 1997).

(2) Specify the indices of policy performance
The most desirable policy option for fishery managers is often the one that has the largest expected economic benefits and has an acceptably high chance of avoiding undesirable biological and economic outcomes. The specification of indices of policy performance indices, however, is often arrived at by discussions among fishery managers, scientists, and industry members. It is usually desirable to achieve a consensus on the set of indices of policy performance with which to evaluate the tradeoffs among the alternative policy options and to limit the set of indices to some reasonably small number, say three to seven. This step thus involves defining: (i) a set of indices with which to evaluate policy performance; and (ii) a time horizon over which to evaluate the potential outcomes of each policy. The indices most often chosen usually include measures of the expected average catch biomass, interannual variability in catch biomass, and probability of depleting the stock below some threshold (Bergh and Butterworth, 1987).

(3) Specify the alternative hypotheses
While the main goal of stock assessment is to evaluate the consequences of alternative actions, it is impossible to unequivocally specify the exact consequences of each action. Instead there are often hypotheses for system dynamics that determine the biological and economic responses of the system to each of the alternative actions that can be taken. These hypotheses consist of structurally different models that could each describe population dynamics. They also consist of alternative values for the parameters in the models. In order to conceptualize uncertainty in the potential outcomes of an action that could be taken, Bayesian theory requires the specification of a set of alternative hypotheses about possible "state of nature". Thus, this step involves specifying the alternative models that could describe population dynamics, and the potential ranges of values for the parameters in the models. The construction of alternative hypotheses is carried out mainly by scientists, though fishermen and managers are sometimes consulted. For example, fishermen can be consulted for their knowledge about the fish stocks, particularly for the construction of hypotheses regarding the spatial structure and migratory behaviour of fish stocks (Punt, 1993b). Most often, only one population dynamics model is selected and a subset of its parameters are treated as uncertain. However, with only a single model, the actual consequences for each alternative policy option could be completely missed by the analysis. Including structurally different models as alternative hypotheses for the true population dynamics reduces this possibility and gives managers a more complete basis upon which to evaluate the potential consequences of the alternative policy options. The following steps are carried out mainly by the stock assessment scientists.

(4) Determine the relative weight of the evidence in support of the alternative hypotheses
This involves using Bayes theorem (see below) to estimate the posterior distributions for the input parameters to each model and the marginal posterior probability for each alternative model, if such alternatives are considered (below and Appendices 1–3). This is by far the most difficult step. Computer programing skills are typically required for this step and the next in order to carry out the quantitative evaluation.

(5) Evaluate the distribution and expected value of each management performance measure for each alternative management policy
This usually involves the following steps: (i) Randomly draw values for parameters from the posterior probability distribution (e.g. r and K in the logistic model); (ii) Using the drawn parameter values, project from the current year the model into the future and apply the policy of interest in each future year to predict its consequencees; (iii) Calculate the performance indices for each policy; (iv) Repeat steps (i) to (iii) many times; (v) Produce a distribution of performance indices for the policy.

(6) Present the results to decision makers
Summarize results for each policy in the form of marginal posterior distributions and expected values for the management quantities of interest. Graphs of these distributions and decision tables can often be helpful for conveying the decision analysis results to decision makers (see example Hilborn *et al.*, 1993a; McAllister *et al.*, 1994).

As the estimation of the posterior distribution for model input parameters is the most difficult step in this framework, this step is discussed in considerable detail further below.

Why the Bayesian concept of probability is more useful than the frequentist concept for decision making under uncertainty

In step 4 above, Bayesian probabilities are required to determine the weight of evidence in support of the alternative hypotheses. This is because in decision analysis (Raiffa, 1968; Berger, 1985) the probabilities assigned to alternative hypotheses must reflect the degree of belief that each hypothesis is true. In this section the authors explain why such probabilities can only be obtained from Bayesian theory and not from conventional frequentist statistical theory.

The manner of viewing probability is the most fundamental difference between the Bayesian and conventional statistical approaches (Arnold, 1990). Under conventional or "frequentist" statistics, probability is conceived from a frequentist viewpoint: for a random event (e.g. the number of heads, H, in 100 coin tosses), a probability distribution can be defined that is based on an expected frequency distribution of the event (e.g. H has a binomial distribution with parameters n, the number of coin tosses, and c, the chance of obtaining a head in a single coin toss) and samples come from a well-defined sample space (e.g. the sample space for H ranges from 0 to 100) (Berger, 1985; Walters and Ludwig, 1994). The random event is usually a set of observations (i.e. data) that are obtained from a random sampling process where the values for the parameters that define the process are fixed and unknown (e.g. if it is uncertain whether it is a "fair" coin, then the unknown parameter would be c). Data can only be considered to be realizations of random variables. Furthermore, a probability can be assigned to the obtaining of one particular set of data if one particular set of values for the parameters that determine the distribution were true (e.g. obtaining H=60 heads in 100 coin tosses if c=0.5); probability distributions cannot be assigned to, or calculated for, parameters or related quantities of interest (e.g. a probability distribution cannot be calculated for parameter c in the binomial model) (Arnold, 1990).

In contrast, under the Bayesian viewpoint, probability is conceived of more generally in terms of degrees of belief or credibility of alternative hypotheses or states of nature in light of the data and prior information (Berger, 1985; Walters and Ludwig, 1994). Alternative hypotheses can be defined in terms of alternative discrete values for a model parameter, regions of parameter space for a set of continuous variables or parameters, or a set of structurally different population dynamics models (Sainsbury, 1988). The values for parameters are still considered to be unknown. However, most fundamentally, a value for a parameter can be considered to be a random variable. A probability can thereby indicate the credibility (e.g. given available data) of some value for a parameter relative to other values in the sample space (if states of nature are discrete). Alternatively it can give the probability that the value for the parameter lies within some range if states of nature are continuous in one or more dimensions and the sample space of each is defined. Moreover, a probability distribution for a parameter or vector of parameters ($\boldsymbol{\theta}$) can be defined based on the goodness of fit of a model to data and on prior information not contained in these data. This distribution is known as a Bayesian posterior probability distribution and it is defined using Bayes theorem (Bayes, 1763) (see below). The uncertainty about a parameter thus is represented by regarding the parameter as a random variable with a particular distribution. The procedure for calculating a posterior distribution is referred to as Bayesian estimation.

Using Bayes theorem to calculate posterior probabilities

McAllister *et al.* (1994) defined a conceptual framework for Bayesian stock assessment that deals with the calculation of a posterior distribution for a vector of parameters in only a single population dynamics model. However, Punt and Hilborn (1997) suggest that there always exists an infinite variety of structurally different models that could serve as competing hypotheses for the true population dynamics, and that it may often be appropriate to calculate posterior probabilities for at least a few of such alternatives. Below, the authors extend the framework of McAllister *et al.* (1994) to situations with structurally different models. However, first a single population dynamics model is dealt with.

Calculating posterior probabilities for parameters in a single model

In order to construct a posterior probability density function (pdf) of model input parameters, a model is fitted to data (e.g. a population dynamics model is fitted to relative indices of abundance). For example, if a logistic model is used, the model input parameters would include the intrinsic rate of increase, r, and the carrying capacity, K, that together with the observed catch biomass series determine the model predicted biomass at the start of the current year (i.e. year in which the policy decision is to be made). An additional parameter would be the constant of proportionality for relative abundance indices, q. A vector or set of model parameters (r, K, q) that are considered jointly will be referred to as $\boldsymbol{\theta}$. A state of nature is one potential realization of the set of all possible values for $\boldsymbol{\theta}$, say $\boldsymbol{\theta}_i$ (e.g. r=0.1, K=1000, and q=0.001).

The posterior probability, $P(\boldsymbol{\theta}_i|\,\text{data})$, can be interpreted as a measure of the credibility for $\boldsymbol{\theta}_i$, the unique vector of values i, given the data obtained and the probability of $\boldsymbol{\theta}_i$ is said to be "conditioned on the data". In a situation in which parameter values are continuous (i.e. defined over some non-zero interval on the real number line) rather than discrete, the posterior probability for a given state of nature $\boldsymbol{\theta}_i$ conditioned on the data, $P(\boldsymbol{\theta}_i|\,\text{data})$, is given by Bayes theorem (Bayes, 1763):

$$P(\boldsymbol{\theta}_i|\text{data})=\frac{P(\boldsymbol{\theta}_i,\text{data})}{P(\text{data})}=\frac{p(\boldsymbol{\theta}_i)L(\text{data}|\boldsymbol{\theta}_i)d\boldsymbol{\theta}}{\int p(\boldsymbol{\theta})L(\text{data}|\boldsymbol{\theta})d\boldsymbol{\theta}} \quad (1)$$

where $P(\boldsymbol{\theta}_i$, data) is the joint probability for $\boldsymbol{\theta}_i$ and the obtaining of the data, P(data) is the probability of obtaining the data, $p(\boldsymbol{\theta})$ is the probability density that conveys the prior probability for $\boldsymbol{\theta}$, and L(data $|\boldsymbol{\theta}_i$) is the probability density that conveys the probability of obtaining the data if the set of values in $\boldsymbol{\theta}_i$ were the actual values. L(data $|\boldsymbol{\theta}_i$) is often referred to as the likelihood function of the data evaluated at $\boldsymbol{\theta}_i$ and is easily calculated (see below). $p(\boldsymbol{\theta})$ conveys the probability for a given state of nature prior to obtaining a set of data that can further our ability to discriminate among alternative $\boldsymbol{\theta}_i$ and is easily calculated.

Calculating marginal posterior probabilities for structurally different models

One of the advantages of the Bayesian approach to stock assessment is that it permits the consideration of structurally different models as alternative hypotheses. The calculation of marginal posterior probabilities for alternative models is fundamental to such considerations. The calculation of such probabilities requires a prior probability, $p(m_j)$, for each alternative model j, a prior probability, $p(\boldsymbol{\theta}_{i,j})$, for the parameter vector i under model j, a likelihood function of the data, L(data $|\boldsymbol{\theta}_{i,j}$), given vector i in model j. In Appendix 1 the authors use Bayes theorem to formulate the joint posterior probability, $P(m_j, \boldsymbol{\theta}_{i,j}$ | data), for model j and one set of values for its parameter vector, $\boldsymbol{\theta}_{i,j}$. The authors also formulate the marginal posterior probability, $P(m_j$ | data), for each model, j.

Comments on the use of Bayesian posterior distributions

Bayesian posterior distributions for model input parameters and alternative models are useful in a variety of ways. They fully summarize all that is known about alternative hypotheses based on the goodness of fit between models and data and prior information. For example, a marginal posterior probability distribution can be calculated for each input parameter to a population dynamics model and also for each quantity of interest derived from the model [e.g. for maximum sustainable yield (MSY) which, in the logistic model, is a function of the input parameters r and K]. Posterior distributions can be used to make inferences about any quantity that is a known function of model input parameters. For example, a joint posterior distribution for population model parameters can be used to calculate the probability that stock biomass in a given year is less than some threshold value. Posterior distributions can also be used to compute Bayesian confidence intervals for quantities of interest (see below). As noted above, posterior probabilities for alternative hypotheses are a standard input to decision analytic procedures for policy evaluation (Berger, 1985). Various methods for numerical integration have recently been used in decision analyses to translate joint posterior distributions for model input parameters into marginal posteriors for the potential consequences of alternative policy options (McAllister *et al.*, 1994; Walters and Punt, 1994; McAllister and Pikitch, 1997, Appendix 2). Finally, in recent applications, marginal posterior probability distributions for management quantities appear to have been effective for conveying uncertainties in scientific advice to fishery managers and industry members (Punt, 1993b; McAllister *et al.*, 1994).

Bayesian confidence intervals (CIs) are more easily interpreted than conventional CIs in that they have an intuitive interpretation that is often incorrectly assigned to conventional CIs. A conventional (e.g. 95%) CI can be correctly interpreted as follows: if the data are obtained randomly and a 95% CI were computed, and these steps were repeated independently a very large number of times, the true (but unknown) value would be expected to fall within the estimated CIs, 95% of the time, i.e. 95% of the CIs will include the true value (Arnold, 1990). In contrast, a 95% Bayesian CI suggests that conditionally on the data, there is a 95% probability that the true value lies within the specified interval.

Using prior probabilities in Bayesian stock assessments

Conventional approaches to estimation (e.g. maximum likelihood) use only the data to estimate model parameters; they do not allow prior information to be used and instead assume that nothing is known about the parameter values prior to estimation. In contrast, there can be considerable information about certain parameters prior to evaluating the data to be used in estimation. Bayesian estimation is conceptually intuitive because this information can be incorporated in the prior probability distribution (prior) for alternative hypotheses and the prior is the starting point of estimation. If there is little information in the data, the posterior distribution reflects the prior. If data are informative, the posterior distribution becomes sharper about one of the hypotheses and the prior loses its influence on the shape of the posterior (see example below).

In a stock assessment, prior probabilities are required for alternative models if more than one model is considered and for input parameters within a model if they are uncertain. It is important to distinguish between input parameters and variables derived from input parameters. For example, as noted above, r and K are input parameters to the logistic model; MSY is a variable (or parameter) derived from r and K. It is recommended that the assignment of prior probabilities to derived variables (or derived parameters) should be avoided in order to avoid (1) the Borel paradox (i.e. use of such priors under different parameterizations of the same model can result in different marginal posteriors for the same model quantities) (Wolpert, 1995; Raftery *et al.*, 1995b), and (2) specifying priors with contradictory assumptions about model input parameters (Punt and Hilborn, 1997).

Further, below recent discussions are extended (McAllister *et al.*, 1994; Walters and Ludwig, 1994; Adkison and Peterman, 1996; Punt and Hilborn, 1997) and some recommendations on the use of prior probabilities in fisheries stock assessment summarized. Before this, informative and non-informative prior distributions are defined, the use of the likelihood function explained, some alternative methods for numerical integration summarized, a key step in Bayesian inference, and an example application with the logistic model provided.

Informative and non-informative priors

In a Bayesian stock assessment, experience with similar fish populations can be used to develop informative prior probability distributions for estimated parameters (McAllister *et al.*, 1994; Raftery *et al.*, 1995a; Punt and Hilborn, 1997). Such information can potentially be obtained for many of the parameters in models typically used in stock assessment. For example, in age-structured models, such experience can be used to construct informative priors for the slope of a stock-recruit function, and the extent of recruitment variability about the stock recruit function (Punt, 1993b; McAllister *et al.*, 1994; McAllister, 1995). Informative priors can also be constructed for the intrinsic rate of increase, r, in the logistic model.

Walters and Ludwig (1994) and Punt and Hilborn (1997) caution against using data from other populations to construct informative priors for parameters that represent abundance scaling and habitat size such as the carrying capacity parameter, K. However, informative priors for such parameters could be constructed using habitat considerations in addition to data from similar populations (e.g. Geiger and Koenings, 1991; Hoenig *et al.*, 1994). For example, K could be re-parameterized as the product of the spatial area occupied by the species, and carrying capacity per unit area, $K = kA$. A prior for, e.g. K could then be constructed if the spatial area occupied, A, could be accurately estimated for the population of interest and several other ecologically and taxonomically similar populations, and fishery data (e.g. catch biomass and abundance indices) were available to estimate k for each of the populations.

In some instances, there may be very little information about some parameters in a model (e.g. the scaling parameter q for commercial catch rate indices). In such cases it may be desirable to use non-informative priors. As the name suggests, non-informative priors are assigned to provide little information relative to the experiment (Box and Tiao, 1973); they convey ignorance (or objectivity) with respect to the parameters of interest and "let the data speak for themselves". In many situations, some of the parameters will have informative priors while others will have non-informative priors. In such instances, non-informative priors are used in combination with informative priors so that when the data contain little information, the marginal posteriors for key parameters reflect their priors (Box and Tiao, 1973; Berger, 1985; McAllister *et al.*, 1994). Non-informative priors are also useful for testing the effect of informative priors vs. the data alone on marginal posterior distributions for model quantities.

The identification of the functional form for a non-informative prior, however, is not always straightforward (Gelman *et al.*, 1995). For example, an apparently plausible non-informative prior for the constant of proportionality, q, for a relative abundance index would be uniform on $q>0$. Similarly, a plausible non-informative prior for carrying capacity, K, might be uniform on $K>0$. However, if the relative abundance data are lognormally distributed, and there is little information in these data (e.g. they have a very large CV), then the marginal posterior distribution for K (when q is integrated from the joint posterior distribution for K and q) will favour smaller values for K (Walters and Ludwig, 1994). Therefore, a prior that is uniform on $q>0$ is not non-informative with respect to K. If the prior for q is to be non-informative with respect to K then the appropriate prior for q is uniform on log q or equivalently $p(q) \propto 1/q$ (i.e. under this latter prior and uninformative data, the marginal posterior for K reflects the prior for K – see Figures 2a and 3a; for further details see Box and Tiao, 1973 and Gelman *et al.*, 1995). This latter prior for q has been adopted as a non-informative prior for q in recent Bayesian assessments (Punt, 1993b; McAllister *et al.*, 1994; Stocker *et al.*, 1994) and Punt and Hilborn (1997) recommend it as a default prior when there is no prior information about q. Generally, a non-informative prior for a "scale" parameter, s, such as the constant of proportionality for abundance indices, and the standard deviation in the normal and lognormal density functions, is $p(s) \propto 1/s$. For a location parameter, L, such as the mean in the normal and lognormal density functions, a non-informative prior is $p(L) \propto 1$ (Gelman *et al.*, 1995).

Likelihood functions for Bayesian stock assessment

In addition to the prior distribution, the likelihood function is the other major component of Bayes theorem

[Equation (1)]. The likelihood function is a probability density function (Arnold, 1990) that can be used to evaluate the goodness of fit between data predicted by one set of values for model parameters and the observed data. The likelihood function gives the probability of obtaining the observed data if a given set of parameter values happened to be true. The likelihood function also specifies how data are distributed and needs to be chosen carefully for the type of data used. It is often most convenient to work with the log of the likelihood function, as this function can sometimes approach extremely small values. It is also most commonly assumed that observations are statistically independent of each other. Therefore, the likelihood function is usually the product of the individual likelihood functions for each datum.

In stock assessment, the types of data to which models are most commonly fitted are relative abundance indices and catch numbers-at-age data (from either surveys or commercial catches). A density function that is most commonly used for relative abundance data is the log-normal density function (see example below). For catch-age data, the multinomial density function has often been applied (Methot, 1990; McAllister and Ianelli, 1997).

Methods for numerical integration and the calculation of marginal probability distributions

The calculation of marginal posterior distributions and expected values for management quantities (e.g. MSY) requires the integration of the joint posterior distribution for population model parameters (e.g. r and K in the logistic model). There are a variety of methods that can be used. One important criterion for choosing a method is the number of parameters that are to be treated as uncertain. Bayesian approaches that admit uncertainty in relatively few parameters perform integration in two main ways. The first is analytically (e.g. Thompson, 1992). To enable analytical integration, the functional forms of the prior and likelihood functions need to be carefully chosen (e.g. "conjugate" priors are often used; Arnold, 1990).

The second method uses a grid of values for the estimated model parameters (Hilborn *et al.*, 1994; Walters and Ludwig, 1994, see example below). For each parameter, a minimum and maximum value is defined and the range is divided up into equally spaced intervals. Each range is chosen so that the support from the data and prior falls within it. The posterior probability is calculated at each equally spaced point in the grid of parameter values. For example, in the logistic model, we would start by looping over the values for K. At each value for K and over all values of r, we compute the likelihood times the prior (or "posterior kernel") at each new pair of values for r and K, and then store the values for r, K, and the kernel in a file or array. The result is a discrete approximation of the joint posterior distribution for r and K, i.e. rather than a continuous distribution. A large random sample (e.g. 5000 draws) of the vectors of parameter values is then taken from this empirical distribution for r and K, with the probability of drawing each vector proportional to its posterior probability (see Appendix 3 for a simple algorithm to take draws from some prespecified probability distribution). The management quantities of interest are then calculated for each drawn set of values and the marginal distribution for each is obtained by normalizing its empirical frequency distribution (i.e. an axis covering the full range of values for the management quantity is divided up into equally spaced bins and the number of randomly drawn values falling into each bin is divided by the total number of draws).

The number of calculations required by this method increases geometrically with the number of parameters. For example, if there are 50 grid points per parameter then the number of calculations required exceeds 6 million when four or more parameters are included. Furthermore, some posteriors have very narrow ranges of parameter combinations with empirical support. In these instances very fine step sizes are required and far more than 50 steps may be required to obtain a reasonably precise sample of the parameter space with support. A modification to this method that reduces the number of calculations required integrates over nuisance parameters (Walters and Ludwig, 1994; Walters and Punt, 1994).

When the number of parameters exceeds three or four, it is usually appropriate to use Monte Carlo approaches for numerical integration. These include Markov Chain Monte Carlo (MCMC) methods (e.g. the Gibbs Sampler) and importance sampling [e.g. the sampling/importance resampling (SIR) algorithm] (Berger, 1985; Rubin, 1987, 1988; Gelfand and Smith, 1990; Smith, 1991; Oh and Berger, 1992; West, 1993; Newton and Raftery, 1994; Raftery *et al.*, 1995a; Kinas, 1996). Monte Carlo Bayesian methods are being applied in fisheries in a growing number of instances (Bergh and Butterworth, 1987; Punt, 1993b; McAllister *et al.*, 1994; Raftery *et al.*, 1995a; McAllister and Ianelli, 1997). The authors briefly describe the SIR algorithm and compare it to some other Monte Carlo Bayesian methods in Appendix 2.

Providing policy advice from fitting a logistic model to c.p.u.e. data for Namibian hake

In order to illustrate the calculation of Bayesian posterior probabilities, and their use in policy evaluation, a

Table 1. Total catch and catch per unit effort (c.p.u.e.) data for Namibian hake in ICSEAF Divisions 1.3 and 1.4 (ICSEAF, 1989). c.p.u.e. unit=tons h^{-1} fished.

Year	Catch (tons)	c.p.u.e.
1964	1815	—
1965	93 510	1.78
1966	212 444	1.31
1967	195 032	0.91
1968	382 712	0.96
1969	320 430	0.88
1970	402 467	0.90
1971	365 557	0.87
1972	606 084	0.72
1973	377 642	0.57
1974	318 836	0.45
1975	309 374	0.42
1976	389 020	0.42
1977	276 901	0.49
1978	254 251	0.43
1979	170 006	0.40
1980	97 181	0.45
1981	90 523	0.55
1982	176 532	0.53
1983	216 181	0.58
1984	228 672	0.64
1985	212 177	0.66
1986	231 179	0.65
1987	136 942	0.61
1988	212 000	0.63

logistic model was fitted to catch per unit effort (c.p.u.e.) data for Namibian hake (*Merlucius capensis* and *M. paradoxus*) (ICSEAF, 1989) (Table 1). The example consists of two interrelated parts: (1) estimating quantities of interest to the management of Namibian hake; and (2) predicting the consequences of alternative policy options.

In the first part, posterior distributions are computed for some important population model quantities for Namibian hake and the relative influence of the data and the priors on the posterior distributions indicated. The quantities include the stock biomass in the final year of the data series (1988) (B_{1988}), the carrying capacity (K), the intrinsic rate of increase (r), the fraction of stock biomass remaining from the beginning of the times series (1964) (B_{1988}/K), and the maximum sustainable yield (MSY). The sensitivities of the marginal posterior probability distributions for these quantities to different prior probability distributions for model input parameters and the length of the time series of data were evaluated. The joint posterior for r and K that was computed in this part was used as input to the decision analysis in the next part.

In the second part, the potential consequences of alternative constant catch quota policy options over a 5-year horizon are evaluated to demonstrate how Bayesian decision analysis results can help fishery managers to more effectively account for uncertainty. Policy performance was evaluated by computing posterior probability distributions for the depletion in the final year (B_{1993}/K) for each policy option. Note that the specifications for the following decision analysis are solely for the purpose of illustration and do not necessarily bear any relationship to any actual policy options, procedures, or quantities used in the management of Namibian hake. For example, there would be several other policy performance indicators to evaluate in an actual policy evaluation. However, we stick to one to keep the illustration simple. Below, the example is presented in terms of the six basic steps in a Bayesian stock assessment that were outlined above.

Step 1: Identifying each alternative management action that can be taken

The constant quota policy options that were evaluated included annual catch quotas set at 200, 250, 300, and 350 kt.

Step 2: Specifying indices of policy performance

The time horizon used to evaluate the potential consequences of the alternative policy options was 5 years (from 1989 to 1993). The index of policy performance that the authors chose to evaluate was the depletion in 1993 [B_{1993}/K, Equation (3)]. The most desirable policy option was identified to be the largest catch quota policy that maintained the population above 0.5 K. The main product of the decision analysis was a marginal posterior probability distribution for B_{1993}/K for each policy. A decision table was constructed to display the decision analysis results.

Step 3: Specifying alternative hypotheses using the logistic model

The dynamic, discrete time form of the logistic model is given by:

$$B_y = B_{y-1} + rB_{y-1}(1 - B_{y-1}/K) - C_{y-1} \quad (3)$$

where B_y is the biomass in year y, r is the intrinsic rate of increase, K is the carry capacity in biomass, and C_{y-1} is the observed catch in year y − 1 (Hilborn and Walters, 1992). It will be assumed that stock biomass in the initial year, B_{1964}, is equal to K. Maximum sustainable yield is given by:

$$MSY = \frac{rK}{4} \quad (4)$$

Using this dynamic model, the authors constructed hypotheses about a derived variable, MSY, because this variable provides an indication of the sustainability of different TAC decisions and the consequences of each alternative TAC policy could vary considerably depending on the actual MSY. The four hypotheses on MSY were: (1) MSY<225 kt; (2) 225 kt<MSY<250 kt; (3) 250 kt<MSY<275 kt; and (4) MSY>275 kt. See the decision analysis results section for a discussion of partitioning parameter space for Bayesian decision analysis.

Step 4: Determining the relative weight of the evidence in support of the alternative hypotheses

The likelihood function used

The likelihood function can be constructed assuming that there is a regular relationship between abundance indices and actual abundance. Thus, it was assumed that observed abundance indices (catch per unit effort, I_y) are directly proportional to stock biomass, lognormally distributed, and independent:

$$I_y \sim \text{lognormal}(\hat{q}\hat{B}_y, \sigma^2) \quad (5)$$

where $\hat{B}_y$, the model-predicted stock biomass in year y, is given by Equation (3), $\hat{q}$ is the estimated catchability coefficient, and σ is the lognormal standard deviation. It is cautioned that the first assumption above may be unreasonable because of potential temporal increases in catching power, among other things. However, it is used because it is simple and makes the illustration easier. If such temporal changes in q were suspected, the Bayesian approach would allow us to consider this possibility as an alternative hypothesis and a posterior probability for it and its alternative (i.e. no temporal change in q) could be computed. Moreover, if c.p.u.e. data were only poorly related to population abundance, estimation performance could deteriorate significantly and results could be very imprecise and biased. The Bayesian approach, like other estimation approaches, would have no solutions to these latter problems.

The likelihood function (or probability density function) of the first year of data, $L(I_{1965}|\boldsymbol{\theta}_k)$, is given by the lognormal density function:

$$L(I_{1965}|\boldsymbol{\theta}_k) = \frac{1}{I_{1965}\sigma\sqrt{2\pi}}\exp\left(-\frac{[\log(I_{1965})-\log(\hat{q}\hat{B}_{1965})]^2}{2\sigma^2}\right) \quad (6)$$

where $\boldsymbol{\theta}_k$ is one potential set of values for the uncertain parameters r, K, and q, and σ is the lognormal standard deviation. If we assume that all observations are independent, the likelihood function of the entire set of data is simply the product of the likelihood function over all years:

$$L(I|\boldsymbol{\theta}_k) = \prod_{y=1965}^{1988}\frac{1}{I_y\sigma\sqrt{2\pi}}\exp\left(-\frac{[\log(I_y)-\log(\hat{q}\hat{B}_y)]^2}{2\sigma^2}\right) \quad (7)$$

The parameter σ was treated as known and set at 0.2 [approximately, the maximum likelihood estimate (MLE)]. σ was fixed at 0.2 because when the model was fitted to shorter data series (e.g. 5 years) σ was not estimable with any reliability. As the likelihood function can result in very small numbers, it is best to work with the log likelihood in the computer program. With some rearrangement and the elimination of some constants, the log likelihood is given by:

$$\log[L(I|\boldsymbol{\theta}_k)] = -\frac{1}{2\sigma^2}\sum_{y=1965}^{1988}\left[\log\left(\frac{I_y}{\hat{q}\hat{B}_y}\right)\right]^2 - \text{const} \quad (8)$$

where const is set to some constant value (see Appendix 3 for details).

The prior probability distributions used

For simplicity, it will be assumed that each of the parameters is independent in the joint prior pdf of $\boldsymbol{\theta}$. Therefore, priors for each of the parameters can be constructed independently and the joint prior is given by:

$$p(\boldsymbol{\theta}_k) = p(r_k)p(K_k)p(q_k) \quad (9)$$

where $p(r_k)$, $p(K_k)$, and $p(q_k)$, and the priors for the parameter values r_k, K_k, and q_k.

A prior distribution for K that is fully non-informative with respect to K and B_y would be uniform on the interval K>0. This means that the marginal posteriors for K and B_y would be flat if the data were fully uninformative. The baseline prior for K will be restricted to: $K \sim U(0.5\text{ mt}, 5\text{ mt})$ (to be referred to as "uniform on K") where the lowest bound is approximately equal to the largest observed catch in the time series. Here, the upper bound is arbitrary but specified so as to have very little influence on posterior means (i.e. the ratio of the posterior density at the bounds of the prior to that at the mode is very small, e.g. $<10^{-5}$). An alternative prior for K that would convey slightly more information about it is:

$$\log(K) \sim U[\log(0.5\text{ mt}), \log(5\text{ mt})] \quad (10)$$

[to be referred to as "uniform on log(K)"]. This implies:

$$p(K_i) \propto \frac{1}{K_i} \quad (11)$$

where $p(K_i)$ is the prior probability for one particular value for K. This prior assigns lower credibility to higher values for K and helps to avoid implausibly large posterior expected values for K when there is little information in the data about K.

A prior for r that is non-informative with respect to r, K, and B_y would be uniform on r>0. The baseline will be restricted to $r \sim U(0.01, 1)$ (to be referred to as "uniform on r") where the lower and upper bounds are considered to be very small and large values for r for a hake, respectively. An informative prior for r could be obtained from catch-effort data for similar stocks. For example, an informative prior for r could be:

$$r \sim \text{Lognormal}(0.4, 0.5^2) \tag{12}$$

(to be referred to as "informative for r").

For the catchability coefficient for the commercial catches, q, there was no information available that could be used to develop an informative prior. Therefore, a non-informative prior was used for q with respect to K and B_y.

$$\log(q) \sim U(-\infty, +\infty) \tag{13}$$

(Box and Tiao, 1973; McAllister *et al.*, 1994). This was the baseline prior used for q and it implies:

$$p(q_k) \propto \frac{1}{q_k} \tag{14}$$

Using a grid-based method to estimate the joint posterior distribution of parameters and K

The simplest approach for estimating a posterior pdf for relatively few parameters is to use a grid-based method (Hilborn *et al.*, 1994; Walters and Ludwig, 1994). A range for each parameter is defined and divided into a number of discrete narrow intervals of equal width. Walters and Ludwig (1994) suggest a minimum of 40 intervals. We then step over each parameter value in the grid for each parameter and calculate the posterior "kernel" at each point in the grid of values for parameter r, K, and q:

$$P(r_i, K_i, q_i \mid I) \alpha p(r_i) p(K_i) p(q_i) L(I \mid r_i, K_i, q_i) \tag{15}$$

where i signifies a unique combination of values for r, K, and q, m is the number of combinations of r, K, and q that are used (e.g. 50 equally spaced values of each parameter might be on the grid giving m=125 000).

However, as q is a nuisance parameter that is of no interest for assessing population status or biology, q will be integrated out of the posterior in all analyses. Walters and Ludwig (1994) showed that under certain conditions gridding over nuisance parameters such as q can be avoided by using a short-cut calculation at each grid point for the key parameters of interest (r and K). Such conditions occur when the nuisance parameters are scale parameters like q and σ in a likelihood function such as the lognormal (as in the current example). For the lognormal likelihood function this short cut to estimating the posterior distribution is implemented by calculating the MLE of q at each point in the grid for r and K. The resulting maximized likelihood (with respect to q) at each grid point, $L^{max(q)}(I \mid r,K,q)$, is directly proportional to the product of the non-informative prior for q and joint likelihood function integrated with respect to q (Walters and Ludwig, 1994):

$$L^{max(q)}(I \mid r,K,q) \propto \int_{q=0}^{\infty} p(q) L(I \mid r,K,q)\, dq \tag{16}$$

where $p(q) \alpha 1/q$ and max(q), the maximum likelihood estimate of q, is given by:

$$\max(q) = \exp\left[\frac{1}{24} \sum_{y=1965}^{1988} \log\left(\frac{I_y}{\hat{B}_y}\right)\right] \tag{17}$$

If there are 50 points on the grid for each parameter, this reduces the number of grid points from 125 000 to 2500. It also reduces the imprecision and "bumpiness" in marginal distributions that results from coarseness in the grid over q because an exact value for the integral in Equation (16) is provided instead. In order to keep notation simple the term $L(I \mid r,K)$ will be defined as:

$$L(I \mid r,K) = \int_{q=0}^{\infty} p(q) L(I \mid r,K,q)\, dq \tag{18}$$

Therefore, instead of calculating the quantity in Equation (15) at each grid point for r, K, and q, we compute at each grid point for r and K:

$$P(r_i, K_i \mid I) \alpha p(r_i) p(K_i) L(I \mid r_i, K_i) \tag{19}$$

Finding the most suitable upper and lower bound, and step-size for the parameter grids is not always a trivial task (Adkison and Peterman, 1996). It is desirable to select an interval that has very low posterior probability (e.g. $<10^{-6}$) outside of it and a step-size that accurately profiles the posterior probability gradients in the zones of support (parameter regions with non-negligible posterior probability density). However, it is also desirable to select a step-size that is as large as tolerable in order to keep the computing time down. A useful starting point for setting the bounds is to pick the extremes of the range of values for the prior or points beyond which there is effectively zero prior probability. The most suitable ranges and step sizes can then be found by trial and error. If the support from the data are well within the range for the prior, then the range for each parameter in the grid search can usually be narrowed. An alternative is to use a non-linear search method to identify the posterior modal values of each parameter and then to adjust the range and step size for each parameter about the mode until the most appropriate grid is identified. In the present example, the grid used for K was 2–5 mt with a step size of 0.01 mt and that for r was 0.2–0.7 with a step size of 0.01. However, when fewer years were used in the data series, the lower bound for K on the grid was set at 0.7 mt.

In summary, the steps in estimating the joint posterior distribution for r and K were as follows. At each grid point for r and K, the model [Equation (3)] was projected from virgin conditions in 1964 ($B_{1964}=K$) to the final year (1988) using the observed catch series [Equation (3)]. If the stock biomass in any year dropped to zero or less the prior was set to 0. At the end of this projection, the MLE for q [Equation (17)], the log likelihood [Equation (8)], and log of the joint prior [Equation (9)] were computed. The values of r, K, the MLE of q, B_{1988}, and the log of $p(r)p(K)L(I|r,K)$ were stored in an output file. The file thereby contained a discrete (rather than continuous) approximation of the joint posterior for r and K. This output file was then used to calculate marginal distributions and expected values for r, K, MSY, B_{1988}, and B_{1988}/K [Equation (20) below]. The file was also used as input to a decision analysis of alternative harvesting policy options, as described later.

Computing marginal posterior probability distributions for management quantities

The marginal posterior probabilities for each quantity of interest are computed by integrating the joint probability with respect to all of the other quantities. The marginal posterior probability that a quantity, such as stock biomass in the final year Y, is within some interval [e.g. $P(B_{Y,j} \leq B_Y < B_{Y,j+1}|I)$] can be obtained by:

$$P(B_{Y,j} \leq B_Y < B_{Y,j+1}|I) = \frac{\sum_{i=1}^{m} H_{i,j} p(r_i) p(K_i) L(I|r_i,K_i)}{\sum_{k=1}^{m} p(r_k) p(K_k) L(I|r_k,K_k)} \quad (20)$$

where m is the number of combinations of r and K that are used, $H_{i,j}=1$ if:

$B_{Y,j} \leq (B_{Y,i}|r_i,K_i) < B_{Y,j+1}$, and $H_{i,j}=0$, otherwise.

The posterior coefficient of variation (CV) for each management quantity, X, was computed by dividing the posterior standard deviation by the posterior expectation or mean for the quantity of interest:

$$CV(X) = \frac{\sqrt{\sum_{i=1}^{m} P(\theta_i|I) X_i^2 - \left[\sum_{i=1}^{m} P(\theta_i|I) X_i\right]^2}}{\sum_{i=1}^{m} P(\theta_i|I) X_i} \quad (21)$$

Step 5: Evaluating the expected value of each performance measure for each alternative management policy

The expected depletion resulting from each policy was calculated under each hypothesized range of values for MSY using the following procedure:

(1) Randomly draw 5000 vectors of $\boldsymbol{\theta}''$ ($\boldsymbol{\theta}''$ = r, K, B_{1988}, MSY) with replacement from the discrete approximation to the posterior distribution of $\boldsymbol{\theta}''$ obtained by the grid based method, with the probability of drawing each vector, $\boldsymbol{\theta}_j''$, being proportional to $p(r_j)p(K_j)L(I|r_j, K_j)$. See Appendix 3 for an algorithm for randomly drawing values from such a distribution.

(2) Using each drawn vector, $\boldsymbol{\theta}_j''$, and applying one of the constant quota policy options in each year, project the biomass dynamic model [Equation (3)] from the year 1988–1993.

(3) In each projection j, calculate the values for MSY_j and $B_{j,1993}/K_j$, determine the bin (i.e. hypothesized range of values) that MSY_j falls into, and under each bin for MSY, add the value for $B_{j,1993}/K_j$ to the sum for that bin. Also, for each range for MSY, sum the number of projections in which the simulated MSY falls within it.

(4) To obtain the unconditional expected value for B_{1993}/K, divide the sum of values for B_{1993}/K by 5000. To obtain the expected value for B_{1993}/K under each hypothesis for MSY, divide the summation of values for B_{1993}/K under each hypothesized range for MSY by the number of values falling in that range for MSY.

(5) To obtain the marginal probability for each hypothesis for MSY, divide the total number of values falling under each range for MSY by 5000.

Step 6: Presenting the results to decision makers

Estimation results

The joint posterior for parameters r and K using the baseline priors indicates a strong negative correlation between parameters r and K (corr = −0.96). Note that the ridge of the posterior surface for r and K descends from lower values of K and higher values of r to higher values of K and lower values of r (Fig. 1). Despite the strong negative correlation between r and K, the data appear to be strongly informative because the marginal posterior distributions for r and K are highly peaked and insensitive to the priors used (Figs 2c and 3c). In the baseline case, the mean values of r and K are 0.35 (CV=0.18) and 3.0 mt (CV=0.14), respectively (Table 2). Ninety-five percent Bayesian confidence intervals (based on the 2.5th and 97.5th percentiles of the marginal posteriors) for r and K lie between 0.24 and 0.48 and 2.3 mt and 3.8 mt, respectively.

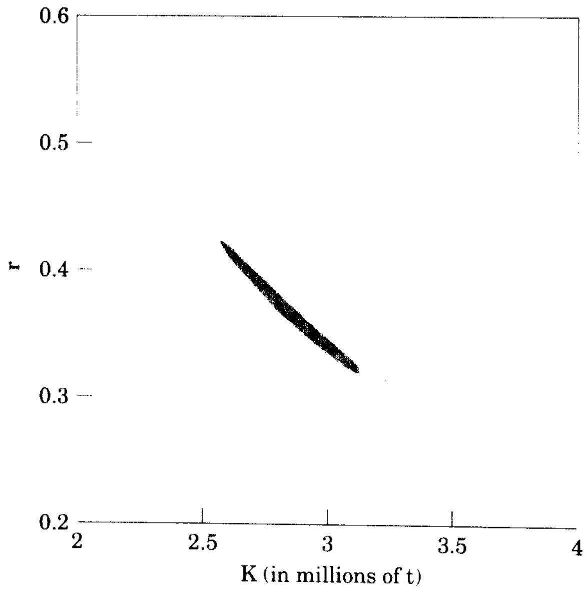

Figure 1. Joint posterior probability distribution for parameter r (intrinsic rate of increase) and K (carrying capacity) using baseline prior: r ~ U(0.01, 1), K ~ U(0.5 mt, 5 mt), log(q) ~ U($-\infty$, $+\infty$).

As expected, the marginal posteriors for r and K showed increasing sensitivity to the priors used as the length of the c.p.u.e. time series was decreased (Figs 2 and 3). For example, when only the first 5 years of c.p.u.e. data were used, the marginal posteriors for r were relatively flat for all cases except those with the informative prior for r (Fig. 2a). The estimated posterior CVs for r were 0.59 in the baseline case and 0.40 in the case with the informative prior for r and uniform prior for K (Table 2). The prior CV for r in the non-informative prior was 0.58, while the prior CV for r in the informative prior for r was 0.50. Therefore, there is almost no information about r in the first 5 years of data.

With the first 15 years of data, the marginal posteriors for r were still quite flat and they differed noticeably depending on the prior used for r (Fig. 2b). For example, the posterior mean of r was 0.56 (CV=0.41) with the uniform priors on r and log K, and 0.46 (CV=0.37) with the informative prior for r and uniform prior on K. With the full 24-year time series, the marginal posteriors for r were highly peaked and very similar under each of the priors (Fig. 2c, Table 2). For example, the posterior means for r were from 0.35 to 0.37 with CVs from 0.16 to 0.18. Not surprisingly, when there were few data, the marginal posterior for r was more sensitive to the prior used for r than for K (Fig. 2a).

When data were few (e.g. with only the first 5 years of c.p.u.e. data), the marginal posteriors for K was, in contrast, more sensitive to the prior used for K than for r (Fig. 3). Furthermore, the marginal posteriors for K were all relatively flat and had long right-hand tails (Fig. 3a). The marginal posteriors for K with the uniform prior on log K, however, favoured smaller values for K than the marginal posteriors for K with the uniform prior on K. For example, the posterior means for K were 2.39 mt (CV=0.45) under the baseline prior, and 1.96 mt (CV=0.47) under the uniform prior on r and log K (Table 2). With the first 15 years of c.p.u.e. data, the left-hand tails and modes of the marginal distributions shift right to higher values and the posterior means for K were 2.63 mt (CV=0.34) and 2.37 mt (CV=0.33), respectively (Fig. 3b, Table 2). The minimum possible value for K increased with increases in the length of the catch series because only larger values of K could result in non-zero values for stock biomass as the catch series increased. With the full time series, the marginal posteriors for K are very peaked and relatively insensitive to the prior used (Fig. 3c, Table 2). For example, the posterior means for K ranged from 2.92 to 3.00 mt with CVs from 0.12 to 0.14.

The marginal posteriors for stock biomass and stock depletion in the final year, and MSY also lost their sensitivity to the prior used as the c.p.u.e. time series increased (Figs 4–6, Table 2). With the first 15 years of data, the marginal posteriors for final stock size, depletion, and, for MSY, appear to be affected less by the choice of a prior for r and K than the marginal posteriors for r and K (Figs 2b, 3b, and 5, Table 2). The distributions for MSY and depletion are narrower than those for r and K, mainly because given values for them can be obtained from many different combinations of values for r and K (e.g. the same MSY can apply to a small productive stock or a large unproductive one).

The marginal posteriors for stock biomass, stock depletion, and MSY also show differing sensitivities to the priors used for r and K. The marginal posterior for the final stock biomass was more sensitive to the prior for K than the prior for r (Fig. 4a, Table 2). This is indicated because the marginal posterior with the prior that is uniform on K and informative for r is very similar to the posterior with the uniform prior on K and r. In contrast, the marginal posteriors that have a uniform prior on log(K) are very similar even though they have different priors for r. For the same reason, the marginal posterior for depletion appears to be more sensitive to the prior on K than the prior on r (Fig. 4b, Table 2).

In contrast, the marginal posterior for MSY was more sensitive to the prior used for r than for K (Fig. 4c, Table 2). With only the first 5 years of c.p.u.e. data, the marginal posteriors for MSY with the informative prior for r were more peaked than those with the flat prior for r. The posterior CV for MSY was 0.60 with the informative prior for r and flat prior for K, while it was 0.76 when the priors for r and K were flat. Therefore, when the data are few, it appears that an informative prior for r can potentially be useful for increasing precision in the estimate of MSY. However, the use of an informative

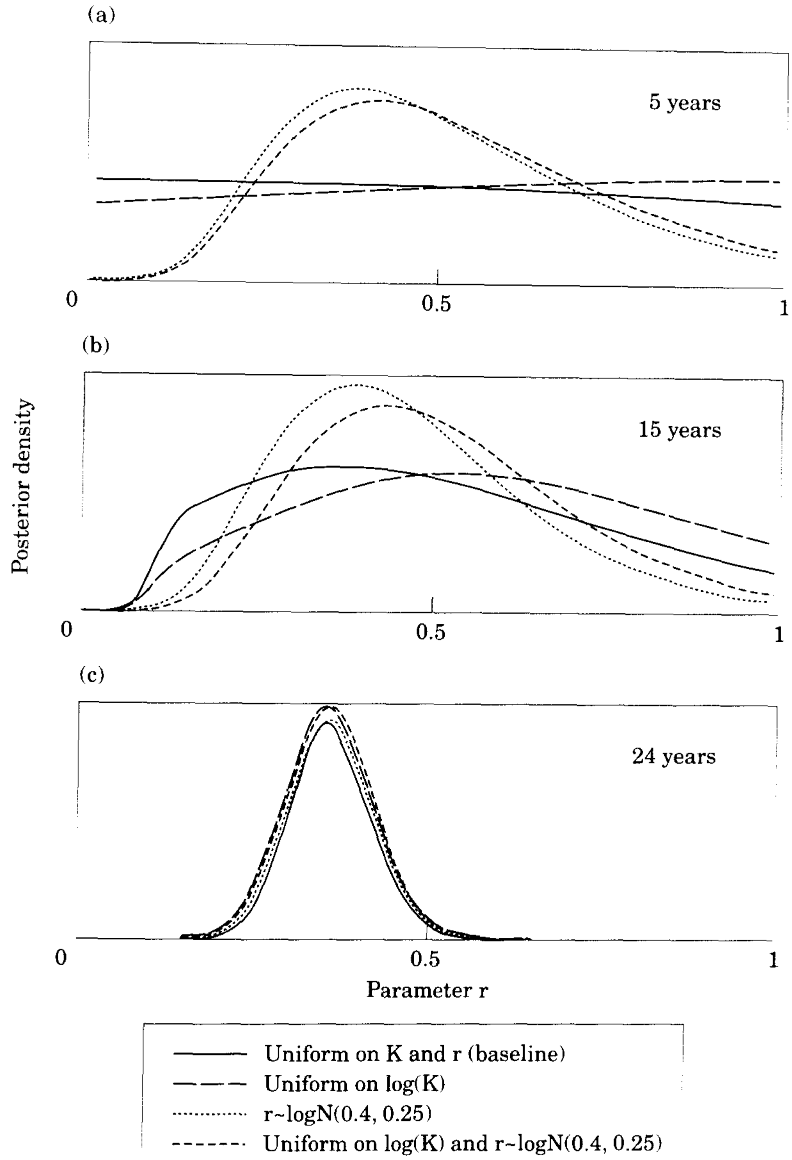

Figure 2. Marginal posterior density functions for r using (a) the first 5 years of catch per unit effort (c.p.u.e.) data, (b) the first 15 years of data, and (c) all 24 years of data and the following priors: (i) baseline, (ii) r – baseline, log(K) ~ U[log(0.5 mt), log(5 mt)], (iii) r ~ lognormal(0.4, 0.5^2), K – baseline, (iv) log(K) ~ U[log(0.5 mt), log(5 mt)], r ~ lognormal(0.4, 0.5^2) (referred to in subsequent figures as the four different priors).

prior for r does not appear to increase the precision in the estimates of depletion (Fig. 4a and 4b, Table 2).

Decision analysis results: evaluating policy performance under uncertainty

The decision analysis results are summarized in a decision table (Table 3). The prior that was used was uniform over r and K. The top row shows the alternative hypotheses in terms of alternative ranges of values for MSY. The second row shows the probability for each hypothesis. Thus, the probability that MSY is: (1) less than 225 kt is 0.02, (2) between 225 and 250 kt is 0.24, and so on. The next row shows the expected value for the ratio of stock biomass in 1988 to K (B_{1988}/K) under each hypothesis. The figure in the last column of this row is the expected value for B_{1988}/K integrated over all hypotheses. The next rows show the expected stock depletion in 1993 under each quota policy and for each hypothesis about MSY. The figure in the last column in each of these rows is the expected depletion for each policy integrated across all hypotheses. Before discussing the results of this decision table we briefly discuss the construction of decision tables to provide advice for fishery managers.

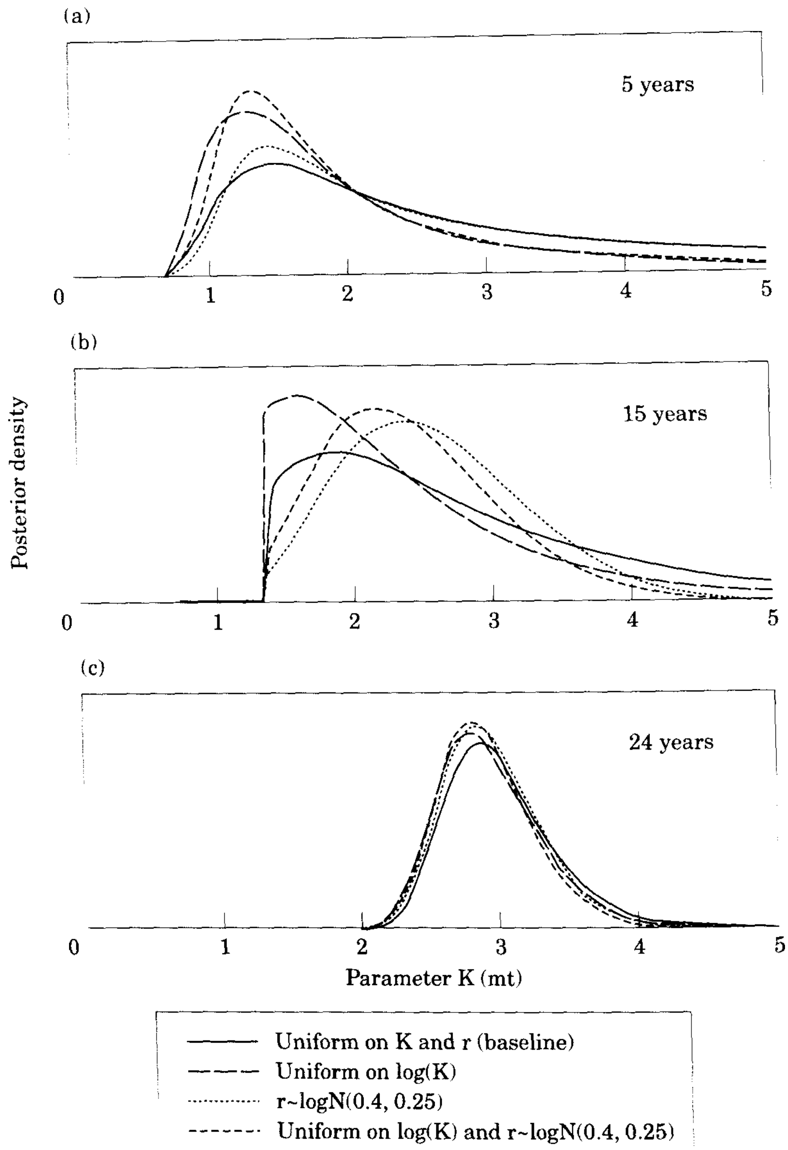

Figure 3. Marginal posterior density functions for K using (a) the first 5 years of catch per unit effort (c.p.u.e.) data, (b) the first 15 years of data, and (c) all 24 years of data and the four different priors.

In a decision table, uncertainty can be conveyed by partitioning the parameter space for a key parameter (e.g. on MSY) into discrete ranges to form alternative hypotheses and the expected consequences under each hypothesis, integrated over all other model parameters, are shown for each policy option. A decision table could also be partitioned over alternative structurally different models with the marginal probability for the model and expected consequences for each policy indicated under each of the alternative models.

The partitioning of parameter space is not an essential part of Bayesian analysis. However, it is useful for conveying uncertainty in the consequences of alternative management actions. The parameter chosen for partitioning should be a key model quantity of interest such as K in a newly exploited stock or MSY in a moderately to heavily exploited stock. There are no set rules for defining where to place the partitions. However, to keep things simple there should be no more than about five alternative hypotheses, and each hypothesis should correspond to an equally sized segment of parameter space, except for the hypotheses on the extremes of the range of values. It is also desirable to partition the parameter space so that the predicted consequences for a given policy option are noticeably different under each adjacent hypothesis, i.e. it is desirable that meaningfully different outcomes are predicted to occur under each alternative hypothesis. Providing these various

Table 2. Posterior means and coefficient of variations (CVs) (in parentheses) for various management quantities. The years refer to the length of the catch per unit effort (c.p.u.e.) time series used. Baseline refers to the use of a prior that is uniform on r and K. U(logK) refers to the use of a prior that is uniform on r and the log of K. inf. p(r) refers to the use of a prior that in informative for r but uniform on K. B_{fin} refers to stock biomass in 1969 for the 5-year c.p.u.e. time series, 1979 for the 15-year c.p.u.e. time series, and 1988 for the full 24-year c.p.u.e. time series. Biomass values [i.e. K, MSY (maximum sustainable yield), and B_{1988}] are in mt.

	Expected value				
	K	r	MSY	B_{fin}	B_{fin}/K
5 yr, baseline	2.39 (0.45)	0.49 (0.59)	0.271 (0.76)	1.78 (0.60)	0.69 (0.19)
5 yr, U(logK)	1.96 (0.47)	0.53 (0.54)	0.238 (0.69)	1.36 (0.66)	0.64 (0.20)
5 yr, inf. p(r).	2.37 (0.45)	0.49 (0.40)	0.283 (0.60)	1.77 (0.60)	0.70 (0.19)
5 yr, inf. p(r), U(logK)	1.96 (0.46)	0.52 (0.39)	0.241 (0.57)	1.36 (0.66)	0.64 (0.20)
15 yr, baseline	2.63 (0.34)	0.49 (0.47)	0.273 (0.20)	0.90 (0.37)	0.34 (0.16)
15 yr, U(logK)	2.37 (0.33)	0.56 (0.41)	0.289 (0.17)	0.81 (0.36)	0.34 (0.17)
15 yr, inf. p(r).	2.61 (0.25)	0.46 (0.37)	0.276 (0.14)	0.89 (0.31)	0.34 (0.16)
15 yr, inf. p(r), U(logK)	2.45 (0.25)	0.51 (0.36)	0.285 (0.13)	0.84 (0.30)	0.34 (0.16)
24 yr, baseline	3.00 (0.14)	0.35 (0.18)	0.259 (0.06)	1.78 (0.14)	0.60 (0.10)
24 yr, U(logK)	2.95 (0.13)	0.36 (0.18)	0.261 (0.06)	1.76 (0.13)	0.60 (0.10)
24 yr, inf p(r).	2.97 (0.13)	0.36 (0.17)	0.260 (0.05)	1.77 (0.13)	0.60 (0.10)
24 yr, inf. p(r), U(logK)	2.92 (0.12)	0.37 (0.16)	0.262 (0.05)	1.75 (0.13)	0.60 (0.10)

conditions are met, it is not necessary that the alternative hypotheses be statistically distinguishable from each other in the real world. For example, the most likely MSY of 250–275 kt for Namibian hake is not statistically distinguishable from the other hypotheses for MSY (Table 3, Fig. 6c). Yet, the consequences under each hypothesis for MSY show clear differences (Table 3). This would demonstrate to fishery managers that the consideration of uncertainty is important since the consequences of the alternative policy options depend strongly on the state of nature assumed. The hypotheses reported in the table should have reasonable probabilities of occurrence, with the most biologically pessimistic hypothesis having a low but non-negligible probability (e.g. 0.30>probability>0.01 if there are 3–5 alternative hypotheses). Punt and Hilborn (1997) suggest that a decision table be produced for each key policy performance index of interest. This sounds reasonable, but in practice, more than a few decision tables is likely to clutter the presentation of results to fishery managers.

The results in the decision table (Table 3) for Namibian hake are now discussed to demonstrate how the decision table can help decision makers to account for uncertainty in decision making and take precautionary action to reduce the chance of occurrence of undesirable outcomes to an acceptably low level (FAO, 1995). The first point to note is that the decision over a constant quota policy for the next 5 years could be quite different if decision makers considered the unpartitioned results only (the unconditional expected values for policy performance) (last column in Table 3) instead of the partitioned results (the expected values under each of the alternative hypotheses). Without partitioning, decision makers might find the 300 kt policy acceptable if they desired the largest catch quote policy that maintained stock depletion above the level expected under long-term harvesting of MSY (i.e. 0.5). After 5 years of TACs of 300 kt, the unconditional expected value for depletion drops from 0.60 in 1988 to 0.52 in 1993.

Now consider the two most pessimistic hypotheses. Under the most pessimistic hypothesis, MSY<225 kt, the expected depletion for the 300 kt policy drops from 0.50 to 0.39. However, decision makers might give this hypothesis little weight seeing its marginal posterior probability is so low (0.02). Under the next most pessimistic hypothesis, that the MSY is between 225 and 250 kt, the marginal posterior probability is 0.24, suggesting that decision makers should take this hypothesis more seriously. This hypothesis suggests that to keep depletion greater than 0.5 over 5 years the 250 kt quota would be more appropriate. Under the most pessimistic hypothesis that is plausible, a depletion of 0.45 is not unreasonable, and this policy would appear to be the most acceptable. Therefore, on all grounds, 250 kt would be most reasonable.

The marginal posterior distributions for the policy indicator, stock depletion, indicate the uncertainty in the outcome of each policy (Fig. 7). They can also show the expected probability of some undesirable event. For example, if it is undesirable to reduce stock size below 50% of K, the expected probability that this will occur

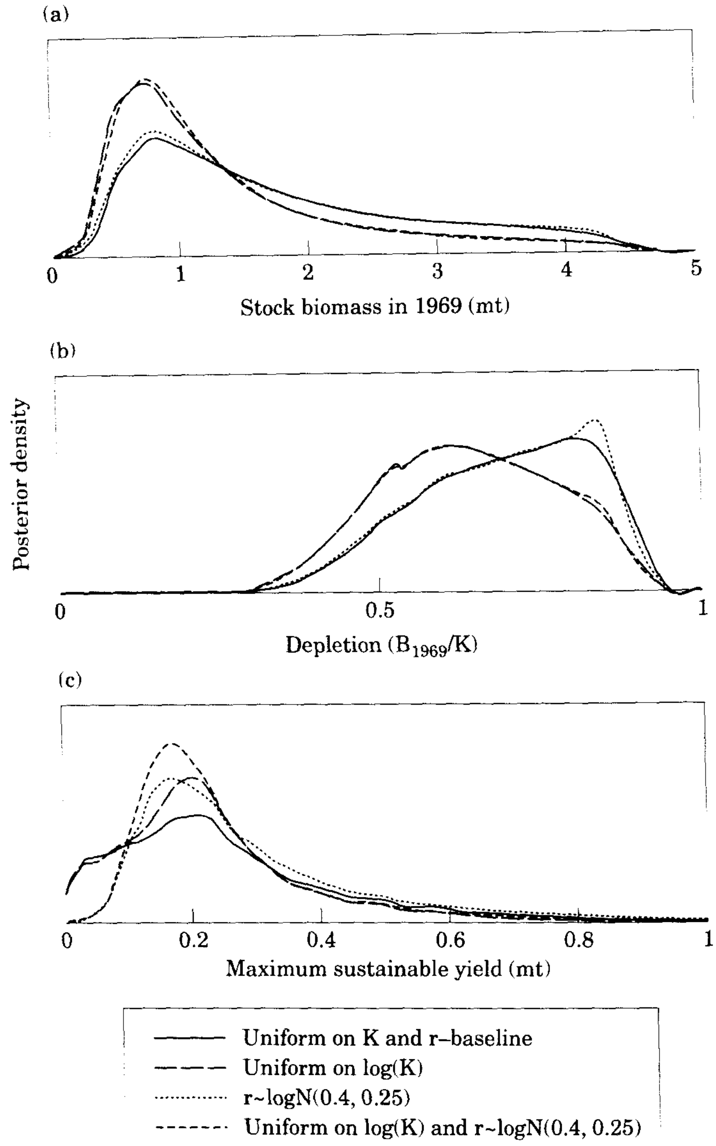

Figure 4. Marginal posterior density functions for (a) stock biomass in 1969, (b) depletion (B_{1969}/K), and (c) maximum sustainable yield using four different priors and the first 5 years in the catch per unit effort (c.p.u.e.) time series.

can be shown graphically for each policy and computed using the marginal distribution for B_{1993}/K. For example, for the 200, 250, 300, and 350 kt constant quota policy options, the probabilities of depleting the stock below 50% of K in 1993 are approximately 0.01, 0.09, 0.38, and 0.82, respectively. Note that the 250 kt policy is also best if you want this probability to be less than 0.10. The marginal posteriors for depletion in 1993 are still highly peaked under each policy as were the marginals for this quantity in 1988 (Figs 6b and 7). This is mainly because the data were highly informative and the population dynamics were assumed to be deterministic. As the data were highly informative, the decision analysis results were practically identical when the different priors were used.

In summary, the example illustrates how fishery managers could account more effectively for uncertainty and take more cautious decisions with the use of a decision table. However, decision tables have only recently been introduced to fisheries stock assessment (Hilborn and Walters, 1992) and there is no clear indication yet that the quality of decisions made has improved with the presentation of results in decision tables. As decision tables are new, it is important for the stock assessment scientists presenting them to ensure that the decision makers learn how to properly interpret decision tables.

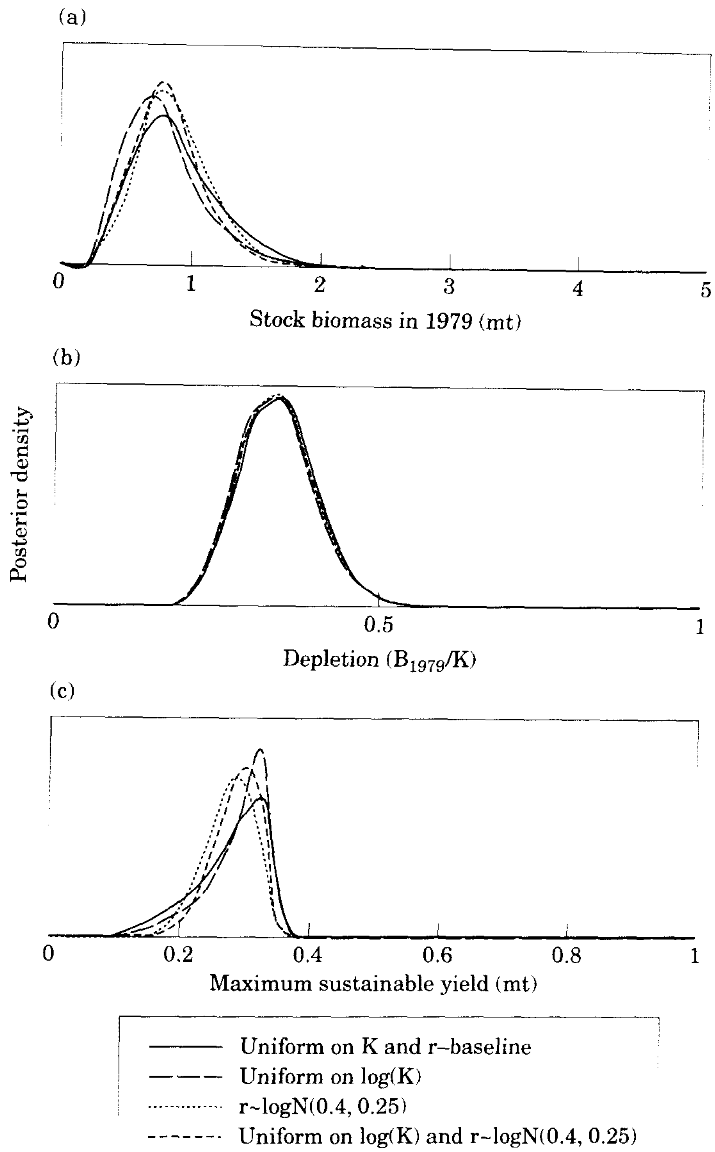

Figure 5. Marginal posterior density functions for (a) stock size in 1979, (b) depletion (B_{1979}/K), and (c) maximum sustainable yield using four different priors and the first 15 years in the catch per unit effort (c.p.u.e.) time series.

Furthermore, we remind readers that the grid-based method in the example is useful only for the simplest of stock assessment problems, i.e. where they are only at most three or four uncertain parameters. For example, with the grid-based method, integration would have been considerably more cumbersome if an informative prior had been available for q. We would no longer have been able to use the computational shortcut that eliminated gridding over q and we would have had to grid over three parameters rather than only two. To deal with more typical fisheries stock assessment problems in which there are multiple uncertain parameters, it is recommended that readers become familiar with Bayesian Monte Carlo methods (Appendix 2). Below some recent developments of Bayesian Monte Carlo methods for fisheries stock assessment are summarized.

Applications of Bayesian Monte Carlo methods to fisheries stock assessment

It is only with the introduction of Bayesian Monte Carlo methods that Bayesian methods have become more widely applied in fisheries stock assessment. Most applications (Francis *et al.*, 1992; Punt, 1993b; McAllister *et al.*, 1994; Stocker *et al.*, 1994; Raftery *et al.*, 1995a; Kinas, 1996; McAllister and Ianelli, 1997) have used importance sampling approaches (Berger, 1985; Rubin,

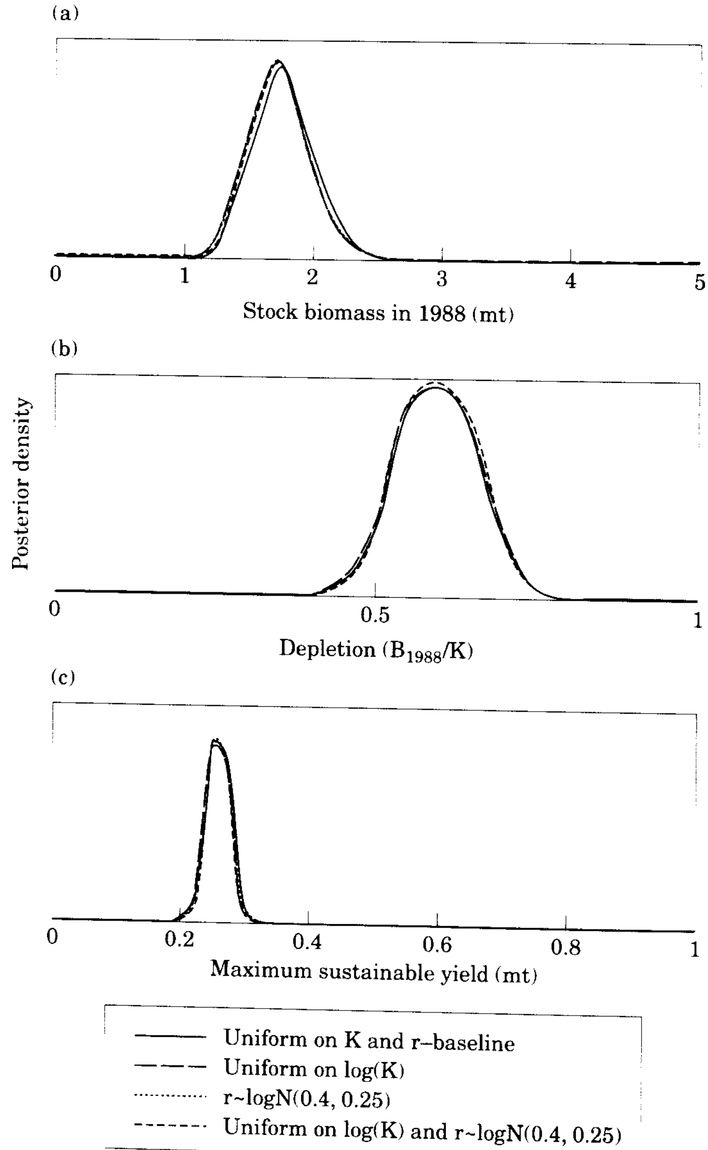

Figure 6. Marginal posterior density functions for (a) stock biomass in 1988, (b) depletion (B_{1988}/K), and (c) maximum sustainable yield using four different priors and the full 24-year catch per unit effort (c.p.u.e.) time series.

1987, 1988; Appendix 2). This is mainly because importance sampling has been relatively simple and straightforward to implement. However, the Hastings-Metropolis algorithm (Metropolis *et al.*, 1953; Hastings, 1970), a Markov Chain Monte Carlo (MCMC) Method (Appendix 2), has also recently been applied.

Fitting population dynamics models to abundance indices

In fisheries stock assessment, Bayesian methods were first applied to fit population dynamics models to abundance indices. In a study that was not explicitly Bayesian, Francis *et al.* (1992) fitted an age-structured model to relative abundance indices and mean length data from a trawl survey for orange roughy (*Hoplostethus atlanticus*) on the Chatham Rise in New Zealand. They applied a method analogous to "simple" importance sampling in order to estimate probability distributions of stock biomass and evaluate the risks of alternative harvesting policy options. Simple importance sampling is a method for numerical integration that is used to estimate, from a joint posterior distribution of the parameter vector $\boldsymbol{\theta}$, marginal posterior distributions and mean values for quantities of interest. A series of independent and identically distributed (iid) random draws of the parameter vector $\boldsymbol{\theta}$ is obtained from a probability density function of $\boldsymbol{\theta}$ called an "importance

Table 3. Decision table showing the expected stock depletion (B_{1993}/K) that could result after 5 years of implementing some alternative constant quota policy options for Namibian hake. Expected values for depletion are shown by state of nature [ranges of values for maximum sustainable yield (MSY)]. The marginal posterior probability for each hypothesis is shown immediately below each hypothesis. B_{1988}/K is the stock depletion in the year (1988) in which the decision is to be made. The expected values for depletion across all hypotheses are shown in the last column.

	MSY (kt)				
	<225	225–250	250–275	>275	
Probability	0.02	0.24	0.59	0.15	Expected
B_{1988}/K	0.50	0.54	0.60	0.68	0.60
200 kt quota	0.52	0.60	0.68	0.74	0.66
250 kt quota	0.45	0.53	0.60	0.67	0.59
300 kt quota	0.39	0.45	0.53	0.59	0.52
350 kt quota	0.32	0.38	0.44	0.51	0.44

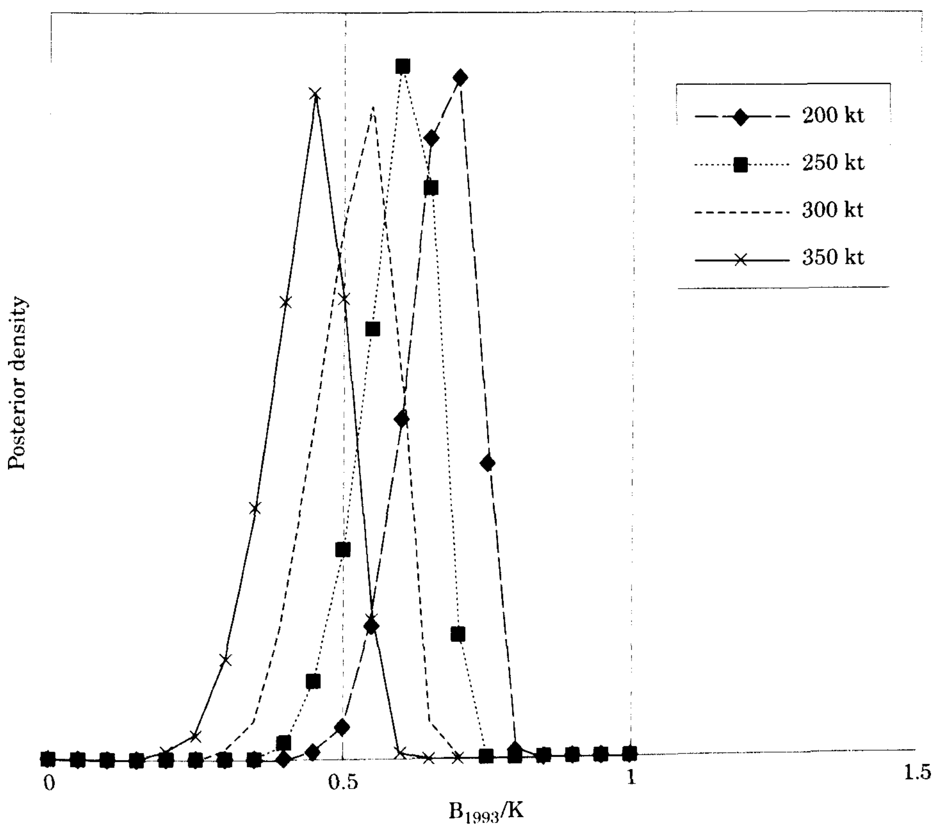

Figure 7. Marginal posterior density functions for B_{1993}/K resulting from the 200, 250, 300, and 350 kt constant quota policy options for years from 1989 to 1993.

function", h(**θ**) (the prior distribution of **θ** was used as h(**θ**) in all applications mentioned in this section; see Appendix 2). For each **θ** drawn, an "importance ratio" [Equation (A2.1), Appendix 2] is calculated and used to provide a discrete approximation of the joint posterior distribution of **θ**.

Francis *et al.* (1992) treated as uncertain only two of several types of population model parameters used (virgin stock size and recruitment anomalies). The importance sampling approach introduced by Francis *et al.* (1992) is easily generalized, and it can deal with uncertainty in multiple parameters. Punt (1993b) extended the work of Francis *et al.* (1992) by increasing the number of parameters treated as uncertain (e.g. by including informative priors for the rate of natural mortality and the extent of recruitment variability), and evaluating the possibility that there were multiple stocks of orange roughy with varying rates of intermixing. However, because of the large number of draws of parameter values required by simple importance sampling (up to millions), it is not practical for most fisheries decision analyses.

Raftery *et al.* (1995a) used an extension of simple importance sampling, the sampling/importance resampling (SIR) algorithm, to estimate probability distributions for quantities in a deterministic population dynamics model for bowhead whale. The addition of the resampling step in SIR, which is done after the initial step of importance sampling and involves far fewer draws, was particularly helpful because it provided a far more efficient method for numerical integration than simple importance sampling (Appendix 2). This made decision analyses of alternative management policy options computationally more tractable. McAllister *et al.* (1994) generalized these approaches (Francis *et al.*, 1992; Punt, 1993b; Raftery *et al.*, 1995a) to provide the flexibility to incorporate uncertainty in recruitment deviates about an estimated stock-recruit function, and informative priors for several model parameters, and to enable a Bayesian decision analysis of alternative harvesting policy options. They estimated a posterior pdf of parameters in an age-structured population dynamics model for New Zealand hoki (*Macruronus novaezelandiae*) by fitting the model to relative indices of abundance and using age-structured data to construct priors on cohort size. They illustrated the use of the posterior distribution for population model parameters in a decision analysis of alternative harvesting policy options.

Fitting population dynamics models to catch-age data and abundance indices

Many stock assessments use catch-age data and VPA-type procedures. It is therefore appropriate to point out some Bayesian methods that can be applied to catch-age data. McAllister and Ianelli (1997) extended the work of McAllister *et al.* (1994) by applying SIR to both catch-age data and relative abundance data to estimate a posterior pdf of population dynamics model parameters. As there is often considerable information about population dynamics in catch-age data from the commercial catch and research surveys, this extension can lead to markedly increased precision in estimates of population model parameters.

The general steps in the stock assessment procedure are the same as those outlined above. However, catch-age data as well as relative abundance indices were added to the likelihood function [Equation (1)], the number of parameters treated as uncertain was increased, and an importance function considerably more elaborate than the prior distribution was applied (Appendix 2). McAllister and Ianelli (1997) illustrated the approach using data for yellowfin sole (*Limanda aspera*) in the eastern Bering Sea. McAllister and Pikitch (1997) used the estimated posterior pdf of population model parameters in a decision analysis of the potential statistical, biological, and economic consequences of alternative designs for the trawl survey there.

The Hastings-Metropolis MCMC algorithm (Metropolis *et al.*, 1953; Hastings, 1970), which involves a random walk over the posterior probability surface (Appendix 2), has also been applied to catch-age data (A. E. Punt, pers. commn, CSIRO, Hobart, Australia; Punt and Hilborn, 1997). Posterior distributions estimated from catch-age data and relative abundance indices using this MCMC method and the SIR algorithm appear to be very similar (McAllister and Ianelli, 1997). Each method has its own advantages and disadvantages (Appendix 2 and below). It is perhaps most prudent to apply both when given the opportunity to see whether the results from the two methods correspond.

Some general recommendations about the use of priors in stock assessment

The use of prior probability distributions in estimation is unique to the Bayesian approach and the development of methods for their construction is currently a key area of research in Bayesian stock assessment. Before concluding, the authors will therefore summarize various recommendations for the construction and use of priors in stock assessment.

To start with, models should be parameterized to facilitate the use of data from other populations (McAllister *et al.*, 1994; Punt and Hilborn, 1997; McAllister and Ianelli, 1997). For example, the Beverton-Holt stock-recruit function is typically formulated in terms of two parameters, α and β as $R=S/(\alpha+\beta S)$, where R is the number of recruits and S is an index of reproductive potential. These two parameters are functionally dependent and prevent comparisons between populations. The Beverton-Holt model can be reformulated in terms of two different parameters, the steepness (the expected ratio of recruitment to "virgin" recruitment when spawning biomass is at 20% of the long-term unexploited or "virgin" level, B_o) and B_o which are functionally independent (Francis 1992). This enables the construction of a prior for steepness using stock-recruit data from similar populations (McAllister *et al.*, 1994). In contrast, the Ricker stock-recruit function (i.e. $R=S\ \exp[\alpha'(1-S/\beta')]$) does not need to be reformulated. Providing that the units of R and S are standardized among populations (e.g. in terms of the age of recruits and the units of reproductive potential), the slope parameter α' is directly comparable among populations and a prior for α' can be estimated from data for several populations (Myers *et al.*, 1996; McAllister and Ianelli, 1997).

In some models, some parameters may be known very precisely (e.g. mass-, and fecundity-at-age in age-structured models) because of extensive sampling. Such parameters can be fixed at their best estimates in order to reduce the number of estimated parameters and to make the estimation more tractable (McAllister *et al.*, 1994; Punt and Hilborn, 1997).

Whenever possible, historical experience from other stocks should be used to construct informative priors in order to improve estimates of model quantities and to avoid debate over the definition of non-informative priors (McAllister *et al.*, 1994; Raftery *et al.*, 1995a; Punt and Hilborn, 1997). However, in most cases, it appears that a mixture of informative and non-informative priors will be appropriate because of a lack of prior information for some parameters and considerable amounts for others.

The functional form of a prior distribution needs to be carefully chosen. A poor choice can cause strong biases in posterior distributions. For example, many parameters (e.g. all of the parameters in the logistic model) are defined only on the positive real number line. In such cases, the lognormal or gamma distributions that include only the positive real number line, could be appropriate as candidate prior distributions for such parameters.

Sometimes it is not clear which prior is most appropriate. For example, different experts might support a different prior distribution for a given parameter. In these cases, the sensitivity of the results to the choice of prior and the implications for management should be evaluated and reported to decision makers (McAllister *et al.*, 1994). In all cases, the procedure used to construct or choose a prior distribution should be documented explicitly in detail and peer reviewed.

Priors for alternative models

Very little attention has been given to assigning priors to structurally different models. This is mainly because there exist few instances in which alternative models have been formally considered in Bayesian stock assessments. Due to computational limitations, it will always be necessary to limit the number of structurally different models, and perhaps it will be necessary to ignore all but one, even though there are other plausible alternatives. Reasons for model selection should always be carefully documented (Butterworth *et al.*, 1996; Punt and Hilborn, 1997). The reasons will likely depend on the types of policy options considered and the level of disaggregation in the fishery data available (e.g. this can help to determine whether an age-structured or a biomass dynamic model should be used, or whether a spatially disaggregated model is needed). Model selection may also depend upon the processes of interest. For example, a model with migration among subpopulations might be considered to be important in some instances (Punt, 1993b). Other models might include trends in environmental conditions, habitat dependent processes, or interspecific and intraspecific competition (Sainsbury, 1988). Note that the level of aggregation in the models specified need not always correspond to the level of detail in the data and information available. This is because it may sometimes be informative to evaluate the potential consequences of alternative policy options under structurally different models, even when data are not available to distinguish among them (Punt, 1993b).

Butterworth *et al.* (1996) suggest guidelines for ranking the plausibility of structurally different models for stock dynamics. The guidelines rank hypotheses based on the availability of data for the population of interest, for similar species in the same and different regions, for any species, and finally solely on theoretical grounds. Butterworth *et al.* (1996) suggest that the qualitative weightings that could be assigned using their guidelines could provide a starting point for assigning prior probabilities to the alternative hypotheses. For example, in absence of any data with the same or other populations and without strong theoretical arguments, it appears to be sensible to give plausible alternative models equal prior probabilities (e.g. Sainsbury, 1988). In contrast, the results of experiments such as those described by Sainsbury (1988) could be used to assign informative prior probabilities to analogous sets of hypotheses about the population dynamics of similar shelf-dwelling demersal assemblages in other tropical regions.

Priors for model parameters

Using data from similar stocks

It may often be reasonable to assume that the value of a parameter for one fish population is similar to that for other populations of the same species in different regions or similar species in the same and different regions. If data for other populations are used to construct a prior distribution, it may be useful to weight the data for each population according to its closeness (e.g. ecological, geographic, and taxonomic) to the one of interest. It would also be useful to incorporate in the prior distribution the uncertainty in parameter estimates obtained from the data for each population (Myers *et al.*, 1993; McAllister *et al.*, 1994). Punt and Hilborn (1997) suggest that meta-analysis (Gelman *et al.*, 1995) could be applied to construct such priors using data for similar stocks. Under this approach, a prior for a parameter in one population is constructed from the joint or "meta-"

analysis of data sets from several similar populations. The goal of this approach is to estimate the distribution for the parameter among the populations of interest and to use this as the prior (Myers *et al.*, 1996).

Traditional Empirical Bayes methods have also been developed to construct priors using data from several populations (Berger, 1985, pp. 96, 167; Hoenig *et al.*, 1994). However, Gelman *et al.* (1995) suggest that the application of meta-analysis using hierarchical probability models to construct priors is preferable to traditional Empirical Bayes methods, which do not use hierarchical probability models. This is mainly because the latter approach does not account for uncertainty in the estimated parameters that specify the prior distribution, while the former approach, with its use of hierarchical probability models, does.

When data from other populations are used to construct a prior distribution, only stocks known to have reasonable stock assessment and data sampling protocols should be selected. The approach is also useful only when data sets are available for stocks that are likely to be representative of the population of interest. It is important to avoid "selection bias", which can occur, for example, when data only exist for large, productive populations, while the one of interest is small and unproductive (Walters and Ludwig, 1994). The reasons and criteria used for including or excluding the data sets used for the prior should be carefully documented.

Using expert judgment

Another approach to constructing priors is to use expert judgment. This can be useful when constructing priors for technical parameters such as the constant of proportionality in an acoustic survey, q (McAllister *et al.*, 1994), or the catchability coefficient in a trawl survey (McAllister, 1995). In these examples, experts who dealt with the surveys of interest were consulted about the various factors that related the biomass estimates given by the surveys to the biomass of the population of interest. The constants of proportionality, q, were modeled as a function of several parameters whose values were uncertain. Each of these parameters was treated as a random variable. For example, an acoustic survey scientist helped to construct a model for q in an acoustic survey for stocks of spawning hoki in New Zealand (Punt *et al.*, 1994), in which q was a function of mean acoustic target strength, the target species identification error, and the fraction of the population in the survey area, among other things. The expert also helped to construct a probability density function of each parameter that conveyed what was known about it. A Monte Carlo simulation was then used to construct a probability distribution for q by taking draws from the pdfs of the factors that together made up q, computing the resulting value for q, and then repeating these steps many times. See Appendix K, McAllister (1995) for an example of constructing an informative prior for q for a trawl survey abundance index.

In an assessment of bowhead whale (*Balaena mysticetus*) population dynamics, Raftery *et al.* (1995a) consulted a variety of sources to construct prior distributions for model input parameters and derived variables. These priors were constructed from biological studies on bowhead fertility, mortality, growth, and other life history parameters, interspecific comparisons of the same, and aerial, land based and acoustic surveys of bowhead abundance and age distribution, among other things. Note, however, that Butterworth (1995), Butterworth and Punt (1996), and Punt and Butterworth (1996) have pointed out several problems with these priors.

Walters and Ludwig (1994) suggest that the use of arguments based on basic biology (e.g. life history attributes) combined with scientific intuition, rather than empirical data, to construct informative priors should be avoided. Common tendencies in using such arguments to specify priors are to allow too much subjective judgment to enter the formulation, to be overly confident, and to specify a prior distribution that is too precise (Berger, 1985; Walters and Ludwig, 1994; Adkison and Peterman, 1996). Priors can also be very biased when important sources of variability are excluded, or when they are constructed as functions of several contributing variables and the variances of the contributing variables are underestimated. Biases can also result if covariances among contributing variables are ignored. Furthermore, the results of several past evaluations of the same stock can also enter the formulation and result in a double usage of the data. These various problems have led Walters and Ludwig (1994) to advocate for fisheries stock assessment the use of non-informative priors instead of priors based on expert judgment. We believe that this may be going too far. Provided that the above tendencies are recognized, well thought-out priors that are based on expert judgment can sometimes help to provide valuable information about plausible values for parameter values. Priors that assign zero probability to some values of a parameter, however, should be avoided unless there is very good reason to do so; if the true value has been assigned zero value in the prior, the posterior probability will also assign zero probability to the true value (Adkison and Peterman, 1996). The influences on results of priors constructed using expert judgment should always be compared with those of non-informative priors and reported to decision makers.

Non-independence among parameters in the prior distribution

Joint prior distributions with dependence among parameters have rarely been applied in fisheries stock assessment. Such priors contain more information than priors with independence among parameters because they suggest that when considered together, certain combinations of values are more or less likely than when parameters are considered independently. For example, Raftery *et al.* (1995a) constructed a joint prior for juvenile and adult survival rates for bowhead whale that reflected the belief that adult survival was greater than juvenile survival. Butterworth (1995), however, pointed out that the correlation in the prior distribution between survival rate and the age at maturity was missing. Priors with non-independence among parameters are appropriate whenever non-independence is suggested by theoretical considerations and empirical evidence. Non-independence may be suggested by technological, life history, or other ecological considerations or by significant correlations among parameters when they are jointly estimated from data not in the likelihood function (McAllister *et al.*, 1994).

Problems that can be encountered with the implementation of Bayesian methods

While there are significant advantages to the use of Bayesian methods, the authors would also like to discuss some of the problems that may be encountered in their application. One of the most contentious aspects of Bayesian estimation is the requirement to specify priors for uncertain model parameters (Berger, 1985). As discussed above, if they are not carefully constructed, priors may be very biased and overly precise. Even if the data are strongly informative, biased priors can potentially result in large biases in estimates of management quantities (Cordue and Francis, 1994; Adkison and Peterman, 1996). Furthermore, the construction of informative priors can be very tedious and entail considerable amounts of effort, debate, and time spent before results can be obtained.

Computing time can be very lengthy if complex models are used, if there are multiple uncertain parameters, and if the data are very extensive. Bayesian applications of Monte Carlo methods to catch-age data can take from a few hours to several days on a Pentium computer, although computing time can be decreased with speedier machines, languages, and algorithms (e.g. for numerical integration).

In some instances, the posterior may be complex (e.g. multimodal – Kinas 1996) and difficult to estimate. For importance sampling, it may sometimes be difficult or impossible to identify an appropriate importance function (Appendix 2). While, if MCMC methods are used, multimodality can go undetected (Newton and Raftery, 1994) (Appendix 2). Both MCMC and importance sampling approaches could be tried to verify that the same distribution can be obtained with different approaches. If a complex posterior is suspected, adaptive importance sampling could be tried (Kinas, 1996).

There can also be problems with the interpretation of posterior distributions (Adkison and Peterman, 1996). For example, a sharp (or highly peaked) marginal posterior distribution for a quantity of interest (e.g. stock biomass) does not necessarily convey a high degree of certainty. It can also be caused by contradictory data or priors that are inconsistent with the data (Adkison and Peterman, 1996). If data sources are contradictory, some have suggested that analyses should be run separately for each source and the results presented separately to decision makers (Richards, 1991; Schnute and Hilborn, 1993). The effects of the prior on the posterior can also be evaluated by comparing the results with those obtained using non-informative priors. Growing experience also suggests that alternative biological models that are fitted to the same data can sometimes give results that are considerably different. This adds a further advantage to formulating two or more structurally different models as alternative hypotheses and calculating marginal posterior probabilites for each of them. However, the use of Bayesian methods for model checking (i.e. evaluating the goodness of fit of the model to the data and priors) is also advisable (Gelman *et al.*, 1995), especially if only one model is evaluated.

Summary of advantages of the Bayesian approach to stock assessment

The Bayesian statistical approach offers an elegant and theoretically consistent framework within which to provide policy advice; Bayesian methods can be used to account for and convey the full range of uncertainties related to the models and parameters used (Berger, 1985; Arnold, 1990; Gelman *et al.*, 1995; Punt and Hilborn, 1997). Bayesian methods can also be used to implement a precautionary approach to fishery management (FAO, 1995). For example, decision tables can be used to choose management actions that have an acceptably low probability of resulting in undesirable outcomes.

Standard decision analysis methods that can be routinely applied in fisheries policy evaluations require the calculation of probabilities for alternative hypotheses (Raiffa, 1968; Berger, 1985). Only the Bayesian approach allows the calculation of probabilities for alternative hypotheses from data for the population of interest and other prior information (e.g. data for similar populations). Providing that they are carefully constructed, informative prior distributions can be helpful for improving the precision in estimates of management quantities of interest. Other *ad hoc* approaches, such as bootstrapping methods that produce sampling distributions of estimated quantities, do not formally use such prior distributions and cannot be used to compute probabilities for alternative hypotheses, though their results are often interpreted as if they were such probabilities (Francis, 1992; Restrepo *et al.*, 1992; Smith *et al.*, 1993). The use of joint posterior distributions can also account for empirically derived covariances among model parameters that may be missed using non-Bayesian Monte Carlo approaches. Unlike other approaches, the Bayesian approach also permits the calculation of probabilities for alternative models. This can be particularly useful when the identification of the "best" policy depends strongly on the model assumed.

However, while Bayesian methods may be "the methods of choice" we also recognize the utility of applying other approaches for dealing with uncertainties in fisheries stock assessment. The latter approaches may sometimes be preferred because they may be quick and easy to implement and have well understood estimation properties (Schnute and Richards, 1995). The implementation of a variety of stock assessment approaches can also sometimes be useful (Punt and Hilborn, 1997); if the alternative methods lead to the same conclusions, more confidence can be given to them.

Acknowledgements

The authors thank Andre Punt, Ellen Pikitch, Jim Ianelli, and Ray Hilborn for helpful discussions. David Agnew, Julie Rousseau, Jim Cannon, and Kai Lorenzen are thanked for helpful comments on earlier drafts of the manuscript. Chris Francis is thanked for his review of this paper which prompted substantial improvements in presentation. An anonymous reviewer is also thanked for his comments. This work was carried out as part of a project within the Fishery Management Science Programme of the UK Department for International Development.

References

Adkison, M. D., and Peterman, R. M. 1996. Results of Bayesian methods depend on details of implementation: an example of estimating salmon escapements. Fisheries Research, 25: 155–170.

Arnold, S. F. 1990. Mathematical statistics. Prentice Hall Inc. Englewood Cliffs, New Jersey. 636 pp.

Bayes, T. 1763. Essay towards solving a problem in the doctrine of chances. Reprinted in Biometrika, 45: 293–315 (1958).

Berger, J. O. 1985. Statistical decision theory and Bayesian analysis. (2nd Ed). Springer-Verlag, New York, NY. 617 pp.

Bergh, M. O., and Butterworth, D. S. 1987. Toward rational harvesting of the South African anchovy considering survey imprecision and recruitment variability. South African Journal of Marine Science, 5:937–951.

Box, G. E. P., and Tiao, G. C. 1973. Bayesian inference in statistical analysis. Addison-Wesley, Reading, Mass. 588 pp.

Butterworth, D. S. 1995. On reservations concerning the results of the Bayesian synthesis analysis, given the approach used to develop prior distributions for natural mortality rates. Report of the International Whaling Commission, 45: 163–164.

Butterworth, D. S., and Punt, A. E. 1995. On the Bayesian approach suggested for the assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales. Report International Whaling Commission, 45: 303–311.

Butterworth, D. S., Punt, A. E., and Smith, A. D. M. 1996. On plausible hypotheses and their weightings, with implications for selection between variants of the Revised Management Procedure. Report of the International Whaling Commission, 46: (In press).

Cordue, P. L., and Francis, R. I. C. C. 1994. Accuracy and choice of risk estimation for fisheries assessment. Canadian Journal of Fisheries and Aquatic Sciences, 51: 817–829.

Food and Agriculture Organization of the United Nations (FAO). 1995. Precautionary approach to fisheries. Document ST/95/FS.61 (E). Rome.

Francis, R. I. C. C. 1992. Use of risk analysis to assess fishery management stategies: a case study using orange roughy (*Hoplostethus atlanticus*) on the Chatham Rise, New Zealand. Canadian Journal of Fisheries and Aquatic Sciences, 49: 922–930.

Francis, R. I. C. C., Robertson, D. A., Clark, M. R., and Coburn, R. P. 1992. Assessment of the QMA 3B orange roughy fishery for the 1992/93 season. NZ Fisheries Assessment Research Document 92/4.

Gelfand, A. E., and Smith, A. F. M. 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Society, 85: 398–409.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. 1995. Bayesian data analysis. Chapman and Hall. 552 pp.

Geiger, H. J., and Koenings, J. P. 1991. Escapement goals for sockeye salmon with informative prior probabilities based on habitat considerations. Fisheries Research, 11: 239–256.

Geweke, J. F. 1989. Bayesian inference in econometric models using Monte Carlo integration. Econometrica, 57: 1317–1340.

Givens, G. H. 1993a. Contribution to discussion on the meeting on 'The Gibbs sampler and other Markov chain Monte Carlo methods'. Journal of the Royal Statistical Society, B, 55: 75–76.

Givens, G. H. 1993b. A Bayesian framework and importance sampling methods for synthesizing multiple sources of evidence and uncertainties linked by a complex mechanistic model. Ph.D. thesis, University of Washington, Dept. of Statistics, Seattle, WA.

Goodyear, C. P. 1995. Red snapper in US waters of the Gulf of Mexico. MIA-95/96-05. Southeast Fisheries Science Center, Miami, FL.

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, N.Y. 570 pp.

Hilborn, R., Pikitch, E. K., and Francis, R. C. 1993a. Current trends in including risks and uncertainty in stock assessment and harvest decisions. Canadian Journal of Fisheries and Aquatic Sciences, 50: 874–880.

Hilborn, R., Pikitch, E. K., McAllister, M. K., and Punt, A. E. 1993b. A comment on "Use of risk analysis to assess fishery management strategies: a case study using orange roughy (*Hoplostethus atlanticus*) on the Chatham Rise, New Zealand. Canadian Journal of Fisheries and Aquatic Sciences, 50: 1122–1125.

Hilborn, R., Pikitch, E. K., and McAllister, M. K. 1994. A Bayesian estimation and decision analysis for an age-structured model using biomass survey data. Fisheries Research, 19: 17–30.

Hoenig, J. M., Warren, W. G., and Stocker, M. 1994. Bayesian and related approaches to fitting surplus production models. Canadian Journal of Fisheries and Aquatic Sciences, 51: 1823–1831.

ICSEAF. 1989. Historical series data selected for Cape hake assessment. ICSEAF Document SAC/89/Doc./3 (10 pp).

Kinas, P. G. 1996. Bayesian fishery stock assessment and decision making using adaptive importance sampling. Canadian Journal of Fisheries and Aquatic Sciences, 53: 414–423.

McAllister, M. K. 1995. Using decision analysis to choose a design for surveying fisheries resources. Ph.D. thesis, University of Washington, School of Fisheries, Seattle, WA.

McAllister, M. K., and Ianelli, J. N. 1997. Bayesian stock assessment using catch-agedata and the Sampling/Importance Resampling Algorithm. Canadian Journal of Fisheries and Aquatic Sciences, 54: 284–300.

McAllister, M. K., and Pikitch, E. K. 1997. A Bayesian approach to choosing a design for surveying fisheries resources: application to the eastern Bering Sea trawl survey. Canadian Journal of Fisheries and Aquatic Sciences, 54: 301–311

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 Sciences, 51: 2673–2687.

Methot, R. D. 1990. Synthesis model: an adaptable framework for analysis of diverse stock assessment data. *In* Proceedings of the symposium on applications of stock assessment techniques to Gadids. Ed. by L. Low. International North Pacific Fisheries Commission Bulletin, 50: 259–277.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. 1953. Equations of state calculations by fast computing machines. Journal of Chemical Physics, 21: 1087–1092.

Mohn, R. K. 1993. Bootstrap estimates of ADAPT parameters, their projection in risk analysis and their retrospective patterns. *In* Risk evaluation and biological references points for fisheries management. Ed. by S. J. Smith, J. J. Hunt, and D. Rivard. Canadian Journal of Fisheries and Aquatic Sciences, Special Publication, 120: 173–184.

Myers, R. A., Bridson, J., and Barrowman, N. J. 1993. Summary of worldwide stock and recruitment data. Canadian Technical Report on Fisheries and Aquatic Sciences, 2023: 327 pp.

Myers, R. A., Mertz, G., and Barrowman, N. J. 1996. Invariants of spawner-recruitment relationships for marine, anadromous, and freshwater species. Document prepared for the ICES Statistical Committee. ICES CM 1996/D:11.

Newton, M. A., and Raftery, A. E. 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. Journal of the Royal Statistical Society, B., 56: 3–48.

Oh, M. S., and Berger, J. O. 1992. Adaptive importance sampling in Monte Carlo integration. Journal of Statistics and Computer Simulation, 41: 143–168.

Punt, A. E. 1993a. The comparative performance of production-model and ad hoc tuned VPA based feedback-control management procedures for the stock of Cape hake off the west coast of South Africa. *In* Risk evaluation and biological references points for fisheries management. Ed. by S. J. Smith, J. J. Hunt, and D. Rivard. Canadian Journal of Fisheries and Aquatic Sciences, Special Publication, 120: 283–299.

Punt, A. E. 1993b. The implications of some multiple stock hypotheses for Chatham Rise orange roughy. NZ Fisheries Assessment Research Document 93/16.

Punt, A. E., and Butterworth, D. S. 1996. Further remarks on the Bayesian approach for assessing the Bering-Chukchi-Beaufort Seas stock of Bowhead whales. Report of the International Whaling Commission, 46: 481–491.

Punt, A. E., and Hilborn, R. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. Reviews of Fish Biology and Fisheries, 7: 35–63.

Punt, A. E., McAllister, M. K., Pikitch, E. K., and Hilborn, R. 1994. Stock assessment and decision analysis of hoki (*Macruronus novaezelandiae*) for 1994. NZ Fisheries Assessment Research Document 94/13.

Raftery, A. E., Givens, G. H., and Zeh, J. E. 1995a. Inference from a deterministic population dynamics model for bowhead whales. Journal of the American Statistical Association, 90: 402–416.

Raftery, A. E., Givens, G. H., and Zeh, J. E. 1995b. Rejoinder. Journal of the American Statistical Association, 90: 427–430.

Raftery, A. E., and Richardson, S. 1996. Model selection via GLIB. *In* Bayesian Biostatistics. pp. 321–353. Ed. by D. A. Berry, and D. K. Dalene. Marcel and Dekker, Inc. 610 pp.

Raiffa, H. 1968. Decision Analysis. Introductory lectures on choices under uncertainty. Addison-Wesley, Reading, MA, 309 pp.

Restrepo, V. R., Hoenig, J. M., Powers, J. E., Baird, J. W., and Turner, S. C. 1992. A simple simulation approach to risk and cost analysis, with applications to swordfish and cod fisheries. Fishery Bulletin, Washington, 90: 736–748.

Richards, L. J. 1991. Use of contradictory data sources in stock assessments. Fisheries Research, 11: 225–238.

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, 1987 June 1–5. pp. 385–402. Ed. by J. M. Bernardo, M. H. Degroot, D. V. Lindley, and A. M. Smith. Clarendon Press, Oxford. 805 pp.

Sainsbury, K. 1988. The ecological basis of multispecies fisheries, and management of a demersal fishery in tropical Australia. *In* Population dynamics (2nd Edn). pp. 349–382. Ed. by J. A. Gulland. John Wiley and Sons Ltd. 422 pp.

Schnute, J. T., and Hilborn, R. 1993. Analysis of contradictory data sources in fish stock assessments. Canadian Journal of Fisheries and Aquatic Sciences, 52: 2063–2077.

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 Sciences, 52: 2063–2077.

Smith, A. F. M. 1991. Bayesian computational methods. Philosophical Transactions of the Royal Society of London, A, 337: 369–386.

Smith, S. J., Hunt, J. J., and Rivard, D. (Editors) 1993. Risk evaluation and biological reference points for fisheries management. Canadian Journal of Fisheries and Aquatic Sciences, Special Publication, 120. 442 pp.

Stocker, M., Hand, C. M., and McAllister, M. K. 1994. Pacific cod. *In* Groundfish stock assessment for the west coast of Canada in 1993 and recommended yield options for 1994. Ed. by M. Stocker. Canadian Technical Report Fisheries and Aquatic Sciences, 1975: 50–92.

Thompson, G. G. 1992. A Bayesian approach to management advice when stock-recruit parameters are uncertain. Fisheries Bulletin, US, 90: 561–573.

van Dijk, H. K., and Kloek, T. 1983. Monte Carlo analysis of skew posterior distributions: an illustrative economic example. The Statistician, 32: 216–233.

Walters, C. J. 1993. Comment on R.I.C.C. Francis: computing probability distributions for risk analysis. Canadian Journal of Fisheries and Aquatic Sciences, 50: 881–882.

Walters, C. J., and Hilborn, R. 1976. Adaptive control of fishing systems. Journal Fisheries Research Board Canada, 33: 145–159.

Walters, C. J., and Ludwig, D. 1994. Calculation of Bayes posterior probability distributions for key population parameters: a simplified approach. Canadian Journal of Fisheries and Aquatic Sciences, 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 Sciences, 51: 946–958.

West, M. 1993. Approximating posterior distributions and mixtures. Journal of the Royal Statistical Society, B, 55: 409–422.

Wolpert, R. L. 1995. Comment on "Inference from a Deterministic Population Dynamics Model for Bowhead Whales". Journal of the American Statistical Association, 90: 426–427.

Appendix 1: Bayes posterior for structurally different models

It may often be of interest to consider uncertainty about competing models for population dynamics. Such models would contain alternative sets of equations to represent alternative hypotheses for population dynamics (e.g. different stock-recruit functions). A model with dynamics j is defined as m_j. Each model may be plausible to varying degrees prior to evaluating a set of data that may help to further discriminate among the alternatives. The relative plausibility for each model prior to evaluating such data can be expressed by the prior probability for the model $p(m_j)$.

The joint posterior probability for model j and a given realization i of its vector of parameters conditioned on the data, $P(m_j, \boldsymbol{\theta}_{i,j}|\text{data})$, is given by:

$$P(m_j,\boldsymbol{\theta}_{i,j}|\text{data}) = \frac{p(m_j)p(\boldsymbol{\theta}_{i,j})L(\text{data}|\boldsymbol{\theta}_{i,j})d\boldsymbol{\theta}_j}{\sum_{l=1}^{N_m} p(m_l)\int p(\boldsymbol{\theta}_l)L(\text{data}|\boldsymbol{\theta}_l)d\boldsymbol{\theta}_l} \quad \text{(A1.1)}$$

where $p(\boldsymbol{\theta}_{i,j})$ is the prior probability density for vector i under model j, $L(\text{data}|\boldsymbol{\theta}_{i,j})$ is the likelihood function of the data evaluated at $\boldsymbol{\theta}_{i,j}$ under model j, and N_m is the number of alternative models considered. The value $p(\boldsymbol{\theta}_{i,j})$ represents the probability for a given set of values for the parameters in model j prior to obtaining a set of data that can further our ability to discriminate among alternative $\boldsymbol{\theta}_{i,j}$. The marginal posterior probability for model j is given by:

$$P(m_j|\text{data}) = \frac{p(m_j)\int p(\boldsymbol{\theta}_j)L(\text{data}|\boldsymbol{\theta}_j)d\boldsymbol{\theta}_j}{\sum_{l=1}^{N_m} p(m_l)\int p(\boldsymbol{\theta}_l)L(\text{data}|\boldsymbol{\theta}_l)d\boldsymbol{\theta}_l} \quad \text{(A1.2)}$$

The integrals for each model can be estimated using one of the Monte Carlo methods for integration (e.g. by importance sampling – p. 263, Berger, 1985; Appendix 2) or by analytic approximation (Raftery and Richardson, 1996).

Appendix 2: Bayesian Monte Carlo methods for numerical integration

In order to obtain expected values and marginal posterior distributions for management quantities, the joint posterior distribution of the population model parameters, $\boldsymbol{\theta}$, must be integrated. When the dimension of $\boldsymbol{\theta}$ is high (e.g. >3), approximations are most easily obtained using various methods of Monte Carlo integration. Of the Bayesian Monte Carlo methods, "simple" importance sampling and the SIR algorithm (Berger, 1985; Rubin, 1987, 1988) have been applied most often in fisheries stock assessment (e.g. Francis *et al.*, 1992; Punt, 1993b; McAllister *et al.*, 1994; Stocker *et al.*, 1994; Raftery *et al.*, 1995a). SIR is an extension of simple importance sampling and is considerably more efficient. SIR entails the following steps (the first three steps constitute "simple" importance sampling): First, an importance function is defined, $h(\boldsymbol{\theta})$. This is a density function of $\boldsymbol{\theta}$ that is as close as possible to the posterior. It must be a pdf from which a large number, m, of independent and identically distributed (iid) draws of $\boldsymbol{\theta}$ can be taken. The tails of $h(\boldsymbol{\theta})$ must also be no thinner (less dense) than the tails of the posterior (Oh and Berger, 1992). Second, many draws of $\boldsymbol{\theta}$ are taken randomly from $h(\boldsymbol{\theta})$. Third, for each draw, $\boldsymbol{\theta}_k$ (k = 1, 2, . . ., m), an importance ratio is calculated:

$$w(\boldsymbol{\theta}_k) = \frac{L(\text{data}|\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k)}{h(\boldsymbol{\theta}_k)} \quad \text{(A2.1)}$$

where $L(\text{data}|\boldsymbol{\theta}_k)$, $p(\boldsymbol{\theta}_k)$, and $h(\boldsymbol{\theta}_k)$ are the likelihood function of the data, the prior, and the importance function evaluated at $\boldsymbol{\theta}_k$, respectively. Importance sampling, the process of generating $\boldsymbol{\theta}_k$ according to $h(\boldsymbol{\theta})$, forms a discrete distribution over $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_m)$ placing mass:

$$F(\boldsymbol{\theta}_k|\text{data}) = \frac{w(\boldsymbol{\theta}_k)}{\sum_{k=1}^{m} w(\boldsymbol{\theta}_k)} \quad \text{(A2.2)}$$

on each $\boldsymbol{\theta}_k$. This distribution approximates the actual posterior. By the strong law of large numbers and under mild conditions, this approximation improves as m increases (Berger, 1985, p. 263).

For some management quantities (e.g. virgin and current stock biomass, MSY, stock depletion), the discrete distribution $F(\boldsymbol{\theta}_k|\text{data})$ over $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_m)$ can be used directly for numerical integration to compute marginal distributions and expected values. However, for management quantities computed in decision analysis, integration is often tractable only by resampling from this discrete distribution. This is because the use of the full set of importance draws, $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_m)$, in a decision analysis can result in excessive computing times. The fourth step is the resampling part of the SIR algorithm. Distributions for quantities of interest that are known functions of $\boldsymbol{\theta}$, can be estimated efficiently, by taking a random sample of size n with $n \ll m$ from the estimated posterior distribution over $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_m)$ (see McAllister *et al.*, 1994; McAllister and Ianelli, 1997 for details). The end result is a set of n draws of $\boldsymbol{\theta}$ from the posterior. These draws can then be used in a decision analysis.

A major task of importance sampling is finding an appropriate importance function (van Dijk and Kloek, 1985; Berger, 1985; Geweke, 1989; Oh and Berger, 1992). A simple candidate for an importance function is the joint prior pdf (Punt, 1993b; McAllister *et al.*, 1994; Raftery *et al.*, 1995a; Kinas, 1996). This prior can be satisfactory in situations in which the data are not highly informative (Punt, 1993b; McAllister *et al.*, 1994; Raftery *et al.*, 1995a). Otherwise, importance functions that are more similar to the posterior than the prior are required. For example, the multivariate Student t distribution has often been used, with the mean based on the posterior modal value of $\boldsymbol{\theta}$ (found by non-linear minimization) and the covariance of $\boldsymbol{\theta}$ based on the Hessian estimate of the covariance at the mode (van Dijk and Kloek, 1985; Geweke, 1989; Oh and Berger, 1992; West, 1993; McAllister, 1995). Log transformations of some parameters have also helped to account for skewness in the posterior (van Dijk and Kloek, 1985; McAllister, 1995). An importance function based on the likelihood function has also proven useful (Newton and Raftery, 1994).

Alternative Monte Carlo approaches

McAllister and Ianelli (1997) point out alternative Bayesian Monte Carlo approaches for numerical integration. These include adaptive importance sampling (AIS) and Markov Chain Monte Carlo (MCMC) methods (Givens, 1993a,b; Kinas, 1996; Punt and Hilborn, 1997). In contrast to SIR, AIS is an iterative importance sampling procedure that can use as an importance function a finite mixture of multivariate pdfs, such as the multivariate Student density (West, 1993; Kinas, 1996). The normalized importance ratios are used to develop the new importance function and successive rounds of importance sampling are conducted such that the successive importance functions converge on the posterior. A potential advantage of AIS over SIR is that AIS may be computationally more efficient than SIR especially for estimating complex (e.g. multi-modal) posteriors (Kinas, 1996).

In contrast, MCMC methods are based on iterative Markovian updating schemes for estimating posteriors and sampling entails a random walk over the posterior probability surface (Metropolis *et al.*, 1953; Hastings, 1970; Punt and Hilborn, 1997). Markovian Chain Monte Carlo methods can be easier to implement because they may require less coding and initial start-up time than importance sampling methods. However, MCMC methods may be computationally less efficient than importance sampling methods (Smith, 1991; Givens, 1993a). Ensuring that the draws of $\boldsymbol{\theta}$ obtained using MCMC so are independent is also not so straightforward as it is with SIR. In addition, there are some conditions in which MCMC may not necessarily converge on the posterior (e.g. when the posterior surface is multimodal or not log-concave, Newton and Raftery, 1994).

Appendix 3: A simple algorithm for random sampling from a probability distribution

Some of the Bayesian methods for numerical integration (e.g. importance sampling and grid-based methods) require an algorithm for taking random draws from a discrete or continuous distribution. A simple approach for taking draws from such a distribution is to use the cumulative distribution function, cdf or $F(\boldsymbol{\theta})$ (Gelman *et al.*, 1995). For the distribution, $p(\boldsymbol{\theta})$, the cdf is defined as: $F(v) = P(\boldsymbol{\theta} < v)$.

We used the procedure below to take draws of $\boldsymbol{\theta}$ from a discrete approximation of the posterior distribution of $\boldsymbol{\theta} = r$, K in the logistic model. The procedure assumes that a discrete approximation of $P(\boldsymbol{\theta}|\text{data})$ has already been obtained (e.g. through importance sampling or a grid-based method). Furthermore, the procedure assumes that the estimate of $P(\boldsymbol{\theta}|\text{data})$ is contained in a file in which each line k contains a vector of parameter values $\boldsymbol{\theta}_k$, and the natural logarithm of the posterior kernel, $\log[L(\text{data}|\boldsymbol{\theta}_k)\, p(\boldsymbol{\theta}_k)]$. If the dimension of the parameter vector $\boldsymbol{\theta}$ is large (e.g. >5) and an importance sampling method has been used, the vector $\boldsymbol{\theta}_k$ can be represented instead by the unique random number "seed" that was used to generate the sequence of values in the vector $\boldsymbol{\theta}_k$ in the importance sampling algorithm. The steps of an algorithm for randomly sampling $\boldsymbol{\theta}$ from a discrete distribution of $\boldsymbol{\theta}$ [e.g. $P(\boldsymbol{\theta}|\text{data})$] are as follows:

(1) Take $n \ll m$ random draws (e.g. n=5000 or 10 000) of U from the uniform distribution on [0, 1] where m is the number of points in the discrete approximation of the posterior distribution $P(\boldsymbol{\theta}|\text{data})$.
(2) Sort the n values of U drawn from U(0, 1) in ascending order to produce the vector $U_1, \ldots, U_n$.
(3) Compute the normalizing constant for the empirical distribution [e.g. the denominator in Equation (1)]:

$$\text{nconst} = \sum_{k=1}^{m} \exp(\log[L(\text{data}|\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k)] - \text{const}) \quad \text{(A3.1)}$$

where const is set to some constant value that rescales the sum of the log likelihood and log prior to values close to zero for parameter values near the mode of the posterior. A suitable value for const therefore is the sum of the log likelihood and log prior at the posterior mode.

(4) Starting with U_1 in the series $U_1, \ldots, U_n$, find the first value in the cumulative posterior probability distribution function that exceeds the value, U_1, i.e. find $TOTP_k$ such that:

$$TOTP_k = \frac{\sum_{i=1}^{k} \exp(\log[L(data|\boldsymbol{\theta}_i)p(\boldsymbol{\theta}_i)] - const)}{nconst} \qquad (A3.2)$$

and

$$TOTP_{k-1} \leq U_1 \leq TOTP_k \qquad (A3.3)$$

The associated parameter vector, $\boldsymbol{\theta}_k$, is considered to be randomly drawn from P($\boldsymbol{\theta}$ data).

(5) Repeat step (4) over $U_2, \ldots, U_n$ to obtain n randomly drawn samples from P($\boldsymbol{\theta}$ data).

Note that it is possible to draw the same value of $\boldsymbol{\theta}_k$ more than once from P($\boldsymbol{\theta}$ data) because more than one value in the series $U_1, \ldots, U_n$ can satisfy the condition A3.3 for a given member k in the posterior series $k=1, \ldots, m$. If a high percentage of the values drawn are of the same vector, $\boldsymbol{\theta}_k$, this can indicate that the posterior is estimated too imprecisely. A useful diagnostic, therefore, is the maximum number or percentage of draws of the same member k in the randomly drawn series $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$. For example, if more than 1% of the draws of $\boldsymbol{\theta}$ are the same vector, $\boldsymbol{\theta}_k$, then a more precise estimate of the posterior P($\boldsymbol{\theta}$ data) may be desirable.