A proposed stock assessment method and its application to bowhead whales, *Balaena mysticetus*

David Poole
Department of Statistics
P.O. Box 354322
University of Washington
Seattle, Washington 98195-4322
E-mail address: poole@stat.washington.edu

Geof H. Givens
Department of Statistics
Colorado State University
Fort Collins, Colorado 80523

Adrian E. Raftery
Department of Statistics
University of Washington
P.O. Box 354322
Seattle, Washington 98195-4322

Abstract.–Monte Carlo simulation from probability distributions is often favored as a means of quantifying the uncertainty in the results of a population analysis. Observed data are combined with simulations from a population model by using subjective distributions for model parameters for which no data are available. The results from such methods can unfortunately be inaccurate unless care is taken in the combination of these simulations and the observed data. A Monte Carlo method was proposed at the 1996 meeting of the Scientific Committee of the International Whaling Commission for the assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales. We show that this method is potentially inaccurate, and as such, it appears to be unsuited to the bowhead application and thus possibly to other similarly structured management problems.

Manuscript accepted 18 March 1998.
Fish. Bull. 97:144–152 (1999).

The variability in parameter estimates from a population analysis is of great interest to stock assessment scientists. Monte Carlo simulation is an intuitively appealing and easily applied strategy for quantifying this uncertainty. Over the past six years, various Monte Carlo assessment methods for the Bering-Chukchi-Beaufort stock of bowhead whales, *Balaena mysticetus*, have been discussed by the Scientific Committee of the International Whaling Commission (IWC). A Bayesian approach (Raftery et al., 1995) was adopted and used as the basis for the IWC assessment of the stock in 1994. The method was developed after a 1991 Scientific Committee (SC) recommendation that methods for taking full account of uncertainty about inputs and outputs to population dynamics models be developed. An alternative maximum likelihood approach was also used for bowheads (Butterworth and Punt, 1995; Punt and Butterworth, 1996). In contrast to the adopted method, the latter approach does not allow for uncertainty in the values of various biological parameters. Rather, it assumes that they are known exactly.

At the 1996 SC meeting, a modified maximum likelihood assessment to account for uncertainty in biological parameters was considered (Punt and Butterworth, 1997). The assessment method was an application of a Monte Carlo approach developed by Restrepo et al. (1991, 1992). Punt and Butterworth (1997) cited an example of the use of the Monte Carlo approach by the International Commission for the Conservation of Atlantic Tunas, and Restrepo et al. (1992) applied their approach to swordfish and cod fishery assessments.

In our paper, we review the proposed Monte Carlo approach, both in general and in the specific bowhead application, and evaluate its performance and its compliance with established statistical principles. We show that in some circumstances the method can provide suboptimal results for bowhead assessment, and it may thus be unsuitable for more general fisheries management problems. In order to illustrate the potential pitfalls of the method, some simulations were performed. The paper is presented solely as a scientific appraisal of the suggested approach, and the examples given are purely illustrative. We show how modifications of the technique can lead to improved performance, but we do not formally propose any alternative assessment methods here.

The Monte Carlo approach

Description

Restrepo et al. (1992) described an approach to quantifying the uncertainty in the results of sequential population analyses for various fish stocks. They motivated their approach by noting that

> Fisheries managers recognize the dangers of accepting parameter estimates without consideration of the variability inherent in the estimates of fish stock status and related parameters… If all sources of error are not appropriately accounted for, then estimates of the uncertainty in the assessment results may be too small.

This Monte Carlo approach (hereafter MCA) proceeds as follows. Probability distributions are used to describe uncertainty in the inputs to an assessment model. These distributions are constructed in two ways:

1 If observed data are available for a specific input, a parametric statistical model for the data is assumed, and the parameters are estimated by maximum likelihood. A parametric bootstrap is then used to obtain a sample of input values. These values are used as values of the input in the assessment model.
2 If no data relevant to the input are available, a subjective "prior" distribution, representing educated guesses about the true value, is placed upon it.[1] A sample from this prior is used in the assessment model.

The second case occurs in many stock assessment procedures. For instance, a prior was needed for natural mortality, M, in the swordfish assessment of Restrepo et al. (1992). Similarly, we required a subjective distribution for the growth rate parameter, MSYR in the bowhead whale example in the section "Application of MCA to bowhead whale assessment."

[1] Such a distribution is effectively a Bayesian prior distribution. However, since the framework of MCA is not Bayesian, Restrepo et al. (1992) do not refer to such a distribution as a prior.

The next step in MCA is to compare simulated model outputs, such as a time trajectory of stock sizes to observed data, in order to formulate a likelihood (assuming lognormal deviations). Many input parameter sets are drawn randomly from the specified input distributions (i.e. either from data-based bootstrap or subjective prior), and for each set, a conditional maximum likelihood estimate (MLE) is calculated for quantities of interest, given the fixed input parameters and the observed data. The simulation distribution of such conditional MLEs is used to quantify uncertainty. The simulation is viewed as translating input uncertainties into output uncertainties.[2]

MCA is suggested for situations where (possibly many) nuisance parameters exist. These are typically the model input parameters for which no informative data are available. The basic strategy is to estimate the quantities of interest (e.g. current stock size and production rate) conditional on values of the nuisance parameters and then to integrate over the prior for the nuisance parameters. The distribution of the conditional estimates of the parameters of interest is then examined for the purposes of inference. For example, if $\hat{\theta}_\gamma$ is an estimator of θ, conditional on nuisance parameters γ, then

$$\hat{\theta}_{(1)} = \int \hat{\theta}_\gamma \, p(\gamma) d\gamma \tag{1}$$

is an MCA estimate of θ, where $p(\gamma)$ is the prior for γ.[3]

In practice, the integral in Equation 1 is not calculated exactly; it is approximated by using the Monte Carlo simulation described above. In the form of an algorithm, MCA proceeds as follows:

1 Obtain the MLE $\hat{\theta}_{\gamma_0}$ from the observed data $\boldsymbol{X}$ and a likely value γ_0.
2 Sample γ^* from the prior $p(\gamma)$.
3 Sample pseudo-data $\boldsymbol{X}^*$ from a distribution with density $f(\boldsymbol{x}; \hat{\psi}(\boldsymbol{X}))$, where $f(\boldsymbol{x}; \psi)$ is a model for the data but not necessarily the assessment model and where $\hat{\psi}(\boldsymbol{X})$ is an estimate of the parameters of this model. $\hat{\psi}(\boldsymbol{X})$ may depend on the results of step 1, namely $\hat{\theta}_{\gamma_0}$ and γ_0, or even on standard MLEs $\hat{\theta}$ and $\hat{\gamma}$. An example of f would be to assume $\boldsymbol{X}^* \sim N(\boldsymbol{X}, \hat{\psi})$, where $\hat{\psi}$ is an estimated dispersion matrix.

[2] Restrepo et al. (1992) also consider uncertainty in the assessment model itself, but this issue is not of primary interest here. In the bowhead whale application, the assessment model is fixed by the IWC.

[3] In general these parameters may be multivariate vectors. For simplicity, we restrict our focus to scalar parameters in the subsequent examples.

4 Find the conditional MLE $\hat{\theta}_{\gamma*}$ given γ^* by using the simulated data $\boldsymbol{X}^*$. Store this value.
5 Repeat steps 2–4 many times to obtain a collection of $\hat{\theta}_{\gamma*}$'s.

MCA inference about θ is then based on the distribution of this collection of $\hat{\theta}_{\gamma*}$'s. Typically, the original $\hat{\theta}_{\gamma_0}$, or the mean or median of the Monte Carlo sample, is used as a point estimate, and the 0.025 and 0.975 quantiles form the bounds of a 95% confidence interval.

The Monte Carlo approach is not a straightforward extension of a sensitivity analysis. Usually, in a sensitivity analysis, a small number of alternative parameter values are tried, and the individual point estimates and confidence intervals obtained by using each parameter set are tabled. Inference usually follows from a single analysis where a particular baseline parameter has been set; the remainder of the table is used to assess how conclusions would change under different modeling assumptions. With MCA, a potentially vast number of parameter values are tried, and an overall confidence interval is obtained by pooling the results of each individual analysis, effectively integrating over the distribution of the nuisance parameters. We show in the rest of this paper that this integration is a source of potential bias.

The MCA technique is potentially highly sensitive to violations of its assumptions. If one is going to express uncertainty in the value of a parameter by means of a probability distribution, then this distribution should be treated as a prior in a fully Bayesian setup. The method given in Equation 1 provides estimators that will not necessarily possess the desirable properties of either Bayesian or ML estimators. A general overview of Bayesian methods is given by Lee (1997).

In a Bayesian framework, the best estimator of θ (with respect to squared error loss) is the posterior mean $E(\theta|\,\text{data})$, therefore it would be better to define the estimator as

$$\hat{\theta}_{(2)} = E\big(\theta\,|\,\text{data}\big) = E\Big[E\big(\theta\,|\gamma,\text{data}\big)\Big] = \int \hat{\theta}_{\gamma}\pi\big(\gamma\,|\,\text{data}\big)\,d\gamma, \tag{2}$$

where $\hat{\theta}_{\gamma}$ = the posterior mean of θ conditional on γ; and
$\pi(\gamma|\,\text{data})$ = the posterior distribution of γ.

One might regard Equation 2 as a general strategy and use it in cases where $\hat{\theta}_{\gamma}$ is not necessarily the Bayesian estimator. In this case, however, the properties of $\hat{\theta}_{(2)}$ are not clear.

A simple point estimation example

Schweder and Hjort[4] first identified potential weaknesses with MCA, and they described two situations where differences between the methods of Equations 1 and 2 arose. The first of these is repeated here: consider a random sample of size n from a normal distribution with mean μ and variance σ^2, denoted $X_i \sim N(\mu, \sigma^2)$ for $i = 1, \ldots, n$. Assume that there is a $N(\mu_0, \tau_0^2)$ prior for μ, where μ_0 and τ_0^2 are the prior mean and variance respectively. Let σ^2 be the parameter for which inference is desired. μ is a nuisance parameter and σ^2 is regarded as fixed. The maximum likelihood estimate of σ^2 conditional on μ is

$$\hat{\sigma}_{\mu}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2.$$

The estimators given by Equations 1 and 2 are then

$$\hat{\sigma}_{(1)}^2 = E_{prior}(\hat{\sigma}_{\mu}^2) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)^2 + \tau_0^2$$

and

$$\hat{\sigma}_{(2)}^2 = E_{posterior}(\hat{\sigma}_{\mu}^2) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_{post})^2 + \tau_{post}^2$$

respectively, where μ_{post} and τ^2_{post} are the posterior mean and variance of the nuisance parameter μ. For this simple case, we have closed form expressions for Equations 1 and 2, and no Monte Carlo sample from the prior for μ is required. Note that in Equation 2, $\hat{\sigma}_{\mu}^2$ (the conditional maximum likelihood estimator of σ^2) is used as $\hat{\theta}_{\gamma}$. In addition we note that $\hat{\sigma}_{(2)}^2$ depends on σ^2 because both μ_{post} and τ^2_{post} are functions of σ^2. In other words, the estimator $\hat{\sigma}_{(2)}^2$ depends on the quantity it is trying to estimate. This occurs when $\hat{\theta}_{\gamma}$ is not the Bayesian posterior mean of θ conditional on γ. In our examples, we simply plugged in the ordinary MLE where needed to remove this dependency. Thus, to evaluate $\hat{\sigma}_{(2)}^2$ here, $\hat{\sigma}_{std}^2$ was used as a plug-in estimate of σ^2 in the expressions for μ_{post} and τ^2_{post}. $\hat{\sigma}_{std}^2$ is the usual MLE of σ^2 and is what would normally be used if conditioning on μ were not of interest.

We know from standard theory that $\hat{\sigma}_{std}^2 \to \sigma^2$ as $n \to \infty$. Because $\tau^2_{post} \to 0$ and $\mu_{post} \to \mu$, we have that $\hat{\sigma}_{(2)}^2 \to \sigma^2$ as $n \to \infty$. $\hat{\sigma}_{(1)}^2$, on the other hand, will converge to σ^2 only if $\mu_0 = \mu$ and $\tau_0^2 = 0$. It follows that $\hat{\sigma}_{(1)}^2$ will, in general, yield accurate estimates only if

[4] Schweder, T., and N. L. Hjort. 1997. Indirect and direct likelihoods and their synthesis—with an appendix on minke whale dynamics. Paper SC/49/AS9 presented to the IWC Scientific Committee, October 1997.

μ_0 is close to μ and τ_0^2 is small. We show later the practical implications of choosing an estimator for which the data either do ($\hat{\sigma}^2_{(2)}$) or do not ($\hat{\sigma}^2_{(1)}$) eventually dominate the prior.

An alternative approach that does not involve conditioning on μ is a fully Bayesian analysis where both μ and σ^2 are random variables. This approach requires the specification of a joint prior on μ and σ^2. One such prior is the indifference or "reference" prior (Jeffreys, 1961)

$$p(\mu,\sigma^2) \propto \frac{1}{\sigma^2}.$$

The resulting marginal posterior mean of σ^2 is the best Bayesian estimator of this parameter and is given by

$$\hat{\sigma}^2_{(3)} = \frac{1}{n-3}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

To investigate the difference between the MCA estimate in Equation 1 and the ad hoc "Bayes" approach in Equation 2, we performed some simple simulations. For each of nine combinations of μ_0 and τ_0^2, a random sample of size n = 1000 was drawn from a $N(\mu$=50, σ^2=100) distribution. In each case, $\hat{\sigma}^2_{(1)}$, $\hat{\sigma}^2_{(2)}$, $\hat{\sigma}^2_{(3)}$, and $\hat{\sigma}^2_{std}$ were determined. The results are shown in Table 1.

Table 1

Simulation results for the example of Schweder and Hjort (See Footnote 4 in the main text).

			Estimators of σ^2 (true value is 100)			
μ_0	τ_0^2	MCA: $\hat{\sigma}^2_{(1)}$	Ad hoc "Bayes": $\hat{\sigma}^2_{(2)}$	Full Bayes: $\hat{\sigma}^2_{(3)}$	Std. MLE: $\hat{\sigma}^2_{std}$
50	1	95.4	94.5	94.7	94.4
	9	102.0	93.0	93.2	92.9
	25	121.0	95.9	96.1	95.8
60	1	198.4	100.4	99.8	99.5
	9	197.2	92.1	92.3	92.0
	25	220.8	99.9	100.1	99.8
70	1	499.8	106.6	103.4	103.1
	9	490.0	106.4	106.5	106.2
	25	530.4	109.7	110.0	109.6

The results show very poor performance for $\hat{\sigma}^2_{(1)}$, and good, similar performances for $\hat{\sigma}^2_{(2)}$, $\hat{\sigma}^2_{(3)}$, and $\hat{\sigma}^2_{std}$. In this simple case, the analyst would presumably never choose MCA or the ad hoc "Bayes" method over optimal estimators, such as $\hat{\sigma}^2_{(3)}$ and $\hat{\sigma}^2_{std}$. The key point of this example is that if conditioning on nuisance parameters is to be used, the strategy presented in Equation 2 appears to be preferable to the MCA estimate in Equation 1. As mentioned earlier, the use of Equation 2 has a severe limitation of its own; we therefore do not regard it as a viable alternative approach.

There is considerable literature on the role of conditioning in inference. Reid (1995) has presented a review of recent developments.

Confidence interval estimation

The example in the section "A simple point estimation example" illustrates poor performance for MCA with regards to point estimation. Similar problems occur when constructing confidence intervals.

Consider the following example: let $X_i \sim U(\gamma-\theta\gamma, \gamma+\theta\gamma)$ for i = 1,…,100, denote a random sample from a uniform distribution with bounds specified by the given functions of θ and γ. Let γ be the nuisance parameter. The unconditional MLEs are $\hat{\gamma}$ = (max X_i + min X_i)/2 and $\hat{\theta}$ = (max X_i – min X_i)/(max X_i + min X_i). The conditional MLE of θ given γ is $\hat{\theta}_\gamma$ = (max X_i –min X_i)/(2γ).

Suppose the nuisance parameter, γ, has a $U(a,b)$ prior, $0 \le a < b$. MCA would proceed as follows:

1 Sample $\gamma^* \sim U(a,b)$.
2 Use the parametric bootstrap (e.g. Efron and Gong, 1983) to sample $X_i^* \sim U(\hat{\gamma}-\hat{\theta}\,\hat{\gamma}, \hat{\gamma}+\hat{\theta}\,\hat{\gamma})$, i = 1, . . . ,100.
3 Find the conditional MLE, $\hat{\theta}_{\gamma^*}$, using the bootstrap data, and conditioning on the current γ^*.
4 Store $\hat{\theta}_{\gamma^*}$ and go to step 1. Use the collection of $\hat{\theta}_{\gamma^*}$'s to obtain a confidence interval with the quantile method.

The ad hoc "Bayes" method, which relies on sampling from the posterior, proceeds with the same steps except that step 1 is replaced by

1. Sample γ^* from its posterior distribution.

Here, if we think of the likelihood as a function of γ only, then the posterior for γ is proportional to

$$\prod_i\left[\frac{1}{2\theta\gamma} I(\gamma-\theta\gamma \le X_i \le \gamma+\theta\gamma)\right] I(a \le \gamma \le b),$$

where I is an indicator function given by

$$I(x) = \begin{cases} 1 \text{ if } x \text{ is True} \\ 0 \text{ if } x \text{ is False.} \end{cases}$$

Because this posterior depends on θ, we plugged in $\hat{\theta}$ in the same way as we did for the ad hoc "Bayes" method in the section "A simple point estimation example."

In a simulation example, we let $\theta = 4$, $\gamma = 4$, $a = 0$, and $b = 8$, so that the true value for γ was at the midpoint of its prior. The MLEs were $\hat{\theta} = 4.06$ and $\hat{\gamma} = 3.88$. The 95% MCA interval for θ was (1.99, 87.55). The interval obtained with the ad hoc "Bayes" method was (3.78, 4.03). The MCA interval was 342 times wider than the second interval, and 2.7 times wider than the range of the observed data.

Note that the MLE $\hat{\theta}$ is not contained in the second interval above. This is an undesirable result of the ad hoc "Bayes" method here. One possible remedy is as follows: instead of simply plugging $\hat{\theta}$ into the posterior for γ each time we sample from it, we could attempt to "integrate over theta" by plugging in a different estimate of θ on each occasion. Each of these plug-in values for θ can be obtained by calculating the MLE of θ from a nonparametric bootstrap sample of the real data X. The confidence interval resulting from this strategy was (3.59, 4.08), so the MLE is now contained in the interval. We stress that this is again an ad hoc solution and we strongly favor standard maximum likelihood or Bayesian methods over either MCA or the ad hoc "Bayes" method.

Finally, the example can also be twisted so that the MCA interval is too narrow. Suppose $X_i \sim U(\gamma - \theta\gamma^2, \gamma + \theta\gamma^2)$. This leads to $\hat{\theta}_\gamma = (\max X_i - \min X_i)/(2\gamma^2)$, a cusp-shaped function of γ over any interval that includes 0. Thus, considering γ priors of the form $U(-a,a)$ for $a > 0$, we observed that MCA leads to the surprising result that the width of a quantile-based confidence interval for θ approaches 0 as a increases, while holding the observed data fixed. In other words, the width of the confidence interval is entirely dependent on the prior, and wider priors lead to narrower MCA confidence intervals.

Application of MCA to bowhead whale assessment

Punt and Butterworth (1997) examined the applicability of MCA in the assessment of the Bering-Chukchi-Beaufort stock of bowhead whales. As with the swordfish assessment of Restrepo et al. (1992), the approach involved generation of pseudodata with a parametric bootstrap. In the bowhead case, the real data consisted of abundance estimates and corresponding CV estimates for several years, and observed age-class proportions. Thus, each MCA simulation consisted of the following:

1 *Bootstrapping of data*. A series of pseudo-abundance estimates is bootstrapped from the observed data (Table 1 in Punt and Butterworth, 1997). Each estimate is assumed to be independent and from a lognormal distribution with mean and CV equal to the observed estimates from that year. Pseudodata for fractions of calves and matures are generated from Table 4 of IWC (1995).
2 *Sampling of biological nuisance parameters from priors*. Parameters such as age at maturity and natural mortality rates are generated from prior distributions from IWC (1995).
3 *Conditional estimation*. Conditional on the values of the nuisance parameters, maximum likelihood estimation is used with an age-structured density dependent population dynamics model to obtain estimates of the parameters of interest: carrying capacity (K) and a productivity parameter (MSYR). The likelihood contains contributions from both the abundance and proportion data.
4 *Uncertainty estimation*. The variation in conditional MLEs is used to represent uncertainty.

Note that K and MSYR are the parameters of interest (denoted by θ in our previous notation), whereas the other biological parameters take the role of γ. The distributions from which they were simulated are the prior distributions. The results of 1000 replications of this procedure are used to form confidence intervals.

A simple population dynamics model

For the purposes of illustration, we applied MCA to a simple population dynamics model (PDM). This is a non-age-structured density dependent PDM given by

$$P_{t+1} = P_t - C_t + 1.5(MSYR)P_t\left(1-(P_t / K)^2\right), \quad (3)$$

where P_t = the population in year t, with $t = 0$ corresponding to the baseline year before commercial hunting started (here 1848);
K (or P_0) = the initial population size or carrying capacity;
MSYR = the maximum sustainable yield rate of production as a proportion of the population aged 1+; and
C_t = the number of whales killed by hunting in year t (known exactly).

This model is much simpler than the BALEEN II PDM (de la Mare and Cooke[5]) used by the IWC for

[5] de la Mare, W. K., and J. G. Cooke. 1993. "BALEEN II: The population model used in the Hitter-Fitter Programs". Unpublished manuscript available from the IWC Secretariat, The Red House, 135 Station Road, Histon, Cambridge, UK CB4 4NP.

bowhead assessment, but it nevertheless captures many of the essential features of the bowhead population. The model is viewed as having two inputs (K and MSYR) and one output (P_{1993}); with fixed values of MSYR and the initial K, Equation 3 is applied recursively until P_{1993} is obtained. We use the term "model input" for input parameters whose true value is uncertain. Since the time series of catches C_t is exactly known, we regard it as a set of constants in Equation 3, rather than as a set of model inputs.

In our simplified version of the bowhead analysis, we treated MSYR as a nuisance parameter (γ in our previous notation) and K was the parameter of interest (the θ from before). The only parameter about which we observed data for a parametric bootstrap was P_{1993}. MSYR was assigned a subjective prior distribution. In terms of implementation, we ran the model "backwards" with P_{1993} and MSYR as inputs and K as output. A Newton-Raphson algorithm was used to solve for K. As a result, we effectively had P_{1993} and MSYR as model inputs, and K (obtained conditionally on MSYR and the data) as the model output.

Implementation of MCA and a Bayesian approach

To evaluate MCA, we examined its performance when the true whale stock status was known. "True" values of K and MSYR were selected, and the PDM was run to obtain the "true" value of P_{1993}. Because K (the parameter of interest here) has a "true" value, we were able to assess the accuracy of the estimates produced by the simulations.

MCA was applied to the simple PDM of Equation 3 in a number of steps. The prior for the nuisance parameter MSYR was gamma(8.2, 372.7), and the likelihood for the observed total population in 1993 was $N(P_{1993}, 626^2)$. These choices were based on IWC consensus (IWC, 1995) and were the same as used in previous work (Raftery et al.[6]).

We assumed that we had a single observation from the likelihood for P_{1993}. In practice such an observation is usually obtained by means of a census. The observation is typically a maximum likelihood estimate of P_{1993}, therefore we denoted it by $\hat{P}_{1993}$.

An original conditional MLE was obtained by conditioning on a "likely" point estimate of MSYR, say MSYR_0. We chose $\text{MSYR}_0 = 0.02$, the mean of the prior for MSYR. The model was then run backwards (i.e. with $\hat{P}_{1993}$ and MSYR_0 as inputs) and the likelihood maximized. The resulting output was the conditional maximum likelihood estimate $\hat{K}_{MSYR_0}$ of K because (given MSYR_0) it lead to the value of P_{1993} that maximizes the likelihood.

The MCA estimation then proceeds as follows:

1. Draw $\hat{P}^*_{1993}$ from $N(\hat{P}_{1993}, 626^2)$. This is the parametric bootstrap from a distribution with mean given by the observed total population in 1993.
2. Draw MSYR^* from the prior for MSYR.
3. Obtain $\hat{K}_{MSYR^*}$ by running the model backwards with $\hat{P}^*_{1993}$ and MSYR^* as inputs.
4. Repeat steps 1–3 many times to form a collection of $\hat{K}_{MSYR^*}$ estimates. Like Punt and Butterworth (1997), we used 1000 replications.
5. Use $\hat{K}_{MSYR_0}$ and the distribution of the $\hat{K}_{MSYR^*}$ estimates to obtain inference about K. Specifically, the distribution of the $\hat{K}_{MSYR^*}$ estimates shows how the conditional MLE of K changes as MSYR is varied according to its prior.

For comparison with MCA, consider a fully Bayesian analysis which involves specifying priors for every model parameter, i.e. K, MSYR, and P_{1993}. This introduces an extra complication in that the input distributions and the model together induce a distribution on the output. There are thus two distributions (the specified prior and the induced distribution) on the output that need to be combined or reconciled in some manner. For our "backwards" implementation of the model here, the priors for MSYR and P_{1993} induce a prior on the output K. This issue has received considerable attention at the IWC and work in the area is ongoing. A possible solution involving logarithmic pooling of the two distributions is discussed in Raftery et al.[6] and Raftery and Poole.[7]

For this example, it was useful to compare a Bayesian approach with MCA, but avoiding the added complexity of the prior incoherence. This could be achieved if we simplified the Bayesian analysis slightly by ignoring the prior on the output K. We had a prior for MSYR and a likelihood for P_{1993} as we did in the MCA implementation above. In addition, we now also had a $N(7800, 1300^2)$ prior for P_{1993}. This was the prior used in Raftery et al.[6] and was again based on IWC consensus. The prior and likelihood were combined to yield a posterior distribution for P_{1993}. Because we had no data on MSYR, its prior was not updated to a posterior. The only operational difference between MCA and the Bayesian method was in the generation of values for P_{1993}: with MCA,

[6] Raftery, A. E., D. Poole, and G. H. Givens. 1996. The Bayesian synthesis assessment method: resolving the Borel Paradox and comparing the backwards and forwards variants. Paper SC/48/AS16 presented to the IWC Scientific Committee, June 1996.

[7] Raftery, A. E., and D. Poole. 1997. Bayesian synthesis methodology for bowhead whales. Paper SC/49/AS5 presented to the IWC Scientific Committee, October 1997.

values were bootstrapped from a distribution whose parameters were determined by maximum likelihood; with Bayes, they were sampled from the posterior.

Simulation results

Simulations were performed by using three sets of "true" parameters as shown in the first three columns of Table 2. The values of MSYR in these three sets of simulations correspond to the 0.5, 0.975, and 0.025 quantiles (respectively) of the gamma (8.2, 372.7) prior for this parameter. In this way, we investigated the performance of the method when the true MSYR is at the center and the boundaries of its prior 95% probability interval. In each case, a value of K was chosen such that extinction did not occur and P_{1993} was positive.

The first set of simulations represents a scenario where the mean of the prior for MSYR happens to coincide with the true value. If we use the prior mean as a point estimate (or "best guess") of MSYR, then our point estimate and the true value are the same. For this set, then, we would expect all assessment methods to provide accurate inference about K. The remaining two sets of simulations represent scenarios where our prior is inaccurate, i.e. the true MSYR is either larger or smaller than the prior mean (which remains unchanged). In these cases, this inaccuracy is naturally going to cause the resulting distribution of the output K to be biased. All assessment approaches will be affected by this bias, particularly with respect to their point estimates of K. Indeed, these parameter sets represent situations that all assessment scientists would like to avoid. The key point of interest is the extent to which a method can be insensitive to poor prior information and still provide somewhat reliable inference.

The results of the simulations are shown in Table 2. For each of the three sets of true parameters, the MCA analysis was run three times by using the 0.025, 0.5, and 0.975 quantiles of the normal likelihood as the observed 1993 population, $\hat{P}_{1993}$. Then, this entire simulation design was replicated 500 times. The quantiles shown are MCA medians for $\hat{K}_{MSYR^*}$ across the 500 replicates, and the coverage rates (last two columns) show the percentage of the 500 replicates for which the estimated 95% MCA or Bayes confidence interval covered the truth.

In the first set of simulations, where the true MSYR and the prior mean $MSYR_0$ were exactly equal, the conditional MLE $\hat{K}_{MSYR_0}$ was very accurate. This is to be expected in this optimistic (if somewhat unlikely) scenario. The MCA confidence intervals provided by $\hat{K}_{MSYR^*}$ covered the true K in all cases, as did the confidence intervals obtained with the Bayesian method. Also, the estimation was fairly insensitive to the accuracy of the estimate $\hat{P}_{1993}$. A difference of 1227 whales in the estimate of P_{1993} (i.e. two standard deviations) resulted in $\hat{K}_{MSYR0}$ changing by less than 100 whales.

In the second set, the true value of MSYR was greater than the prior mean. This resulted in K having been overestimated. Here, both MCA and Bayesian results were biased by the use of the same bad prior, but the MCA coverage was worse. If follows that the suboptimal behavior of MCA cannot be attributed solely to the choice of prior. The 95% MCA intervals provided poor coverage of the truth, worse when $\hat{P}_{1993}$ was accurate than when it was too low. For the Bayesian method, this difference was not as great. The coverage of the Bayesian intervals was somewhat better in two cases, particularly when $\hat{P}_{1993}$ was too large.

In the final set, the true value of MSYR was at the low end of the prior interval, and we observed that K

Table 2

Simulation results for the simple bowhead whale population dynamics model (PDM).

Parameter set					MCA quantiles, $\hat{K}_{MSYR^*}$			Coverages	
MSYR	K	P_{1993}	$\hat{P}_{1993}$	$\hat{K}_{MSYR_0}$	0.025	0.5	0.975	MCA (%)	Bayesian (%)
0.02	14,700	8,733	7,506	14,640	11,230	14,360	18,870	100	100
			8,733	14,700	11,260	14,400	19,160	100	100
			9,960	14,780	11,330	14,450	19,540	100	100
0.04	11,300	9,896	8,669	14,700	11,250	14,400	19,120	68	66
			9,896	14,780	11,290	14,470	19,450	55	63
			11,123	14,880	11,580	14,560	19,840	0	23
0.01	18,500	6,971	5,744	14,570	11,270	14,300	18,430	38	54
			6,971	14,620	11,260	14,340	18,710	80	88
			8,198	14,670	11,270	14,380	19,000	99	99

was underestimated by about 4000 whales for all three values of $\hat{P}_{1993}$. The estimates of K were inaccurate regardless of the accuracy of $\hat{P}_{1993}$. Although not shown, this inaccuracy holds for the Bayesian method as well as for MCA. Both methods provided the best coverage when $\hat{P}_{1993}$ was too large, but the Bayesian method provided better coverage than MCA in the other two cases. Again, these results separated coverage problems attributable to inaccurate priors from additional performance degradation apparently introduced by opting for MCA analysis.

As a final point, these simulations and others that we ran suggest that the MCA estimates of K are heavily dependent on the prior distribution for MSYR, to the point of being undesirably insensitive to the true values of K, P_{1993}, and the data $\hat{P}_{1993}$. This behavior is more extreme than would be the case if the prior was updated to a posterior in a fully Bayesian framework. These results concur with the results in the section "A simple point estimation example," where the MCA estimates were greatly influenced by the prior mean and variance. When the prior is accurate and precise, MCA may perform well, as will Bayesian techniques. Bayesian techniques seem to weight the data more heavily in relation to the prior, than does MCA.

Relation between MCA and bootstrap

The MCA approach described in the section "Implementation of MCA and a Bayesian approach" included bootstrapping the abundance data unconditionally. A standard (conditional) bootstrap would proceed by resampling the residuals (e.g. Efron and Gong, 1983, p. 43) from the model fit. The unconditional approach used in the bowhead application introduced excess variability because bootstrapped pseudodata varied about the observed data, which in turn varied about the model fit.

The general MCA approach might be viewed as an approximation to a bootstrap that is unconditional on the model fit. In such applications (e.g. Smith and Gavaris, 1993), the interpretation of the stochasticity thereby introduced must be carefully considered if it differs from the data stochasticity that causes estimation uncertainty. When, as is permitted with MCA, the unconditional approach simulates from a subjective prior rather than from data, the method is not a bootstrap because the simulation reflects stochasticity other than that introduced by data used for estimating the parameter.

Even in the case when sufficient data are available to permit parametric bootstrap simulation of all inputs (case 1 in the section "The Monte Carlo approach"), MCA does not reduce to a parametric bootstrap of the desired estimator. A parametric bootstrap expresses sampling uncertainty about a statistic $R(\boldsymbol{X}, F(\theta))$, where $\boldsymbol{X} \sim F$, by observing the distribution of $R(\boldsymbol{X}^*, \hat{F}(\theta))$, where $\hat{F}$ is an estimate of F that depends on the data $\boldsymbol{X}$, and $\boldsymbol{X}^* \sim \hat{F}$. Variability in R is due to $\boldsymbol{X}$. A parametric bootstrap arises when a model $\hat{F}(\theta) = F(\theta)$ is fitted, or less desirably $\hat{F}(\hat{\theta}) = G(\hat{\gamma})$ in some applications of MCA. In this case, $\hat{\theta}$ or $\hat{\gamma}$ should be estimated from the data, $\boldsymbol{X}$, whose stochasticity induces sampling variability in $R(\boldsymbol{X}, F(\theta))$. However, with MCA, $\hat{\theta}$ or $\hat{\gamma}$ is estimated from different data, not the data on model output parameters, although it is the uncertainty associated with estimators of output parameters that is desired. Even in this case, the sampling distributions used are effectively data-based priors, and MCA relies on the unusual approach of integrating a conditional maximization of the likelihood over the prior.

Conclusion

Theoretical investigation and simulation show that the combination of Bayesian and conditional maximum likelihood techniques used by MCA has the potential to yield quite variable or biased results, or both, though it can perform well in ideal circumstances. In some situations, a fully Bayesian or classical ML solution can be obtained by small modifications to MCA, and the optimal properties of these more standard methods are well known. For more complex problems, Bayesian and ML solutions are sometimes more difficult to obtain than is an MCA solution. However, as our examples illustrate, MCA can result in unreliable inference even in simple situations. MCA integrates a conditional maximization of the likelihood over the prior, whereas a fully Bayesian approach integrates a conditional mean. If one uses what is effectively a Bayesian prior, then it is suboptimal to use it in a non-Bayesian inference framework.

We have seen how MCA produces estimates with excessive bias. However, there may exist classes of assessment problems where, owing to some feature that is identifiable in advance, the extra bias is acceptably small. MCA could be applied to such problems because the excess bias would not cause MCA results to differ much from results produced by either fully Bayesian or ML methods. Our bowhead whale and simple examples clearly do not belong to such a class. Furthermore, in the general case, the extent to which MCA might err is not controllable or estimable by the analyst. Although MCA can produce good estimates in some applications, a method that can also go badly wrong is risky when one does not know the extent of problems in any particular application. When priors are used, the usefulness of Bayesian approaches (when obtainable) would seem to be greater than MCA, because the effect of the priors is washed out with increasing data (contrary to what happens with MCA). When priors are not required, either Bayesian or MLE techniques might be usefully applied.

Acknowledgments

This research was supported by the North Slope Borough, Alaska, and the National Oceanic and Atmospheric Administration (through the National Marine Mammal Laboratory and the Alaska Eskimo Whaling Commission). The authors are grateful to Victor Restrepo and John Hoenig for comments on an earlier draft and would also like to thank Doug Butterworth, Carey Priebe, Andre Punt, Tore Schweder, Judy Zeh, and four anonymous reviewers for their helpful comments and suggestions. However, the views expressed are solely the authors'.

Literature cited

Butterworth, D. S., and A. E. Punt.
1995. On the Bayesian approach suggested for the assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales. Rep. Int. Whal. Comm. 45:303–11.

Efron, B., and G. Gong.
1983. A leisurely look at the bootstrap, the jackknife, and cross-validation. The American Statistician 37:36–48.

IWC (International Whaling Commission).
1995. Report of the Scientific Committee. Rep. Int. Whal. Comm. 45:53–222.

Jeffreys, H. S.
1961. Theory of probability (3rd edition). Oxford Univ. Press, Oxford, 447 p.

Lee, P. M.
1997. Bayesian statistics: an introduction (2nd edition). John Wiley and Sons, New York, NY, 344 p.

Punt, A. E., and D. S. Butterworth.
1996. Further remarks on the Bayesian approach for assessing the Bering-Chukchi-Beaufort Seas stock of bowhead whales. Rep. Int. Whal. Comm. 46:481–91.
1997. Assessments of the Bering-Chukchi-Beaufort Seas stock of bowhead whales (*Balaena mysticetus*) using maximum likelihood and Bayesian methods. Rep. Int. Whal. Comm. 47:603–618.

Raftery, A. E., G. H. Givens, and J. E. Zeh.
1995. Inference from a deterministic population dynamics model for bowhead whales (with discussion and rejoinder). J. Am. Statist. Assoc. 90:402–430.

Reid, N.
1995. The roles of conditioning in inference. Statistical Science 10:138–157.

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

Restrepo, V. R., J. E. Powers, and S. C. Turner.
1991. Incorporating uncertainty in VPA results via simulation. Collect. Vol. Sci.. Pap. ICCAT 35:355–61.

Smith, S. J., and S. Gavaris.
1993. Evaluating the accuracy of projected catch estimates from sequential population analysis and trawl survey abundance estimates. *In* S. J. Smith et al. (eds.), Risk evaluation and biological reference points for fisheries management, p. 163–172. Can. Spec. Publ. Fish. Aquat. Sci. 120.