



**Inference from a Deterministic Population Dynamics Model for Bowhead Whales**

Adrian E. Raftery; Geof H. Givens; Judith E. Zeh

*Journal of the American Statistical Association*, Vol. 90, No. 430 (Jun., 1995), 402-416.

Stable URL:

<http://links.jstor.org/sici?sici=0162-1459%28199506%2990%3A430%3C402%3AIFADPD%3E2.0.CO%3B2-O>

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

*Journal of the American Statistical Association* is published by American Statistical Association. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/astata.html>.

---

*Journal of the American Statistical Association*  
©1995 American Statistical Association

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact [jstor-info@umich.edu](mailto:jstor-info@umich.edu).

©2003 JSTOR

# Inference from a Deterministic Population Dynamics Model for Bowhead Whales

Adrian E. RAFTERY, Geof H. GIVENS, and Judith E. ZEH\*

---

We consider the problem of inference about a quantity of interest given different sources of information linked by a deterministic population dynamics model. Our approach consists of translating all the available information into a joint premodel distribution on all the model inputs and outputs and then restricting this to the submanifold defined by the model to obtain the joint postmodel distribution. Marginalizing this yields inference, conditional on the model, about quantities of interest, which can be functions of model inputs, model outputs, or both. Samples from the postmodel distribution are obtained by importance sampling and Rubin's SIR algorithm. The framework includes as a special case the situation where the pre-model information about the outputs consists of measurements with error; this reduces to standard Bayesian inference. The results are in the form of a sample from the postmodel distribution and so can be examined using the full range of exploratory data analysis techniques. Methods for comparing competing population dynamics models are developed, based on a generalization of the Bayes factor idea. A key quantity used by the International Whaling Commission (IWC) in making decisions about bowhead whales, *Balaena mysticetus*, is the replacement yield, RY. Information about the species is of three main types: recent census information, historical catch records, and evidence about birth and death rates. These are combined using a special case of the Leslie matrix population dynamics model. Our method yields full inference about RY and also sheds light on other, sometimes controversial, questions of scientific interest. These ideas are also applicable to many simulation models in other areas of science and policy making. Software to implement these methods is available from StatLib.

KEY WORDS: Bayes factor; Bayesian synthesis method; Importance sampling; Leslie matrix; SIR algorithm.

---

## 1. INTRODUCTION

The bowhead whale, *Balaena mysticetus*, is a large baleen whale. The Western Arctic (Bering–Chukchi–Beaufort Seas) stock of bowheads is the largest remaining one. It has been the object of subsistence hunting on a small scale by the Eskimo peoples of the area for many centuries. The stock was discovered by Yankee whalers in 1848, and the ensuing massive commercial hunt greatly reduced its numbers. Over time, the increased difficulty and cost of whaling due to depletion of the stock, the replacement of whale oil by kerosene, and the decreased use of baleen corset stays discouraged commercial whaling and led to its de facto end in 1915. Since 1946, the stock has been legally protected from commercial whaling by the International Convention for the Regulation of Whaling.

The Convention does, however, allow continued limited aboriginal subsistence whaling, with a quota to be set by the International Whaling Commission (IWC) at a level low enough to allow the stock to recover from the effects of commercial whaling. A key quantity for setting the quota in a given year is the *replacement yield* (RY) in that year, namely, the greatest number of whales that could be taken without the population decreasing. This is unknown and is subject to considerable uncertainty. Because it is important that the quota not exceed this unknown value, a conservative value or “lower bound” is sought. The IWC Scientific Committee (SC) has interpreted this to mean the lower 5% point of a

posterior distribution of RY given available information (IWC 1992); this should reflect all important sources of uncertainty.

There are three main types of information relevant to RY: recent surveys, historical whaling records, and biological information. Recent surveys yield current and recent estimates of abundance and rates of increase, historical whaling records yield a time series of whaling kills, and the biological information provides knowledge about birth and death rates. We now describe these three sources of information.

The recent surveys started in 1978. Since then, there has been an effort to count the whales every spring as they migrate from the Bering Sea to the Beaufort Sea. At Point Barrow they pass close to shore, and this is where the counting takes place. From 1978 to 1984 there were only visual counts by ice-based observers. From 1985 onwards these counts were supplemented by data on times and locations of bowhead vocalizations, recorded using hydrophones, throughout the migration period. The data from these censusing efforts have yielded a posterior distribution for the 1988 population size (Raftery and Zeh 1993) and an estimate of the rate of increase in the bowhead population from 1978 to 1988, which we denote by ROI (Zeh, George, Raftery, and Carroll 1991). The historic kill record has been validated by the IWC (IWC, 1992).

The biological information about birth and death rates comes from photoidentification (Miller et al. 1992); examination of dead animals (George, Philo, Suydam, Tarpley, and Albert 1992); photogrammetric data on the proportions of mature animals and calves in the population (Withrow and Angliss, 1992); photogrammetric data on growth rates (Koski, Davis, Miller, and Withrow 1992); estimates of age derived from variations in  $\delta^{13}\text{C}$  carbon isotope ratios in baleen (Schell, Saupe, and Haubenstock 1989; Withrow, Burke, Jones, and Brooks 1992), which provide information about

---

\* Adrian E. Raftery is Professor of Statistics and Sociology and Judith E. Zeh is Research Associate Professor of Statistics, Department of Statistics, University of Washington, Seattle, WA 98195. Geof H. Givens is Assistant Professor of Statistics, Department of Statistics, Colorado State University, Fort Collins, CO 80523. This research was supported by the North Slope Borough, Alaska, the State of Alaska (through the Alaska Department of Fish and Game), and the National Oceanic and Atmospheric Administration (through the National Marine Mammal Laboratory and the Alaska Eskimo Whaling Commission). The authors are grateful to Jeff Breiwick and Charles Fowler for helpful discussions, and to the editor, the associate editor, and two anonymous referees whose very detailed comments greatly improved the manuscript.

age at sexual maturity; and knowledge of other whale species (Best 1991). Most of this biological information is very imprecise.

It is common practice to combine all this information using a *population dynamics model* (PDM), which itself embodies additional biological information. The PDM's used by the IWC are deterministic accounting models in which births are added and natural deaths and kills are subtracted (Sec. 3). They require as inputs age-specific natural mortality and fertility rates that may be density-dependent, the kill record, and the initial population size, and they output the population for each year, broken down by age and sex.

Assigning values to a PDM's many inputs is hard, and taking account of the uncertainty associated with them is even harder. A typical approach is informal, as follows. A first guess at the input parameters is made and the model is run. If the corresponding outputs seem implausible, then the inputs are adjusted and the model rerun; this is repeated until a satisfactory set of inputs has been found. Finally, a sensitivity analysis is performed in which the inputs are varied; this involves several runs of the model with different combinations of "central," "high," and "low" values of the inputs. If the outputs of primary interest are relatively insensitive to the inputs, then all is well. This is often not the case, however, and sometimes the range of values of the outputs of interest is used informally as a rough confidence interval.

There is little justification for such an interpretation. At the 1991 IWC SC meeting, the bowhead whale subcommittee proposed a lower bound for RY based on this approach, but this was rejected by the full SC on the grounds that it was not statistically valid. As a temporary measure, the SC based its recommendations on an estimate of the lower bound for RY that ignored the biological information and took account only of the current abundance information. Recognizing this to be unsatisfactory, the SC recommended the development of new methods that would enable statistically valid inferences about RY to be based on all the available information. This article describes our effort to carry out this recommendation.

Our approach starts by expressing available information about both model inputs and outputs as a joint probability distribution which we call the *premodel distribution*. The model restricts the premodel distribution to a submanifold; the restricted distribution is called the *postmodel distribution*. Inference about any quantity of interest is made by marginalizing the postmodel distribution, from which a sample is drawn by importance sampling (Sec. 2). This yields inference about the main quantity of interest, RY, as well as other quantities of biological interest, in the process shedding some light on recent controversies (Sec. 5). The approach provides informal ways of checking that the model is not in conflict with the other available information. Competing models may also be compared using Generalized Bayes Factors (Sec. 2.4).

The approach may be viewed as a generalization of standard Bayesian inference to deal with deterministic simulation models. It includes standard Bayesian inference as a special case; see Section 2.3. Givens, Raftery, and Zeh (1993) discussed the advantages of the present approach relative to the

ways of drawing conclusions from deterministic models that have been standard in past work of the IWC SC. Software to implement the methods is available from StatLib; see Section 6.4.

Special attention is paid to the specification of the premodel distribution. At its 1991 meeting, the IWC SC carried out a comprehensive assessment of the status of bowhead whales. The resulting report (IWC 1992) describes current scientific beliefs about model inputs and outputs in considerable detail, together with the basis for them. The report is remarkable in documenting how reasonable agreed prior distributions can emerge from often conflicting expert opinions; see Section 4.

This article is just one example of the general problem of inference from simulation models, which statisticians only recently have begun to consider seriously. Simulation models are different in concept from standard statistical models such as linear regression: they attempt to capture the underlying mechanisms explicitly, are often deterministic, and tend to involve large numbers of inputs and outputs and complex relationships between them. Thus properly taking account of parameter and model uncertainty is even more crucial than with standard statistical models, but it is rarely done.

Speed (1983) was the first to point out the importance of this area for statisticians, spurred by his observation that scientists at CSIRO (Australia's national research organization) were using simulation models more often, abandoning statistical models in the process. In addition to PDM's, he gave as examples a sheep growth and production simulation model, a model for predicting the nitrogen fertilizer requirement of wheat crops, a water-balance model, and a model describing the effects of light and vapor on apple leaves. There *has* been much work on model validation (see, for example, Caswell 1976 for applications in ecology, Hughes 1984 for military applications, and Guttorp and Walden 1987 for applications in geophysics).

Far less attention has been paid to formally taking account of uncertainty about inputs, although Hodges (1987, 1991) and Bankes (1991) have argued strongly that it is vital to do so, based on extensive experience with simulation models at the RAND Corporation. Bankes (1991) has emphasized and deplored the tendency of such models to be large and to continue growing throughout their history; our suggestion at the end of Section 5 indicates how the present approach may be used to *simplify* a complex model.

## 2. INFERENCE FROM DETERMINISTIC SIMULATION MODELS

### 2.1 The Premodel and Postmodel Distributions

We denote by  $\theta$  the set of model inputs and by  $\phi$  the set of model outputs about which we have information independent of the model. We denote by  $\psi$  the set of quantities of interest, which may be model inputs, model outputs, or functions of both and may or may not be included in  $\phi$  or  $\theta$ . We denote by  $p(\theta, \phi)$  the joint *premodel distribution* of  $\theta$  and  $\phi$ , which summarizes all available information about  $\theta$  and  $\phi$  except that embodied in the model itself. In the bowhead whale application, model inputs include birth and death

rates, model outputs include current population size, and a main quantity of interest is the replacement yield; see Section 3.

The model defines a mapping from the set of possible values of  $\theta$  to the set of possible values of  $\phi$ ; we denote this mapping by  $\theta \rightarrow \Phi(\theta)$ . Thus given the model, the joint distribution of  $\theta$  and  $\phi$  is simply the restriction of the premodel distribution to the submanifold  $\{(\theta, \phi) : \phi = \Phi(\theta)\}$ , namely

$$\begin{aligned} \pi(\theta, \phi) &\propto p(\theta, \Phi(\theta)) && \text{if } \phi = \Phi(\theta), \\ &\propto 0 && \text{otherwise.} \end{aligned} \tag{1}$$

We refer to  $\pi(\theta, \phi)$  as the *postmodel distribution*. The marginal postmodel distribution of  $\theta$  is then

$$\pi^{[\theta]}(\theta) \propto p(\theta, \Phi(\theta)). \tag{2}$$

For marginal and conditional distributions, we use superscripts in square brackets to show to what the distribution applies. Thus, for example,  $p^{[\theta]}(\cdot)$  denotes the marginal premodel distribution of the inputs,  $\pi^{[\phi]}(\cdot)$  denotes the marginal postmodel distribution of the outputs, and  $p^{[\phi|\theta]}(\cdot|\cdot)$  denotes the conditional premodel distribution of the outputs given the inputs.

Inference about  $\psi$  is based on its postmodel distribution,

$$\Pr_{\text{post}}[\psi \in B] = \int_{H(B)} \pi^{[\theta]}(\theta) d\theta, \tag{3}$$

where  $H(B) = \{\theta : \Psi(\theta, \Phi(\theta)) \in B\}$  for any set  $B$  and  $\theta \rightarrow \Psi(\theta, \Phi(\theta))$  denotes the mapping from  $\theta$  to  $\psi$  induced by the model.

### 2.2 An Importance Sampling Approach

We evaluate the integral in Equation (3) by importance sampling. In general, importance sampling provides a way of evaluating the ratio of integrals

$$I = \frac{\int_{\Gamma} h(u) du}{\int_{\Gamma} g(u) du}, \tag{4}$$

where  $h$  is a function whose integral over its domain is finite and nonzero and  $\Gamma$  is a set. We choose a random variable with the same domain as  $h$  and probability density function  $g$  such that  $g(u)/h(u)$  exists for all  $u$  in the domain. We simulate  $k$  points  $u_1, \dots, u_k$  from  $g$  and form importance sampling weights  $r_i \propto h(u_i)/g(u_i)$ . We then estimate  $I$  by

$$\hat{I} = \frac{\sum_{i:u_i \in \Gamma} r_i}{\sum_{i=1}^k r_i}. \tag{5}$$

Under quite general regularity conditions,  $\hat{I}$  is a consistent estimator of  $I$  in the sense that  $\hat{I} \rightarrow I$  in probability as  $k \rightarrow \infty$ .

We apply this to Equation (3) by replacing  $u$  by  $\theta$ , replacing  $h$  by  $\pi^{[\theta]}$ , and replacing  $\Gamma$  by  $H(B)$ . We use the premodel marginal distribution of  $\theta$  as the importance sampling function, so that we replace  $g(u)$  by  $p^{[\theta]}(\theta)$ ; this often will be easy to simulate from, especially if most of the components of  $\theta$  are premodel independent. Thus the importance sampling weights are  $r_i = p^{[\phi|\theta]}(\Phi(\theta_i)|\theta_i)$ . These will generally be easy to calculate; for example, if  $\theta$  and  $\phi$  are premodel

independent, then  $r_i = p^{[\phi]}(\Phi(\theta_i))$ , which is just the premodel density of  $\phi$  evaluated at  $\Phi(\theta_i)$ . Then  $\Pr_{\text{post}}[\psi \in B]$  can be evaluated using Equation (5).

The discrete distribution with values  $\{\theta_i\}$  and associated probabilities proportional to  $r_i$  is an estimate of the postmodel marginal distribution of  $\theta$ ,  $\pi^{[\theta]}(\theta)$ . An approximate sample from  $\pi^{[\theta]}(\theta)$  may be obtained by drawing a second-stage sample of size  $l$  from this discrete distribution. Proposed by Rubin (1987, 1988), this is known as the SIR algorithm. The approximation becomes exact as  $k/l \rightarrow \infty$ .

The premodel distribution of  $\theta$  is usually much more diffuse than its postmodel distribution, and so  $p^{[\theta]}(\theta)$  is not a good importance sampling function. Because of this, many draws from  $p^{[\theta]}(\theta)$  are needed to yield a good estimate, and each value drawn requires a separate run of the model. Fortunately, our PDM runs quickly, so this is not a problem. We have had good results with  $k = 200,000$  first-stage samples and  $l = 5,000$  second-stage samples.

### 2.3 Use as an Estimation Tool

There are several points to note about this approach:

1. It enables one to take into account information about model outputs as well as model inputs, and the information about model outputs need not be independent of the information about model inputs.

2. It includes as a special case the situation where the premodel information about  $\phi$  consists of observations of  $\phi$  with measurement error, denoted by  $D$ . Then  $p(\theta, \phi) \propto p^{[D|\phi]}(D|\phi)p^{[\theta]}(\theta)$ , where  $p^{[\theta]}(\theta)$  denotes the premodel evidence about  $\theta$  and  $\pi^{[\theta]}(\theta) \propto p^{[D|\phi]}(D|\Phi(\theta))p^{[\theta]}(\theta)$ . This falls within the framework of standard Bayesian inference, in that  $p^{[\theta]}(\theta)$  may be regarded as the prior distribution of  $\theta$ ,  $D$  as the data,  $p^{[D|\phi]}(D|\Phi(\theta))$  as the likelihood of  $\theta$ , and  $\pi^{[\theta]}(\theta)$  as the posterior distribution of  $\theta$ . Our framework is more general, in that prior information about the model outputs or predictions,  $\phi$ , can also be included by replacing  $p^{[\theta]}(\theta)$  by  $p(\theta, \phi)$ , and the prior information about the outputs  $\phi$  is not constrained to be independent of the information about the inputs  $\theta$ .

As an example, consider the highly simplified (and unrealistic) PDM where the population increases at a known rate  $r$  per year, so that  $P_t = (1 + r)^t P_0$ . The only input is  $\theta = P_0$ . Suppose that evidence exists only about the single output  $\phi = P_t$  and that the quantity of interest is  $\psi = P_0$ . Then the model specifies that  $\phi = a\theta$ , where  $a = (1 + r)^t$ . The premodel information consists of biological information about  $\theta = P_0$  represented by the distribution  $p^{[\theta]}(\theta)$ , and a survey that led to an estimate  $D$  of  $\phi = P_t$  such that  $D = \phi + \varepsilon$ , where  $\varepsilon \sim N(0, \sigma^2)$ . The premodel information about  $\theta$  is independent of that about  $\phi$ . The premodel distribution is then  $p(\theta, \phi) \propto p^{[\theta]}(\theta)N(D; \phi, \sigma^2)$ , and the postmodel distribution is the restriction of this to the subspace  $\{(\theta, \phi) : \phi = a\theta\}$ , namely  $\pi(\theta, \phi) \propto p^{[\theta]}(\theta)N(D; a\theta, \sigma^2)$ .

If this were treated as a standard Bayesian inference problem with parameter  $\theta$ , data  $D$ , and likelihood given by  $D \sim N(a\theta, \sigma^2)$ , then the prior for  $\theta$  would be  $p^{[\theta]}(\theta)$  and the posterior density of  $\theta$  would be proportional to  $p^{[\theta]}(\theta)N(D; a\theta, \sigma^2)$ . Thus the posterior distribution is equal to the post-

model distribution. But the premodel distribution is *not* the same as the prior distribution; rather, it is a version of the prior times the likelihood that is unrestricted by the model structure.

3. Changing the premodel distribution: Suppose that one has carried out this procedure for one premodel distribution,  $p_1(\theta, \phi)$ , and wants results for another,  $p_2(\theta, \phi)$ . One may, of course, rerun the procedure from scratch, replacing  $p_1$  by  $p_2$ . As suggested by Smith and Gelfand (1992), one may also resample from the final sample of size  $l$  obtained the first time with  $p_1$ , using weights  $p_2(\theta, \Phi(\theta))/p_1(\theta, \Phi(\theta))$ . This is more convenient, because it does not require any new runs of the model. It will work well if  $p_2$  is less diffuse than  $p_1$  and is "covered" by it. This can be used to investigate sensitivity to the premodel distribution, because if the whole procedure is carried out initially with a more diffuse premodel distribution, then results for any more concentrated premodel distribution within its range can be calculated easily without requiring any more model runs. Software to do this for the bowhead whale PDM has been developed, as described by Givens et al. (1994). The software is available from StatLib; see Section 6.4.

## 2.4 Model Comparison Using Generalized Bayes Factors

Suppose that we have two competing PDM's. We will generalize the standard Bayes factor to compare such models. The standard Bayes factor (Jeffreys 1961) for comparing two models  $M_1$  and  $M_2$  given data  $D$  is  $p(D|M_1)/p(D|M_2)$ , the ratio of the (predictive) probabilities of the data under the two models. Thus the Bayes factor compares how well the two models predict the data.

With PDM's, we compare competing models on the basis of how well they predict the outputs. If we knew the outputs exactly (i.e., if we knew that  $\phi = \phi^*$ , say), then the corresponding ratio would be  $q_1(\phi^*)/q_2(\phi^*)$ , where  $q_j(\phi^*)$  is the predictive probability of  $\phi^*$  under the PDM  $M_j$  ( $j = 1, 2$ ). This is

$$q_j(\phi^*) = \int_{C_j(\phi^*)} p_j^{[0]}(\theta) d\theta, \quad (6)$$

where  $C_j(\phi^*) = \{\theta : \Phi_j(\theta) = \phi^*\}$ ; the subscript  $j$  denotes the PDM  $M_j$  for  $j = 1, 2$ .

But we do not usually know  $\phi$  exactly; rather, we have uncertain information about its value, represented by its premodel distribution. Let  $E$  denote the marginal premodel information about  $\phi$  represented by the distribution  $p_j^{[\phi]}(\phi)$ . Then  $q_j(\phi^*)$  is replaced by

$$\begin{aligned} p_j(E) &= \int q_j(\phi) p_j^{[\phi]}(\phi) d\phi \\ &= \int p_j^{[\phi]}(\Phi_j(\theta)) p_j^{[0]}(\theta) d\theta, \end{aligned} \quad (7)$$

by Equation (6) and the fact that  $p^{[0]}(\theta) = \int p(\theta, \phi) d\phi$ . For this comparison to be valid, it must be based on the same outputs for the two PDM's, but not necessarily on the same inputs. The marginal premodel distribution of the outputs

should also be the same for the two models. It can be shown that  $p_j(E) = \int_{C_j} p(\theta, \phi) d\theta d\phi$ , where  $C_j = \{(\theta, \phi) : \Phi_j(\theta) = \phi\}$ , so that  $p_j(E)$  is just the total joint premodel probability of inputs and outputs given the model. This implies that we compare models on the basis of how precisely we can state their inputs and how well they predict the outputs.

We call the ratio  $G_{12} = p_1(E)/p_2(E)$  the Generalized Bayes Factor for the PDM  $M_1$  against the PDM  $M_2$ . By analogy with widely used rules of thumb for standard Bayes factors (Jeffreys, 1961, Appendix B), we may view the comparison as providing evidence for  $M_1$  against  $M_2$  not worth more than a bare mention if  $G_{12}$  is between 1 and  $\sqrt{10}$ , positive evidence if  $G_{12}$  is between  $\sqrt{10}$  and 10, strong evidence if  $G_{12}$  is between 10 and 100, and very strong evidence if  $G_{12}$  exceeds 100.

A simple Monte Carlo estimate of the integral in Equation (7) is

$$\hat{p}_j(E) = \frac{1}{k} \sum_{i=1}^k p_j^{[\phi]}(\Phi_j(\theta_i)), \quad (8)$$

where the  $\theta_i$  form a sample from the marginal premodel distribution of  $\theta$  under the PDM  $M_j$ . Other estimates based on more general importance sampling numerical integration algorithms for the integral in Equation (7) may be obtained by analogy with the estimates of (ordinary) Bayes factors from posterior simulation of Newton and Raftery (1991; 1994, sec. 7). Some of these are given explicitly by Raftery et al. (1992).

Raftery et al. (1992) described how to compute Generalized Bayes Factors from Equation (8) for comparing  $M_1$ , the Leslie matrix PDM considered in this article, with  $M_2$ , a general-purpose PDM called Hitter-Fitter often used by the IWC (de la Mare 1989). The Generalized Bayes Factor is  $G_{12} = 1.5$ , indicating that the data favor the Leslie matrix model over the Hitter-Fitter model, but only to a very small extent.

Givens, Zeh, and Raftery (1993) used this the approach to compare  $M_2$  with a third model  $M_3$ , in which the original Western Arctic bowhead whale population in 1848 consisted of two subpopulations with different summer feeding grounds. Commercial whaling would have annihilated one subpopulation by about 1870. Their result was  $G_{32} = 1.7$ , providing some (but very weak) support for the two subpopulation hypothesis.

Generalized Bayes Factors can be used to take account of uncertainty about model structure, as distinct from the uncertainty about model inputs with which we are mainly concerned in this article; see Section 2.5.

## 2.5 Model Uncertainty

In this article we focus on accounting for uncertainty about the inputs of a PDM conditionally on its structure. But uncertainty about its structure may also be important. At least three kinds of uncertainty about model structure may be important: uncertainty about which of several competing PDM's is most appropriate, uncertainty about a particular assumption of the model, and uncertainty about whether there might be some unsuspected but serious error in the model or the calculations.

If there are two competing PDM's, we have shown in Section 2.4 how they may be compared using Generalized Bayes Factors. If this comparison is not decisive, a combined inference can be based on a weighted average of the postmodel distributions from the two models, with the ratio of the weights equal to the Generalized Bayes Factor. This accounts for uncertainty about the structure of the PDM and carries over to the case where there are more than two models. It is analogous to the standard Bayesian way of accounting for model uncertainty in standard statistical models (Kass and Raftery 1995); this often has better out-of-sample predictive ability than inference based on a single selected model (Madigan and Raftery, 1994; Raftery et al. 1993).

An important special case of this arises when there is uncertainty, or disagreement, about a particular model assumption and when the results are sensitive to that assumption. Then a weighted average of the resulting postmodel distributions can lead to a generally acceptable combined postmodel distribution. (For an example of this in the IWC context, see Raftery and Schweder 1993.)

A major, but little discussed, source of uncertainty is the possibility that the PDM itself, or the calculations based on it, may be in serious error for currently unsuspected reasons. This may be of particular concern if the PDM is relatively new or untested; Wade (1994) gave several striking recent examples from other fields of science. Policy makers may wish to take account of this possibility when making decisions; a simple way of doing so is based on the following thought experiment. Suppose we learned that the model was seriously flawed, but that there was no time to build a new model before a decision had to be made; then it might well be reasonable to revert to the premodel distribution and base decisions on that. If such errors occur with probability  $\epsilon$ , then this suggests using the "hyper-postmodel distribution,"  $\pi^*(\theta, \phi) = (1 - \epsilon)\pi(\theta, \phi) + \epsilon p(\theta, \phi)$ . Samples from  $\pi^*(\theta, \phi)$  are available from our importance sampling method with little additional effort.

A similar argument can be applied to standard statistical models, yielding a "hyper-posterior" distribution that is a mixture of the posterior and prior distributions. Schnute and Hilborn (1993) have suggested a different, non-Bayesian way of accounting for this source of uncertainty.

### 3. THE POPULATION DYNAMICS MODEL

#### 3.1 Definition of the Model

The PDM that we consider was developed for bowhead whales by Breiwick, Eberhardt, and Braham (1984) and is a special case of the well-known one-sex age-structured Leslie matrix population projection model (Leslie 1945, 1948; Lewis 1942). A fairly general form of this model is as follows. Let  $n_{xt}$  be the number of females aged  $x$  next birthday on January 1 of calendar year  $t$ , where  $t = 0$  is the initial year, here 1848. Then the model is specified by the equations

$$n_{1,t+1} = \sum_{x=1}^{\infty} f_{xt}(n_{xt} - c_{xt}) \quad (9)$$

and

$$n_{x+1,t+1} = s_{xt}(n_{xt} - c_{xt}) \quad (x = 1, 2, 3, \dots), \quad (10)$$

where  $f_{xt}$  is the average number of female calves that survive to age 1 born in year  $t$  to a female aged  $x$ ,  $s_{xt}$  is the natural survival rate of females aged  $x$  in year  $t$ , and  $c_{xt}$  is the number of females aged  $x$  killed by hunting in year  $t$ .

This can be written in matrix form as

$$\mathbf{N}_{t+1} = \mathbf{A}_t(\mathbf{N}_t - \mathbf{C}_t), \quad (11)$$

where  $\mathbf{N}_t = (n_{1t}, n_{2t}, \dots)^T$ ,  $\mathbf{C}_t = (c_{1t}, c_{2t}, \dots)^T$ , and  $\mathbf{A}_t = (\mathbf{A}_{txy} : x, y = 1, 2, \dots)$  is a doubly infinite square matrix defined by

$$\begin{aligned} \mathbf{A}_{txy} &= f_{yt} & \text{if } x = 1, \\ &= s_{yt} & \text{if } x = y + 1, \\ &= 0 & \text{otherwise.} \end{aligned} \quad (12)$$

As it stands, the model has an infinite number of parameters, and Breiwick, et al. (1984) proposed the following restrictions for the bowhead whale case.

*Mortality.* It is assumed that an immature survival rate,  $s_0$ , prevails from age 1 to age  $a$  and that a mature (adult) survival rate applies from age  $a + 1$  onward. To approximate senescence, it is assumed that all individuals aged  $w$  at time  $t$  die before time  $t + 1$ . Mortality is assumed to be constant over time and in particular to involve no density dependence, so that  $s_{xt}$  does not depend on  $t$ . Thus we have

$$\begin{aligned} s_{xt} &= s_0 & x = 1, \dots, a, \\ &= s & x = a + 1, \dots, w - 1, \\ &= 0 & x = w. \end{aligned} \quad (13)$$

Calf mortality is accounted for in the specification of fertility.

*Fertility.* Fertility is assumed to be constant with respect to age between the age at sexual maturity,  $m$ , and age  $w - 1$ . It is assumed to be density dependent with a functional form corresponding to a modified logistic growth curve. First parturition is assumed to occur one year after the age at sexual maturity. This yields

$$\begin{aligned} f_{xt} &= 0 & x = 1, \dots, m, \text{ and } w, \\ &= f_t = f_0 + (f_{\max} - f_0)[1 - (P_t/P_0)^z] & x = m + 1, \dots, w - 1, \end{aligned} \quad (14)$$

where  $P_t$  is the female population size at the beginning of year  $t$ ;  $f_{\max}$  is the maximum fertility, attained when the stock is near extinction; and  $z$  is the density dependence parameter. Assuming that the population was in equilibrium in 1848 ( $t = 0$ ), before the start of commercial whaling, yields a value for  $f_0$  by solving the matrix equation  $\mathbf{A}_0\mathbf{N}_0 = \mathbf{N}_0$ , namely

$$f_0 = (1 - s)/[(s_0/s)^a(s^m - s^{w-1})]. \quad (15)$$

One estimate of  $f_t$  is shown in Figure 1 with  $f_0 = .0057$  (corresponding to  $s = .98$ ,  $s_0 = .94$ ,  $a = 6$ ,  $m = 20$ , and  $w = 60$ ),  $f_{\max} = .1$ , and  $z = 5$ . Note that  $f_{xt}$  is the product of the number of female calves per mature female with the calf

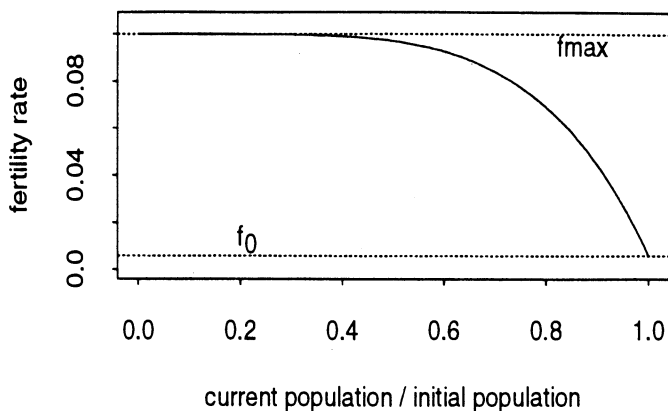


Figure 1. A Typical Estimate of the Fertility Rate  $f_t$  as a Function of Current Population Divided by Initial Population,  $P_t/P_0$ .

survival rate. We have no information that would enable us to separate fertility from calf survival. By including calving rate and calf survival in a single term, we acknowledge that density dependence may occur through changes in reproductive rate, changes in first-year survival, or both.

**Hunting mortality.** The tabulation of hunting kills by year adopted by the IWC is given in table 1 of annex E (IWC 1992). It is shown in Figure 2; note the very heavy mortality in the first five years of commercial whaling.

The lengths of whales killed in the past three decades suggest that the recent subsistence harvest has selected predominantly immature whales. Historically, the commercial catch was probably biased toward larger, mature animals. The model assumes that from 1848 through 1914, 80% of the annual harvest was from sexually mature whales and 20% from immature whales. From 1915 to the present, the model reverses these proportions. It is assumed that half of those caught in each year and age class were females.

With these assumptions, the original model of Breiwick et al. (1984) first divides the catch into two shares: one for immature whales and one for mature whales. It then distributes each share among its corresponding age classes in proportion to the relative abundance of each class at the beginning of the year. With this distribution of the annual catch, it is possible to obtain negative age class counts without population extinction. We have modified the original model so that the number removed from each age class is never more than the current class size, and what remains of the mature and immature catch shares is redistributed proportionally as before but among nonempty age classes.

The model requires values of the eight input parameters,  $s_0$ ,  $s$ ,  $a$ ,  $w$ ,  $m$ ,  $f_{max}$ ,  $z$ , and  $P_0$ , as well as hunting mortality by year. Given these, it outputs a full age distribution of the female population for each year. It is assumed that the sex ratio is 1:1, so doubling this gives the total population.

### 3.2 Yield Quantities and their Relationships to the Model

Several quantities used by the IWC for making policy decisions are related to the inputs and outputs of this PDM (Allen 1976; Butterworth and Best 1990; Cooke 1987). One

is maximum sustainable yield (MSY). Once an unexploited stock of size  $P_0$  begins to be exploited, it can sustain indefinitely any level of catch less than the MSY. The MSY level (MSYL) is the lowest population level at which MSY is attained, expressed as a proportion of  $P_0$ . Under the assumption of density dependence in reproductive rate and/or calf survival, the population increases at a higher rate when it has been reduced below  $P_0$  than when it is at or near its carrying capacity and thus prevented by environmental limitations from increasing. Thus MSYL is less than 1; it has often been assumed to be .6 by the SC.

For a protected species like the bowhead, replacement yield (RY) is a key management concept. RY is the catch from the recruited stock that, if taken, would leave the recruited population at the same level at the beginning of the next season (IWC 1988). For bowheads, we assume that the recruited stock consists of all whales age at least 1 year. The maximum sustainable yield rate (MSYR) is defined as RY at MSYL, expressed as a proportion of the population at MSYL. We define MSYL and MSYR in terms of the total population age 1 or older, as proposed by Butterworth and Punt (1992).

Relationships between the model parameters and MSYL and MSYR are induced by the characteristic equation of the Leslie matrix and by the density-dependence equation in (14). The characteristic equation of the Leslie matrix,  $A_t$ , is

$$\lambda^{m+1} - s\lambda^m - s_0^a s^{m-a} f_t [1 - (s/\lambda)^{w-m-1}] = 0, \quad (16)$$

where  $\lambda$  is the eigenvalue or “population multiplier” so that  $N_{t+1} = \lambda N_t$  by (11) (Breiwick et al. 1984).

If time  $t$  is such that  $P_t/P_0 = \text{MSYL}$  and if  $s_0$ ,  $s$ ,  $a$ ,  $m$ ,  $w$ , and MSYR are known, then we can find the fertility rate at MSYL,  $f_{\text{MSYL}}$ , by setting  $\lambda = 1 + \text{MSYR}$  and solving for  $f_t$  in Equation (16). Similarly,  $f_0$ , the fertility rate of a stable unexploited stock, can be found by setting  $\lambda = 1$  in Equation (16). The only remaining unknown quantities in Equation (14) are  $f_{\text{max}}$  and  $z$ , and, given  $f_{\text{max}}$ , we can solve Equation (14) to find the corresponding value of  $z$ . Note that by definition,  $f_{\text{max}} \geq \max\{f_0, f_{\text{MSYL}}\}$ .

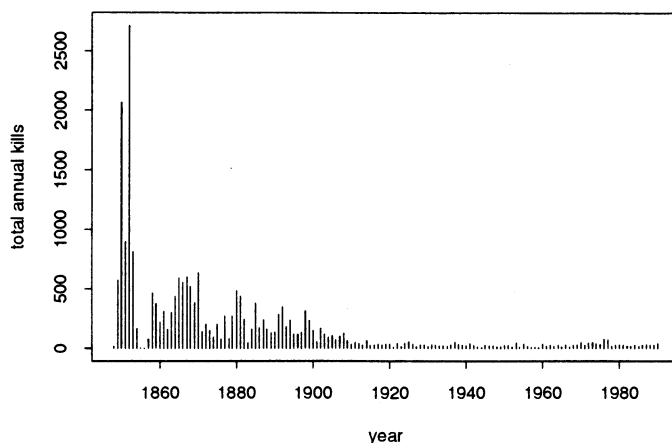


Figure 2. Hunting Mortality, 1848–1990.

## 4. THE PREMODEL DISTRIBUTION

### 4.1 Sources of Information About Model Inputs

The model inputs are all biological parameters about which there is considerable uncertainty. This is because births and natural deaths are rarely observed and ages are very hard to determine, so birth and death rates must be inferred from indirect evidence. There has been considerable research, most of which is summarized in a book titled *The Bowhead Whale* (Burns, Montague, and Cowles 1993).

In May 1991 a special subcommittee of the IWC SC, including a high proportion of all those who have done research on the bowhead whale, met for a week to carry out a comprehensive assessment of the species. This Bowhead Assessment Meeting (BAM) reached agreement on reasonable ranges for many biological parameters. Some extracts from its report (Annex E of IWC 1992) are included here to give an idea of the reasoning leading to our premodel distributions.

Some information about the input parameters comes from known links between them, MSYL, and MSYR (see Sec. 3.2). A set of input parameters is self-consistent if its elements do not contradict these known relationships. Without enforcing self-consistency, it would be possible, for example, to specify priors for survival and fertility rates that would force MSYL values that contradict the desired MSYL prior.

We write the premodel distribution of the inputs as

$$p^{[\theta]}(\theta) \propto p(\theta_0)p(f_{\max}|\theta_0)p(\text{MSYL}|\theta_0, f_{\max}), \quad (17)$$

where  $\theta_0 = (s_0, s, a, m, w, P_0, \text{MSYR})$ ,  $p(f_{\max}|\theta_0) \propto p'(f_{\max}|\theta_0)1_{\{f_{\max} \geq \max(f_0, f_{\text{MSYL}})\}}$ , and  $p(\text{MSYL}|\theta_0, f_{\max}) \propto p'(\text{MSYL}|\theta_0, f_{\max})1_{\{4 \leq \text{MSYL} \leq 9\}}$ . The premodel evidence about the different components of  $\theta_0$  comes from different sources, so we assume these to be premodel independent, except for  $s$  and  $s_0$ .

In (17),  $p(f_{\max}|\theta_0)$  and  $p(\text{MSYL}|\theta_0, f_{\max})$  express the evidence about  $f_{\max}$  and MSYL while also enforcing self-consistency through the dependencies introduced. Equation (17) thus ensures that each joint sample of the input parameters is self-consistent and reasonable in the light of all sources of evidence. The input parameter  $z$  does not appear in (17), because  $z$  is treated as a fixed function of MSYL given  $(\theta_0, f_{\max})$ , and the conditional evidence about  $z$  or MSYL may be expressed in either form (see Sec. 3.2). Because we have evidence about MSYL, we prefer this form.

Figure 3 shows the marginal premodel distributions of the inputs before and after enforcing self-consistency.

### 4.2 Adult and Juvenile Survival Rates, $s$ and $s_0$

The BAM report says:

*Adult mortality:* Eberhardt and Breiwick (1992) used values of .01 and .02. No direct data on mortality rates are available for bowheads, but . . . similar low values had been estimated for the killer whale (Olesiuk et al. 1990). The subcommittee noted that although estimates for other baleen whales are higher than this, those for right whales are similar to those used in Eberhardt and Breiwick (1992). Consequently, values of .01 and .02 were chosen for the HITTER/FITTER runs. Ohsumi noted that an alternative approach to specifying a mortality rate for the runs would be to subtract the estimate of increase accounting for hunting mortality (3.4% from Zeh et al. 1991) from the estimate of gross annual

reproductive rate (5.2% from Withrow and Angliss 1992). This also supports rates near .02.

*Juvenile mortality:* No direct evidence is available. Eberhardt and Breiwick (1992) used a range of .04 to .08 under the assumption that juvenile mortality would be higher than adult mortality; the subcommittee adopted that range for use in this assessment.

The survival rate is equal to 1 minus the mortality rate. We thus adopted distributions with most, but not all, of their probability in the ranges .92–.96 for  $s_0$  and .98–.99 for  $s$ , such that  $s_0 \leq s$ . For the juvenile survival rate we adopted a  $N(.94, .03^2)$  distribution truncated at zero and 1. For the adult survival rate conditional on the juvenile rate, we adopted a  $N(.98, .03^2)$  distribution truncated at the juvenile rate and at 1.

### 4.3 Age at Sexual Maturity, $m$

This has been one of the most difficult and controversial parameters. The BAM report says:

Eberhardt and Breiwick (1992) used values of 13, 15, and 17 years. Aerial photogrammetric measurements indicate that the average length at maturity is approximately 13 meters, and evidence from the landed catch suggests a length at sexual maturity of about 14 meters (Withrow and Angliss 1992). Earlier papers (e.g., Breiwick and Braham 1990) assumed that whales reach this length at ages of 9 years or less. However, Eberhardt and Breiwick (1992) increased this age to 13–17 years, citing recent evidence from carbon isotope aging of baleen plates (Schell et al. 1989) that bowheads grow very slowly, and reach 13 meters at ages between 18 and 20 years. The subcommittee had before it more recent results for baleen aging and photogrammetry (Withrow et al. 1991; Koski et al. 1992), suggesting that bowheads grow very slowly, and that an upper limit exceeding 17 years could also be considered.

Koski et al. (1992) examined 69 pairs of length measurements of identified whales photographed in different years. Their results suggest that bowheads are at least 20 years old at sexual maturity. They had few measurements of whales less than 9.5 meters long, but their analyses suggest that bowheads require roughly 19 to 24 years to grow from 9.5 to 13 m, the approximate length at sexual maturity.

If young bowheads grow rapidly, like most mammals, they may be only 2 or 3 years old at 9.5 m. But both the photogrammetric and the carbon isotope data suggest that bowheads may grow slowly for several years after weaning. The data of Schell et al. (1989) suggest that some 9.5-meter bowheads may be around 8 years old. Taken together with the photogrammetric results, this suggests that age at first parturition for those whales could exceed 30.

But discussions at the BAM suggested that such old ages at first parturition would be unusual and that ages as low as 10 years were more consistent with data on other baleen whales. To accommodate all this evidence, we used a discrete uniform premodel distribution on the integers 10, 11, . . . , 30 for  $m$ .

### 4.4 MSYR

The beliefs of the IWC SC concerning MSYR are summarized in the values used in simulation trials of the Revised Management Procedure (IWC 1992). Most of the trials used .01, .025, .04, and .07 relative to the *mature* component of the population. The irregular spacing of these values suggest



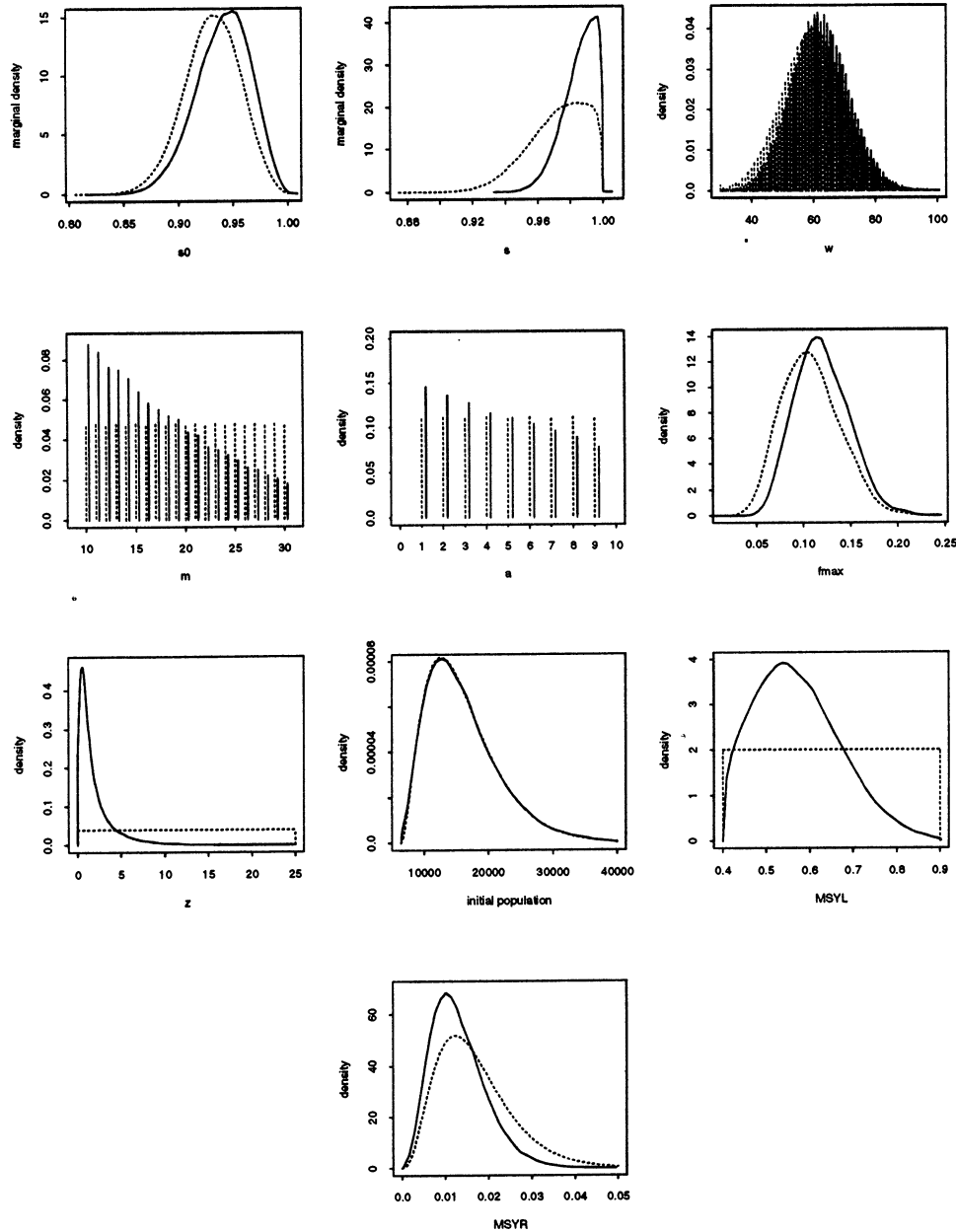


Figure 3. Marginal Premodel Distributions of the Model Inputs: Annual Juvenile Survival Rate,  $s_0$ ; Annual Adult Survival Rate,  $s$ ; Age at Senescence,  $w$ ; Age at Sexual Maturity,  $m$ ; Age at End of Juvenile Mortality,  $a$ ; Maximum Fertility,  $f_{\max}$ ; Density-Dependence Parameter,  $z$ ; Initial Population Size,  $P_0$ ; MSYL; and MSYR. The dashed lines represent the premodel distributions before enforcing self-consistency; the solid lines these distributions after enforcing self-consistency.

that values in the range .01–.04 were considered more likely than higher values. A “random parameters” trial used MSYR values down to .001, and some members of the SC have argued that MSYR may exceed .07.

To give most weight to the 1%–4% range but allow values very near zero or above 7%, we chose a gamma premodel distribution for MSYR relative to the mature population. It was converted to MSYR relative to the  $1^+$  population, as described by Givens et al. (1993).

#### 4.5 Maximum Fertility, $f_{\max}$

We express the evidence about  $f_{\max}$  via  $p(f_{\max}|\theta_0) \propto p'(f_{\max}|\theta_0)1_{\{f_{\max} \geq \max(\hat{f}_0, f_{\text{MSYL}})\}}$ . We consider  $p'(f_{\max}|\theta_0)$

first. We derive this prior for  $f_{\max}$  from data on the minimum possible average calving interval,  $i_{\min}$ , and the annual calf survival rate,  $s_0^*$ , using the relationship

$$f_{\max} = s_0^*/(2i_{\min}). \tag{18}$$

The BAM discussion synthesized the results of three studies (George et al. 1992; Miller et al. 1992; Rugh, Miller, Withrow and Koski 1992) that used photographic reidentification methods and observed the number of pregnant females taken in the Eskimo hunt (Raftery and Davis 1992). The conclusion was that bowheads may calve at intervals of four to seven years, and that under optimal conditions the calving interval may be as short as three years. The data pointed to

four years as the most plausible integer value of the average calving interval. Crude birth rate estimates from photogrammetry data (Withrow and Angliss 1992) also suggest a four-year calving interval. We adopted a gamma (3.25, 1.5) distribution for  $i_{\min} - 2$ , so that the premodel mean of  $i_{\min}$  is  $2 + 3.25/1.5 = 4.17$ .

The calf mortality rate of killer whales is estimated to be in the range 37–50% (Olesiuk, Biggs, and Ellis, 1990). We know of no direct evidence for bowheads. Under the assumption that even under optimal conditions, calves must have mortality at least as high as 1 year olds, we specified the distribution of  $(s_0^* | s_0)$  to be uniform between .75 and  $s_0$ . The resulting density for  $f_{\max}$  yields the marginal distribution shown in Figure 3 by the dashed line.

Because  $f_{\max} \geq f^*(\theta_0) = \max\{f_0, f_{\text{MSYL}}\}$ , we drew  $f_{\max}$  from  $p'(f_{\max} | \theta_0)$  truncated below at  $f^*$  and then rejected the multivariate draw with probability  $c_{f^*}$ , where

$$c_{f^*}(\theta_0) = \int_0^{f^*} p'(f_{\max} | \theta_0) df_{\max}. \quad (19)$$

This rejection step corrects the sampling probability so that  $f_{\max}$  is drawn with probability  $p(f_{\max} | \theta_0)$ , which enforces the link between  $f_{\max}$  and  $\theta_0$  required for self-consistency while summarizing the available evidence. Figure 3 shows  $p(f_{\max} | \theta_0)$  with the solid line.

#### 4.6 MSYL and the Density Dependence Parameter, $z$

Given MSYL,  $f_{\max}$ , and  $\theta_0$ , we can solve for  $z$  as noted in Section 3.2. Thus we focus here on sampling MSYL. The BAM adopted .6 and .7 for MSYL. We prefer to use broader constraints on MSYL, because very little is known about this parameter.

Fowler (1987) has estimated a linear relationship between MSYL and the log maximum rate of population increase per generation, using data on many animals including whales, elephants, mice, and bacteria. The generational rate of increase can be calculated from the stable age distribution (Breiwick et al. 1984), the generation length, and the maximum net per capita rate of increase. These quantities in turn can be calculated from  $s_0$ ,  $s$ ,  $a$ ,  $m$ ,  $w$ , and  $f_{\max}$ . Because Fowler's relationship is a regression equation with a random error term, we derived a conditional density for MSYL from the regression model, convolved it with the estimated Gaussian error distribution to obtain  $p'(\text{MSYL} | \theta_0, f_{\max})$ , and restricted the resulting distribution to  $.4 \leq \text{MSYL} \leq .9$ . The conditional density  $p(\text{MSYL} | \theta_0, f_{\max}) \propto p'(\text{MSYL} | \theta_0, f_{\max}) 1_{\{.4 \leq \text{MSYL} \leq .9\}}$  summarizes the link between MSYL and the other variables as well as expert opinion regarding the reasonable limits of MSYL. This marginal distribution for MSYL (and, therefore, for  $z$ ) is shown in Figure 3 by the solid line.

#### 4.7 Other Inputs and Outputs

The rationales for the premodel distributions of the other inputs,  $a$ ,  $w$ , and  $P_0$ , were given by Raftery et al. (1992) and Givens et al. (1993). For the age at the end of juvenile mortality,  $a$ , we used a discrete uniform distribution on the

integers  $\{1, \dots, 9\}$ . For the age at senescence,  $w$ , we adopted a discretized  $N(60, 10^2)$  distribution. For the unexploited population size,  $P_0$ , we used the premodel distribution derived by Givens et al. (1993) from catch and effort data in the early years of the commercial hunt.

The model outputs are numbers of whales of each age in each year. There is considerable premodel information from the 1978–1988 censuses about recent values of quantities that can be calculated from the model outputs. The premodel distribution for the 1988 population size,  $P_{1988}$ , is the Bayes empirical Bayes posterior distribution from Raftery and Zeh (1993). The premodel distribution for the 1978–1988 annual rate of population increase, ROI, is that of  $e^{.03053 + .01240t_7} - 1$ , where  $t_7$  is a random variable that has a  $t$  distribution with 7 degrees of freedom. This is based on the results of Zeh et al. (1991).

Because percent calves, percent immature, and percent mature sum to 1, we have specified the premodel distribution of the 1988 population age distribution in terms of percent calves and percent mature only. The length data from Withrow and Angliss (1992) provide information about the age distribution. They presented counts of calves, immatures, and matures for 1985, 1986, 1989, and 1990. We assume that the age distribution did not change appreciably between 1985 and 1990. Treating the counts for these four years as independent, we estimated the means, variances, and covariance of percent calves and percent mature. Based on this, we adopted a bivariate normal distribution for percent calves and percent mature with means 5.8 and 43.6, standard deviations 3.6 and 3.9, and correlation .23.

### 5. RESULTS: THE POSTMODEL DISTRIBUTION

Of the  $n = 200,000$  draws from  $p^{(\phi)}(\theta)$ , 69,247 resulted in model trajectories that had nonzero probability according to  $p^{(\phi)}(\phi)$ . The final sample of  $m = 5,000$  draws consisted of 1,459 unique points, and no single point was chosen more than 26 times. These results provided smooth estimates of the postmodel density of quantities of interest. The procedure did not have much Monte Carlo variability; in a replicate run, 69,748 simulations had  $p^{(\phi)}(\phi) > 0$ , and there were 1,478 unique points in the final sample.

Figures 4, 5, and 6 show marginal histograms of the postmodel sample with estimates of the marginal postmodel densities superimposed. The latter were calculated by nonparametric kernel density estimation with a Gaussian kernel using the maximal smoothing span of Terrell (1990). Initial marginal premodel distributions are shown by dotted lines. The marginal postmodel distribution of the quantity of primary interest, RY for 1990, is shown in Figure 4. The 5% point of the postmodel sample is 57 (and 58 in the replicate run). The postmodel mode, or most probable value, is 104.

For many inputs and outputs, such as  $P_{1988}$  in Figure 4, the premodel and postmodel distributions were similar, indicating that the model provided little information about these quantities. For others, the premodel and postmodel distributions were quite different. Perhaps the most important of these is the 1978–1988 rate of increase (Figure 4). Although these results strongly support the conclusion of Zeh et al. (1991) that the population has been increasing, it

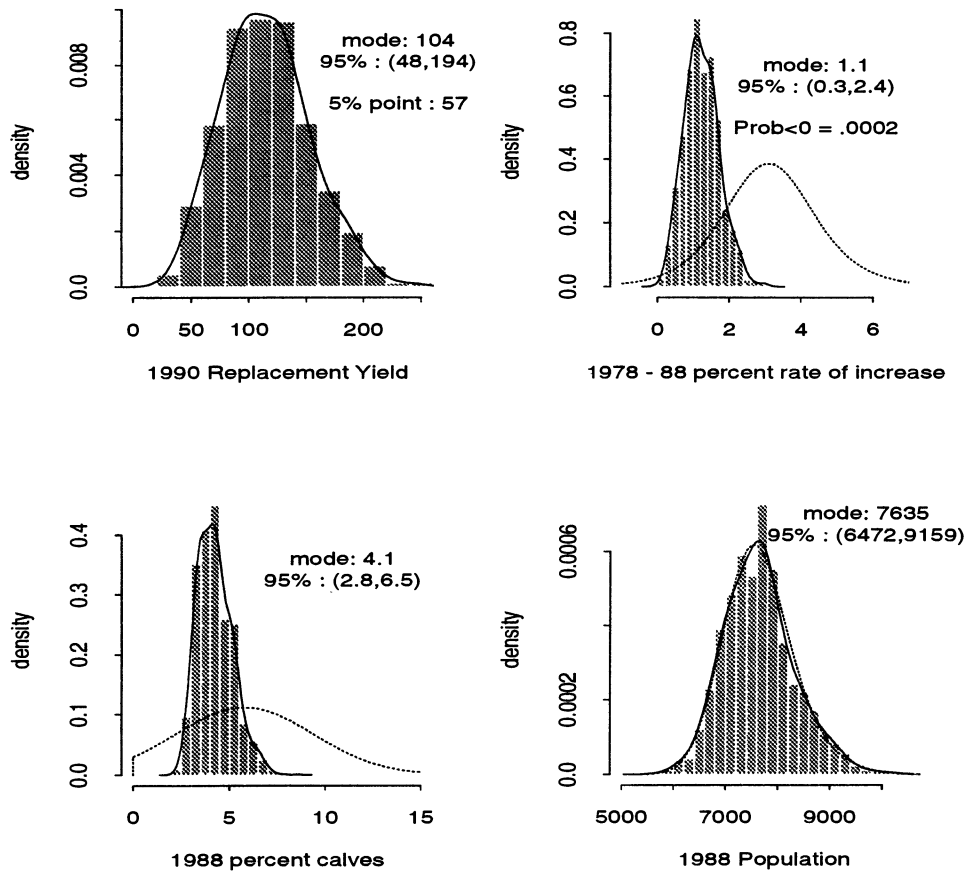


Figure 4. Marginal Postmodel Distributions of Some Model Outputs: RY, ROI, Percent Calves, and  $P_{1988}$ . Postmodel density estimates (solid lines) and initial marginal premodel distributions (dotted lines) are superimposed.

seems that the rate of increase is at the low end of the interval based on the visual census data, with a postmodel mode of about 1%, compared with the premodel mode of about 3%. But the two distributions are not in conflict, and the postmodel 95% interval is completely contained in the premodel 95% interval. The 1988 percentage of calves is also most likely to be in the lower part of its premodel distribution.

The results shed some light on the controversy about age at sexual maturity (Fig. 5). They point to higher values: the most likely value is 21, and nearly all the values favored by the combined evidence (i.e., for which the postmodel probability is higher than the premodel probability) are between 16 and 25. Figure 5 also indicates that the lowest values in the premodel ranges of  $f_{max}$  and  $P_0$  are unlikely, whereas the highest values of MSYR are unlikely. But the postmodel distributions of  $f_{max}$  and MSYR, like those of the survival parameters and MSYL, are not very different from the premodel distributions after self-consistency has been enforced.

The initial population size,  $P_0$ , was probably in the range 15,000–23,000. Thus the stock is still depleted; the postmodel mode of depletion level, defined as  $P_{1988}/P_0$ , is 42% (Fig. 6). The stock was probably once depleted to about 15% of its equilibrium level, or about 3,000 whales.

Figure 7 (p. 413) shows the ratio of the postmodel variance to the premodel variance for several model inputs and outputs. This shows which quantities the model tells us most about. A value close to 1 says that the model is not providing

much additional information, whereas a value much smaller than 1 indicates that the model is substantially reducing uncertainty. Figure 7 shows that the model is most informative about  $P_0$ , % calves and ROI.

Figures 4 and 5 provide an informal check on the fit of the model; that is, on its consistency with the premodel information. All the postmodel histograms have modes well within the range of the premodel distributions. This indicates that the selected trajectories are consistent with all the premodel information, and so the model is able to accommodate all the available evidence.

Figures 4 and 5 also allow us to perform some informal sensitivity analysis to the premodel specification. Changing the premodel distribution so as to assign less weight to values that have virtually no weight under the postmodel distribution would change the results very little. Thus, for example, changing the premodel distribution of  $P_0$  so as to exclude values below 15,000 and above 23,000 would have essentially no effect on the results.

Figure 8 (p. 413) shows the median trajectory for the total population size with a pointwise confidence band. This is qualitatively consistent with existing beliefs and results and with the premodel evidence.

Figure 9 (p. 413) shows the results of a cluster analysis on the quantities of interest using the connected hierarchical clustering method (Everitt 1980), where the distance measure is chosen to be 1 minus the absolute postmodel corre-

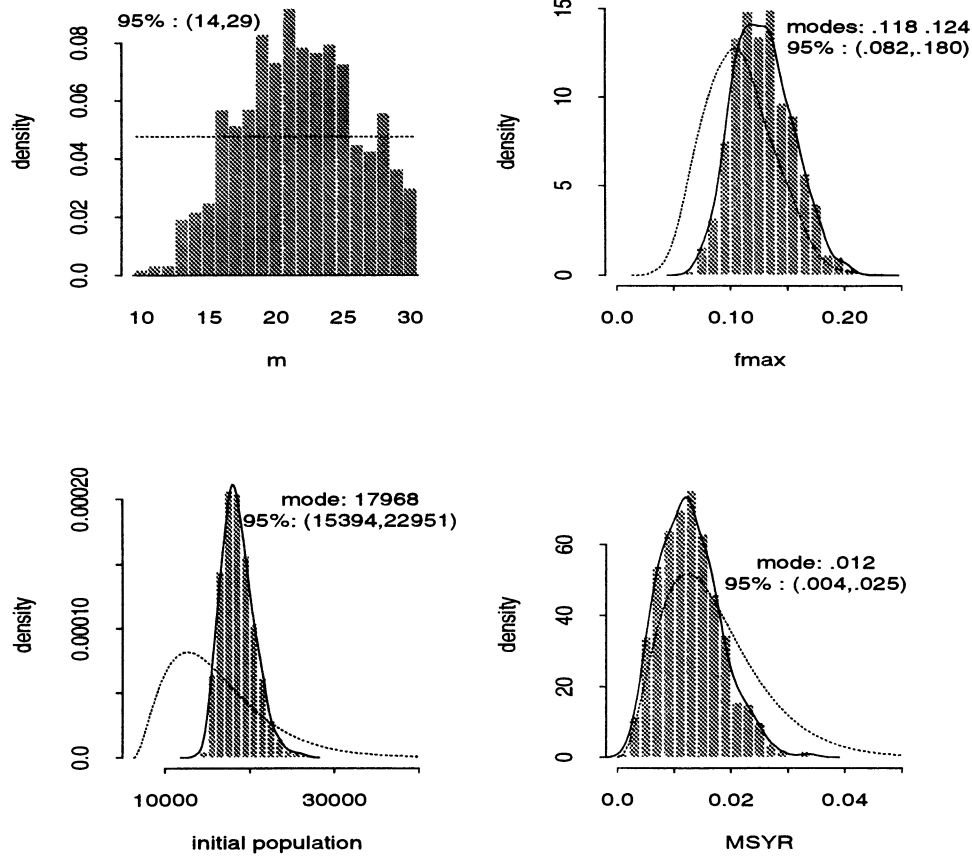


Figure 5. Marginal Postmodel Distributions of Some Model Inputs:  $m$ ,  $f_{max}$ ,  $P_0$ , and MSYR.

lation. It shows that the quantity of primary interest, RY, is far more highly correlated with  $P_0$ , MSYR, and ROI than with the other model inputs and outputs. Jittered bivariate scatterplots for RY, ROI, MSYR, and  $P_0$  are shown in Figure 10 (p. 414); some of the relationships are nonlinear. These results suggest that modest changes in the pre-model distributions of other inputs and outputs would have relatively little effect on inference about RY. Figure 9 also shows that there are some important dependencies between other model inputs and outputs.

The dependencies evident in Figures 9 and 10 suggest that the model could be much simplified to one involving only four variables (instead of the present 15), perhaps linked by equations empirically estimated from the present output,

without losing much accuracy in terms of inference about RY. This is also useful for designing future research: research that refines our knowledge about  $P_0$ , MSYR, or ROI is likely to be most useful.

## 6. DISCUSSION

### 6.1 General Comments and Related Work

We have developed a framework for combining different sources of information about bowhead whales using a deterministic PDM. The method consists of translating all the available information into a joint premodel distribution on both inputs and outputs, and then restricting this to the submanifold defined by the model to obtain the joint postmodel

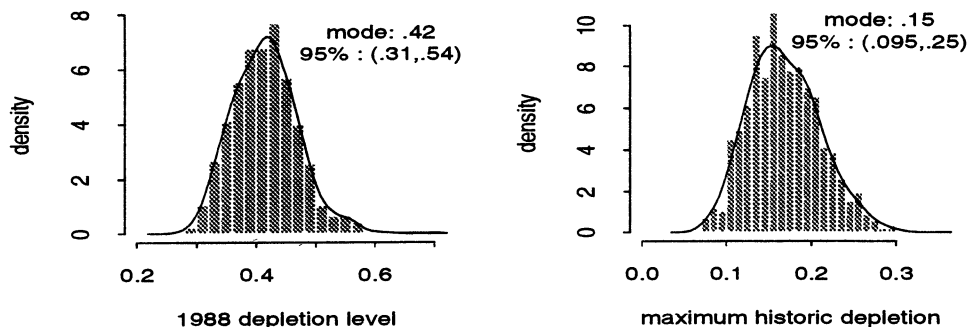


Figure 6. Marginal Postmodel Distributions of 1988 Depletion Level,  $P_{1988}/P_0$ , and Maximum Historic Depletion.

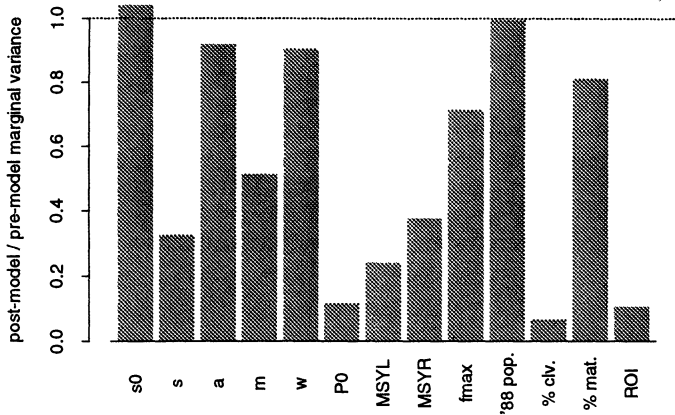


Figure 7. Ratio of Postmodel Variance to Premodel Variance for Selected Inputs and Outputs.

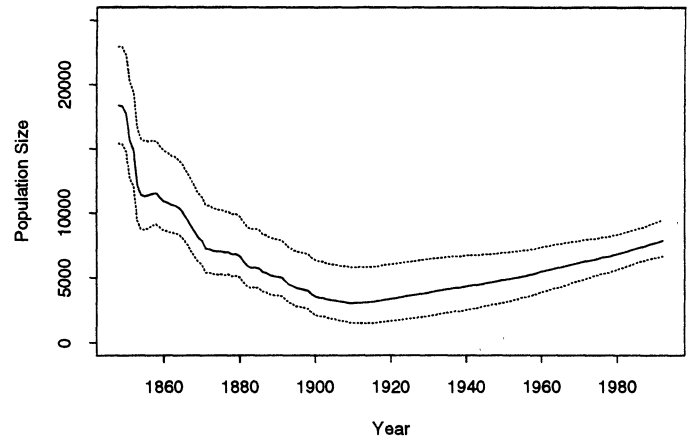


Figure 8. The Pointwise Median Trajectory With Pointwise 95% Intervals.

distribution. Samples from the latter can be obtained by importance sampling and Rubin's SIR algorithm.

The method yields inference, conditional on the model, about model inputs and outputs and functions of these, which takes full account of uncertainty about these quantities and takes full advantage of available information. The results are in the form of a sample from the postmodel distribution, and so this can be explored using the full range of modern multivariate exploratory data analysis techniques. This leads to simple informal model checking methods, some examples of which are given in Section 5. It also suggests how the model might be simplified. It is easy to recompute the results quickly for different premodel distributions, without any additional model runs.

The methods yield full inference about the quantity of primary interest for IWC decision-making about bowheads, namely RY, and they also shed some light on several other questions of scientific interest.

Patwardhan and Small (1992), Wolpert, Reckhow, and Steinberg (1992) and Small and Escobar (1992) have, like us, considered Monte Carlo methods for taking into account parameter uncertainty in simulation models. They considered only the situation where information about model outputs consists of measurements of the outputs with random error. Then, as pointed out in Section 2, the problem can be reduced to a standard Bayesian one with a prior, a likelihood, and a posterior, and standard methods apply. Our framework includes this situation as a special case.

Ragen (1991), building on work of Smith and Polacheck (1984), has used an alternative approach to accounting for uncertainty in PDM's. This approach consists of running the model for all combinations of the model inputs on a grid and eliminating those runs whose outputs do not satisfy certain criteria. This yields a valid solution to the problem, provided that premodel distributions of all the inputs and outputs are independent uniform distributions. It is very expensive computationally, and if applied to our problem would probably require more than one billion runs of the model, which would be prohibitive. Our approach allows much more general premodel distributions and also requires much less computer time.

### 6.2 Improvements

There are several ways in which this work could be improved. We are using a bad importance sampling function (i.e., the premodel distribution of model inputs), because it is easy to sample from. This means that we have to make a lot of model runs or end up with a bad estimate of the post-model distribution that is heaped on a few points. Because our PDM runs fast that is not a problem for us here, but for more complex models it would lead to difficulties. We have been experimenting with adaptive importance sampling (Smith, Skene, Shaw, and Naylor 1987; West 1991) to refine the initial importance sampling function. We have been exploring modifications of this that use local kernel density estimates (Givens 1993; Givens and Rafferty 1993).

The PDM itself could be made more realistic. For example, mortality and age at sexual maturity could be made density dependent, and uncertainty about the historic kill could be included explicitly. The model comparison methods in Section 2.4 could be used to assess whether such elaborations are useful.

Clearly, a stochastic PDM would be more realistic than the deterministic one that we are currently using. A deter-

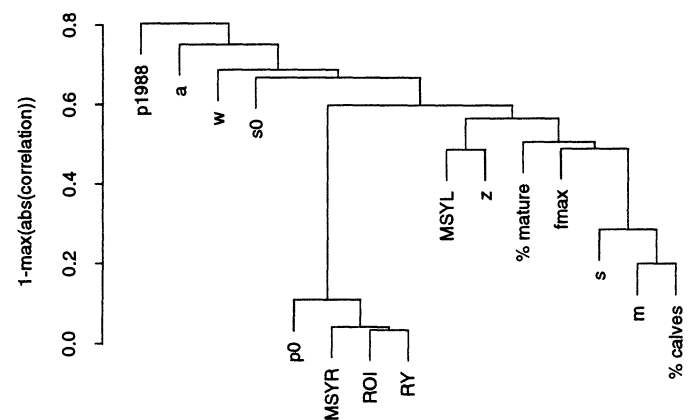


Figure 9. Cluster Analysis on Postmodel Sample of Inputs and Outputs, Using the Connected Clustering Method With Distance Equal to 1 Minus the Absolute Postmodel Correlation.

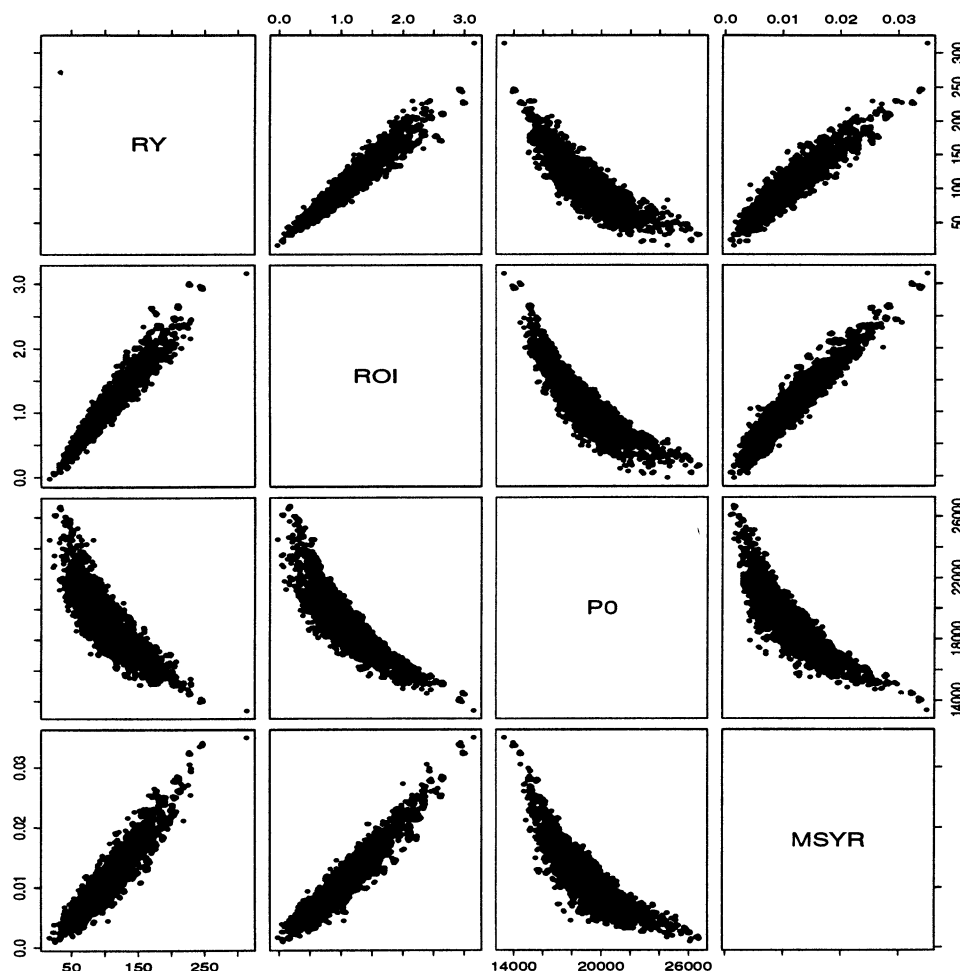


Figure 10. Bivariate Jittered Scatterplots of the Postmodel Distribution of  $RY$ ,  $ROI$ ,  $P_0$ , and  $MSYR$ .

ministic model is much simpler, and its use is based on the assumption that only a small part of the overall uncertainty is due to random variation in the numbers of births and deaths. The IWC has always taken this view. If all whales have the same age-specific fertility and death rates, then the numbers of births and deaths for each age in each year are independent binomial random variables. If this is so, then random variation probably does account for little of the overall uncertainty, in which case deterministic and stochastic models would give similar results. In this sense, ignoring random variation can be viewed as just another approximation or modeling assumption.

Our approach could take into account such random variation quite easily by integrating over trajectories as well as inputs in Equation (3). This can be done by simulating one or several trajectories for each value of  $\theta$  drawn. If birth and death rates vary between whales, then the modeling is harder and the random variation more important. It would be useful at least to assess the sensitivity of our results to such heterogeneity.

### 6.3 Other Approaches

As an alternative to the SIR algorithm described in Section 2.2, a rejection sampling method might be considered for generating the postmodel sample; this was suggested by Smith

and Gelfand (1992) in the context of standard Bayesian inference (but not of simulation models). But for the present problem, the maximum likelihood problem is intractable, so that their suggestion of using the prior scaled up by the likelihood evaluated at the maximum likelihood estimator (MLE) as the initial sampling envelope would not be feasible.

Simulation models often require a lot of computer time, and this limits the number of runs that can be made. To get the most out of a limited number of runs, a nonstochastic way of doing the required numerical integration, such as quadrature, may be best (Naylor and Smith 1982). A difficulty with this is that the number of model runs required by most quadrature methods increases geometrically with the number of input parameters; for our problem, a practical minimum would be  $3^{10} = 59,049$  runs. Raftery and Zeh (1993, app. 4) described a more economical quadrature method (partial iterated three-point Gauss-Hermite quadrature) that uses the output from a standard sensitivity analysis and needs a number of runs that is only a polynomial function of the number of input parameters; here it could be implemented with only 201 runs. A generalization of the method of O'Hagan (1991) might also be useful.

We have described a way of simulating from a high-dimensional target distribution. Another way of doing this is by a Markov chain Monte Carlo method, such as the Gibbs

sampler (Gelfand and Smith 1990) or, more generally, the Metropolis–Hastings algorithm (e.g., Smith and Roberts 1993). But the Gibbs sampler may well work poorly here because of the strong dependencies induced by the model (Fig. 10). Parameter transformations might help, but they would require prior knowledge about the dependencies, which is not generally available. More sophisticated Markov chain Monte Carlo methods, such as those described by Besag and Green (1993), might also be useful but could well be unnecessarily complicated. For the present problem, the approach described here works well and is simpler to apply.

#### 6.4 Software

Software to implement the methods described here is available free of charge by sending the e-mail message “send baysyn from general” to statlib@stat.cmu.edu. This is a large Fortran program specially designed for the bowhead application, but it could be adapted to other applications.

Software to carry out the reweighting procedure for changing the prior without rerunning the whole procedure is also available by sending the message “send bergs from general” to statlib@stat.cmu.edu. A detailed description of this methodology and the program was given by Givens et al. (1994); see also Section 2.3. This is also a Fortran program.

These programs are made available purely as templates for the convenience of other researchers. Validated versions of these programs for official use in the assessment of whale stocks are maintained by the Secretariat, International Whaling Commission, The Red House, Station Road, Histon, Cambridge CB4 4NP, U.K., to whom enquiries may be directed.

[Received June 1993. Revised June 1994.]

#### REFERENCES

- Allen, K. R. (1976), “A More Flexible Model for Baleen Whale Populations,” *Report of the International Whaling Commission*, 26, 247–263.
- Allen, K. R. (1980), *The Conservation and Management of Whale Stocks*, Seattle: University of Washington Press.
- Banks, S. (1991), “Exploratory Modeling and the Use of Simulation for Policy Analysis,” Report N-3093-A, RAND Corporation.
- Besag, J. E., and Green, P. J. (1993), “Spatial Statistics and Bayesian Computation,” *Journal of the Royal Statistical Society*, Ser. B, 55, 25–37.
- Best, P. B. (1991), “Comparative Life History of the Balaenids, Excluding *B. mysticetus*,” Paper SC/43/PS13 presented to the International Whaling Commission, May 1991.
- Breiwick, J. M., and Braham, H. W. (1990), “Historical Population Estimates of Bowhead Whales: Sensitivity to Current Population Size,” *Report of the International Whaling Commission*, 40, 423–426.
- Breiwick, J. M., Eberhardt, L. L., and Braham, H. W. (1984), “Population Dynamics of Western Arctic Bowhead Whales (*Balaena mysticetus*),” *Canadian Journal of Fisheries and Aquatic Sciences*, 41, 484–496.
- Burns, J. J., Montague, J. J., and Cowles, C. J. (eds.) (1993), *The Bowhead Whale*, Special Publication Number 2, The Society for Marine Mammalogy.
- Butterworth, D. S., and Best, P. B. (1990), “Implications of the Recovery Rate of the South African Right Whale Population for Baleen Whale Population Dynamics,” *Report of the International Whaling Commission*, 40, 433–447.
- Butterworth, D. S., and Punt, A. E. (1992), “The Scientific Committee . . . Agreed That the MSY Rate Would Most Likely Lie Between 1 and 4%—But Which MSY Rate?” *Report of the International Whaling Commission*, 42, 583–591.
- Caswell, H. (1976), “The Validation Problem,” in *Systems Analysis and Simulation in Ecology*, ed. B. C. Patten, New York: Academic Press, pp. 313–325.
- Cooke, J. G. (1987), “Catch History and Current Status of the Northwest Pacific Bryde’s Whale (*Balaenoptera edeni*),” Paper SC/39/Ba4 presented to the International Whaling Commission Scientific Committee, June 1987.
- de la Mare, W. K. (1985), “On the Estimation of Mortality Rates From Whale Age Data, With Particular Reference to Minke Whales (*Balaenoptera acutorostrata*) in the Southern Hemisphere,” *Report of the International Whaling Commission*, 35, 239–250.
- (1989), “The Model Used in the Hitter and Fitter Programs (Program: Fitter.SC40),” Report of the Scientific Committee, Annex L, *Report of the International Whaling Commission*, 39, 150–151.
- Eberhardt, L. L. (1985), “Assessing the Dynamics of Wild Populations,” *Journal of Wildlife Management*, 49, 997–1102.
- Eberhardt, L. L., and Breiwick, J. M. (1992), “Impact of Recent Population Data on Historical Population Levels of Bowhead Whales Inferred From Simulation Studies,” *Report of the International Whaling Commission*, 42, 485–489.
- Everitt, B. (1980), *Cluster Analysis* (2nd ed.), New York: Halsted Press.
- Fowler, C. W. (1981), “Density Dependence as Related to Life History Strategies,” *Ecology*, 62, 602–610.
- (1987), “A Review of Density Dependence in Populations of Large Mammals,” in *Current Mammalogy*, Vol. 1, ed. H. H. Genoways, New York: Plenum, pp. 401–441.
- Gelfand, A. E., and Smith, A. F. M. (1990), “Sampling-Based Approaches to Calculating Marginal Densities,” *Journal of the American Statistical Association*, 85, 398–409.
- George, J. C., Philo, L. M., Suydam, R., Tarpley, R., and Albert, T. F. (1992), “Summary of the 1989 and 1990 Subsistence Harvests of Bowhead Whales (*Balaena mysticetus*) by Alaskan Eskimos,” Paper SC/43/PS18 presented to the International Whaling Commission Scientific Committee, May 1991.
- Givens, G. H. (1993), “Contribution to Discussion on the Meeting on ‘The Gibbs Sampler and Other Markov Chain Monte Carlo Methods,’” *Journal of the Royal Statistical Society*, Ser. B, 55, 75–76.
- Givens, G. H., and Raftery, A. E. (1993), “Local Adaptive Importance Sampling for Multivariate Densities With Strong Nonlinear Relationships,” Technical Report 260, University of Washington, Dept. of Statistics.
- Givens, G. H., Raftery, A. E., and Zeh, J. E. (1993), “Benefits of a Bayesian Approach for Synthesizing Multiple Sources of Evidence and Uncertainty Linked by a Deterministic Model,” *Report of the International Whaling Commission*, 43, 495–500.
- (1994), “A Reweighting Approach for Sensitivity Analysis Within the Bayesian Synthesis Framework for Population Assessment Modeling,” *Report of the International Whaling Commission*, 44, 377–384.
- Givens, G. H., Zeh, J. E., and Raftery, A. E. (1993), “Assessment of the Bering–Chukchi–Beaufort Seas stock of bowhead whales using the Hitter–Fitter model in a Bayesian synthesis framework,” Paper SC/45/AS1 presented to the International Whaling Commission Scientific Committee, April, 1993.
- Guttorp, P., and Walden, A. T. (1987), “On the Evaluation of Geophysical Models,” *Geophysical Journal of the Royal Astronomical Society*, 91, 201–210.
- Hodges, J. S. (1987), “Uncertainty, Policy Analysis and Statistics” (with discussion), *Statistical Science*, 2, 259–291.
- (1991), “Six (or so) Things You can do With a Bad Model,” *Operations Research*, 39, 355–365.
- Hughes, W. P. (1984) (ed.), *Military Modeling*, Alexandria, VA: Military Operations Research Society.
- International Whaling Commission (1988), “Bering–Chukchi–Beaufort Sea Bowhead Population Trend Simulations,” Appendix 4 of Annex H (Protected Species), Report of the Scientific Committee, *Report of the International Whaling Commission*, 38, 116.
- (1992), “Report of the Scientific Committee,” *Report of the International Whaling Commission*, 42, 51–270.
- Jeffreys, H. (1961), *Theory of Probability* (3rd ed.), Cambridge, U.K.: Cambridge University Press.
- Kass, R. E., and Raftery, A. E. (1995), “Bayes Factors,” *Journal of the American Statistical Association*, 90, 773–795.
- Koski, W. R., Davis, R. A., Miller, G. W., and Withrow, D. E. (1992), “Growth Rates of Bowhead Whales as Determined From Low-Level Photogrammetry,” *Report of the International Whaling Commission*, 42, 491–499.
- Leslie, P. H. (1945), “On the Use of Matrices in Certain Population Mathematics,” *Biometrika*, 33, 183–212.
- (1948), “Some Further Notes on the Use of Matrices in Population Mathematics,” *Biometrika*, 35, 213–245.
- Lewis, E. G. (1942), “On the Generation and Growth of a Population,” *Sankhya*, 6, 93–96.

- Madigan, D., and Raftery, A. E. (1994), "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window," *Journal of the American Statistical Association*, 89, 1535-1546.
- Miller, G. W., Davis, R. A., Koski, W. R., Crone, M. J., Rugh, D. J., Withrow, D. E., and Fraker, M. A. (1992), "Calving Intervals of Bowhead Whales—An Analysis of Photographic Data," *Report of the International Whaling Commission*, 42, 501-506.
- Naylor, J. C., and Smith, A. F. M. (1982), "Applications of a Method for the Efficient Computation of Posterior Distributions," *Applied Statistics*, 31, 214-225.
- Newton, M. A., and Raftery, A. E. (1991), "Approximate Bayesian Inference With the Weighted Likelihood Bootstrap," Technical Report 199. University of Washington, Dept. of Statistics.
- (1994), "Approximate Bayesian Inference With the Weighted Likelihood Bootstrap" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 57, 3-48.
- O'Hagan (1991), "Bayes-Hermite Quadrature," *Journal of Statistical Planning and Inference*, 29, 245-260.
- Olesiuk, P., Bigg, M. A., and Ellis, G. M. (1990), "Life History and Population Dynamics of Resident Killer Whales (*Orcinus orca*) in the Coastal Waters of British Columbia and Washington State," in *Individual Recognition of Cetaceans*, Special Issue 12, eds. P. S. Hammond, S. A. Mizroch, and G. P. Donovan, Cambridge, U.K.: International Whaling Commission, pp. 209-243.
- Patwardhan, A., and Small, M. J. (1992), "Bayesian Methods for Model Uncertainty Analysis With Applications to Future Sea Level Rise," *Risk Analysis*, 12, 513-523.
- Raftery, A. E., and Davis, R. A. (1992), "Bowhead Whale Calving Intervals," Appendix 2 to Annex E of the Report of the Scientific Committee, *Report of the International Whaling Commission*, 42, 153.
- Raftery, A. E., Givens, G. H., and Zeh, J. E. (1992), "Inference From a Deterministic Population Dynamics Model About Bowhead Whale, *Balaena mysticetus*, Replacement Yield," Technical Report 232, University of Washington, Dept. of Statistics.
- Raftery, A. E., Madigan, D., and Hoeting, J. A. (1993), "Model Selection and Accounting for Model Uncertainty in Linear Regression Models," Technical Report 262, University of Washington, Dept. of Statistics.
- Raftery, A. E., and Schweder, T. (1993), "Inference About the Ratio of Two Parameters, With Application to Whale Censusing," *The American Statistician*, 47, 259-264.
- Raftery, A. E., and Zeh, J. E. (1993), "Estimation of Bowhead Whale, *Balaena mysticetus*, Population Size" (with discussion), in *Bayesian Statistics in Science and Technology: Case Studies*, eds. J. S. Hodges, R. E. Kass, C. Gatsonis, and N. Singpurwalla, New York: Springer-Verlag, pp. 163-240.
- Ragen, T. J. (1991), "Maximum Net Productivity Level Estimation for the Northern Fur Seal (*Callorhinus ursinus*) Population of St. Paul Island, Alaska," research report, National Marine Mammal Laboratory, Seattle, WA.
- Rubin, D. B. (1987), Comment on "The Calculation of Posterior Distributions by Data Augmentation", by M. Tanner and W. H. Wang, *Journal of the American Statistical Association*, 82, 543-546.
- (1988), "Using the SIR Algorithm to Simulate Posterior Distributions," in *Bayesian Statistics 3*, eds. J. M. Bernardo, et al., Oxford, U.K.: Clarendon Press, pp. 395-402.
- Rugh, D. J., Miller, G. W., Withrow, D. E., and Koski, W. R. (1992), "Calving Intervals of Bowhead Whales Established Through Photographic Reidentifications," *Journal of Mammalogy*, 73, 487-490.
- Schell, D. M., Saupe, S. M., and Haubenstock, N. (1989), "Bowhead Whale (*Balaena mysticetus*) Growth and Feeding as Estimated by  $\delta^{13}\text{C}$  Techniques," *Marine Biology*, 103, 433-443.
- Schnute, J. T., and Hilborn, R. (1993), "Analysis of Contradictory Data Sources in Fish Stock Assessment," *Canadian Journal of Fisheries and Aquatic Science*, 50, 1916-1923.
- Small, M. J., and Escobar, M. D. (1992), "Bayesian Monte Carlo Methods for Mechanistic Models," in *Bayesian Statistics in Science and Technology: Case Studies*, eds. J. S. Hodges, R. Kass, C. Gatsonis, and N. Singpurwalla, New York: Springer-Verlag, pp. 271-286.
- Smith, A. F. M., and Gelfand, A. E. (1992), "Bayesian Statistics Without Tears: A Sampling-Resampling Perspective," *The American Statistician*, 46, 84-88.
- Smith, A. F. M., and Roberts, G. O. (1993), "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society*, Ser. B, 55, 3-23.
- Smith, A. F. M., Skene, A. M., Shaw, J. E. H., and Naylor, J. C. (1987), "Progress With Numerical and Graphical Methods for Practical Bayesian Statistics," *The Statistician*, 36, 75-82.
- Smith, T. D., and Polacheck, T. (1984), "The Population Dynamics of the Alaska Fur Seal: What Do We Really Know?" Northwest and Alaska Fisheries Center Processed Report 84-15, available from the National Marine Mammal Laboratory, Seattle, WA.
- Speed, T. (1983), "Modelling," *CSIRO DMS Newsletter*, 96, 1-2.
- Terrell, G. R. (1990), "The Maximal Smoothing Principle in Density Estimation," *Journal of the American Statistical Association*, 85, 470-477.
- Wade, N. (1994), "Method and Madness: When Experts Err," *The New York Times Magazine*, February 6, 1994, p. 12.
- West, M. (1992), "Modelling With Mixtures," in *Bayesian Statistics 4*, eds. J. M. Bernardo et al., Oxford, U.K.: Oxford University Press, pp. 503-524.
- Withrow, D., and Angliss, R. (1992), "Length Frequency of Bowhead Whales From Spring Aerial Photogrammetric Surveys in 1985, 1986, 1989 and 1990," *Report of the International Whaling Commission*, 42, 463-467.
- Withrow, D., Burke, R., Jr., Jones, L., and Brooks, J. (1992), "Variations in  $\delta^{13}\text{C}$  Carbon Isotope Ratios in Bowhead Whale Baleen Plates Used to Estimate Age," *Report of the International Whaling Commission*, 42, 469-473.
- Wolpert, R. L., Reckhow, K. H., and Steinberg, L. J. (1992), "Bayesian Hierarchical Models for Transport and Fate Modeling," in *Bayesian Statistics in Science and Technology: Case Studies*, eds. J. S. Hodges, R. Kass, C. Gatsonis, and N. Singpurwalla, New York: Springer Verlag, pp. 241-296.
- Zeh, J. E., George, J. C., Raftery, A. E., and Carroll, G. M. (1991), "Rate of Increase, 1978-1988, of Bowhead Whales, *Balaena mysticetus*, Estimated From Ice-Based Census Data," *Marine Mammal Science*, 7, 105-122.