Alternative Bayesian Synthesis Approaches to 
Bering-Chukchi-Beaufort Seas Bowhead Whale 
Stock Assessment: Uncertainty in Historic Catch 
and Hitting with Fixed MSYR 

Geof H. Givens and Sandra E. Thompson¹ 

1 May 1995 

¹Geof H. Givens is Assistant Professor of Statistics and Sandra E. Thompson is Graduate Research Assistant, both at Colorado State University, Fort Collins, CO 80523.
Abstract

We report the results of two investigations requested by the Scientific Committee regarding alternative methods for Bayesian synthesis assessment of the Bering-Beaufort-Chukchi Seas stock of bowhead whales (*Balaena mysticetus*).

The first series of studies examines the effect of assuming that the historic pelagic catch record is known without error. We propose a model for variability in the historic catch record and we investigate the potential influence of catch record uncertainty on assessment results. We also examine the influence of systematic bias in the record. Our analyses indicate that variability in the kill record is not a dominant source of uncertainty in the Bayesian synthesis results. Consistent, systematic bias in the record does effect conclusions about initial stock size, depletion, and maximum sustainable yield. There is evidence against the hypothesis that the true historic catch was substantially greater than previously assumed, but the combined analysis can not rule out the possibility that the true catch was less than previously assumed.

The second series of studies examines an alternative method for fitting population trajectories from the BALEEN II (de la Mare and Cooke, 1993) population dynamics model. This method, known as ‘hitting with fixed MSYR’, avoids the need to make assumptions about the unexploited stock size. Our analyses indicate that the use of this method has a substantial impact on the results. This impact may be due to a constraint which prevents population extinction even when biological parameters would otherwise favor extinction. Our analyses indicate that the differences between methods are not due primarily to the assumptions about initial stock size used by the Scientific Committee in their original analysis. Addition of the Scientific Committee evidence about initial stock size to the hitting analysis does not change the results. Our analyses also show that assumptions about population model inputs pre-determine the results of a Bayesian synthesis analysis for the hitting method to a larger extent than for the original method.
# Contents

LIST OF SYMBOLS AND ACRONYMS ................................................. ii

1 INTRODUCTION AND SUMMARY .................................................. 1

I CATCH RECORD UNCERTAINTY .................................................. 2

2 THE HISTORIC CATCH RECORD .................................................. 2
   2.1 Setup ........................................................................ 4
   2.2 A Model for Uncertain Kills, 1848-1914 ............................ 5

3 MONTE CARLO INVESTIGATION .................................................. 9
   3.1 Method ....................................................................... 9
   3.2 Results ...................................................................... 10

4 ESTIMATING KILL RECORD BIAS ................................................. 14
   4.1 Method ....................................................................... 14
   4.2 Results ...................................................................... 14

5 SENSITIVITY TO KILL RECORD BIAS ........................................... 17
   5.1 Method ....................................................................... 17
   5.2 Results ...................................................................... 17

6 CONCLUSIONS ......................................................................... 21

II HITTING WITH FIXED MSYR .................................................... 21

7 HITTING AND PROJECTION ........................................................ 21

8 COMPARISON OF RESULTS ....................................................... 22

9 COMPARATIVE INFLUENCE OF EVIDENCE ................................ 29
   9.1 Results ....................................................................... 29

10 SENSITIVITY OF HITTING TO KILL RECORD BIAS ................. 40

11 CONCLUSIONS ....................................................................... 43

ACKNOWLEDGEMENTS ............................................................... 43

REFERENCES ........................................................................... 44
LIST OF SYMBOLS AND ACRONYMS

We use certain symbols and acronyms to represent quantities related to bowhead stock assessment. This terminology has been used in previous bowheads assessments considered by the Scientific Committee (e.g. Givens, et al. 1995b).

\%
% calves \hspace{1em} 1985-1992 average annual percent calves in population
% mature \hspace{1em} 1985-1992 average annual percent mature in population
\textit{m} \hspace{1em} age at sexual maturity
MCM \hspace{1em} maximum historic ratio of calves to mature females
MSY \hspace{1em} maximum sustainable yield
\textit{MSYL}_{(mat)} \hspace{1em} maximum sustainable yield level, relative to the mature component
\textit{MSYR}_{(mat)} \hspace{1em} maximum sustainable yield rate, relative to the mature component
\textit{P}_0 \hspace{1em} initial stock size
\textit{Q} \hspace{1em} catch limit developed by Givens et al. (1995c)
ROI \hspace{1em} 1978-1993 average annual rate of population increase
RY \hspace{1em} replacement yield
\textit{s} \hspace{1em} adult annual survival rate
1 INTRODUCTION AND SUMMARY

At its 1994 meeting, the Scientific Committee (SC) of the International Whaling Commission (IWC) expressed interest (IWC, 1995) in several investigations of the Bayesian synthesis methodology (Givens, 1993; Raftery et al., 1995; Givens et al., 1995b) which it has used for assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales. Specifically, the SC:

1. Recommended that variability associated with the historic pelagic kill estimates be investigated, and taken account of in the modeling process, and

2. Agreed that an alternative use of the Bayesian synthesis method which incorporated 'hitting with fixed MSYR' could be examined.

This paper presents results from investigations of both topics. The investigations are detailed and the results lengthy, therefore we draw forward the key conclusions from our work, listed below:

- Interannual variability can be estimated for the historic bowhead catch record, despite the apparent lack of relevant data.

- The results of Bayesian synthesis bowhead stock assessment are generally not much affected by incorporating interannual variability in the historic catch record.

- Variation in the initial stock size may be an approximate and adequate surrogate for interannual variation in the historic catch record, for the purposes of Bayesian synthesis bowhead assessment.

- Bayesian synthesis estimates of initial stock size, maximum sustainable yield, and depletion are sensitive to systematic, constant bias in the historic catch record.

- Bayesian synthesis estimates of replacement yield, maximum sustainable yield rate, recent rate of population increase, and most other parameters are not very sensitive to systematic, constant bias in the historic catch record.

- The available bowhead data indicate that it is unlikely that the true historic catch was substantially greater than previously assumed, but cannot rule out the possibility that the true catch was less than previously assumed unless assumptions about initial stock size do not vary with changes in the historic catch record. Since management advice
would present a greater risk to the stock if the former possibility was supported by the data, sensitivity of assessment results to systematic kill record bias is less of a concern than it would otherwise be.

- The ‘hitting with fixed MSYR’ method for Bayesian synthesis assessment produces very different results than the ‘projection’ method used previously by the SC.

- The hitting method prevents population extinction even for model trajectories where the biological parameters would otherwise favor extinction. The effect of this constraint is difficult to isolate, but is believed to be at least partially responsible for the differences between hitting and projection.

- The use of pre-model evidence about initial stock size does not appear to be the cause of the differences between hitting and projection.

- The addition of pre-model evidence about initial stock size into the hitting approach does not change the results, compared to when this evidence is ignored and hitting is used.

- When the hitting method is used, the assessment results are much more pre-determined by the uncertain evidence about population model inputs, compared to when the projection method is used.

- The data on population model outputs are much more influential for the projection method than for the hitting method.

- Bayesian synthesis results for the hitting and projection methods both change in the same ways when systematic kill record bias is assumed, however the hitting method exhibits less Monte Carlo variability.

Part I
CATCH RECORD UNCERTAINTY

2 THE HISTORIC CATCH RECORD

Bockstoce and Botkin (1983) compiled annual data on pelagic bowhead whaling in the Bering, Chukchi, and Beaufort Seas during 1849-1914, gathering information on not only
the number of bowheads struck and killed, but also the number of whaling ships present and
the number and weather conditions of days that were spent whaling. The authors examined
approximately 25,000 pieces of information, including approximately 4,000 log books and
journals from the U.S., Canada and Australia. They found record of 2,716 different whaling
voyages, and of these, only 516 voyages included records of all kills by the ship and records of
weather conditions throughout the voyage. These 516 voyages will be called ‘documented’;
the remaining voyages are ‘incompletely documented’. Bockstoce and Botkin (1983) (p.
111) state that the 516 documented voyages were “spread relatively evenly” throughout the
period of commercial whaling. The authors examined the data for recording bias towards
either successful or unsuccessful voyages, and concluded none was apparent.

Under the simplistic assumption that all ships had equal probability of being discovered
by Bockstoce and Botkin, the expected number of ships not discovered by Bockstoce and
Botkin is less than one (Lewontin and Prout, 1956), given the search effort and results they
report. Certainly, the probability of discovering record of a ship varies between ships, and
hence the number of undiscovered ships is probably larger. However, Bockstoce and Botkin
(1983) believe that they have discovered at least 99% of the bowhead whaling voyages during
the period of commercial harvest from 1848-1914. For this reason, we will consider the annual
numbers of voyages to be known exactly. We also condition on the number of voyages which
were documented for each year. We treat as random the (unknown) numbers of whales killed
on incompletely documented voyages.

Bockstoce and Botkin (1983) estimate the total number of pelagic kills in a year by
dividing the number of whales killed on documented voyages by the proportion of all voyages
in that year which were documented. If the total number of voyages in a year is incorrect,
then the estimate of the number of whales killed is also incorrect. Conditioning on the
number of documented voyages can lead to errors, too. If the total number of all voyages
or documented voyages in a year is small, a change in one of these numbers can cause a
large change in the proportion of documented voyages, which is used as a correction factor
to estimate total pelagic kill. If the number of pelagic kills in a year is zero, the addition
of a single documented voyage with any kill could cause the estimated total pelagic kill to
be inflated to substantially more than zero. Generally, however, we feel that the amount of
uncertainty in the total number of voyages and documented voyages in each year is small
compared to the uncertainty in the number of whales killed per voyage. Therefore, we
concentrate on the latter source of uncertainty.

Breiwick et al. (1984) provide data on the shore-based kill of the stock during the
same period. The numbers of shore-based kill are small, and we have no indication of the uncertainty in these data. We assume that variation in the shore-based kill is small relative to the uncertainty of to the undocumented pelagic kill.

2.1 Setup

Let the years between 1848 and 1914 be indexed by $t$. The SC recommended investigation only of uncertainty in the historic pelagic catch. We know of no data which could be used to estimate uncertainty after 1914, therefore we confine our uncertainty analysis to the period of commercial pelagic whaling, 1848-1914, and treat the later kill data as exact. Annual kills are small after 1914, and recent kills are known almost exactly, so this approach should be reasonable and meet the SC request.

Suppose that in year $t$, there were $N_t$ whaling voyages, of which $n_t$ were documented, and that these documented voyages yielded pelagic kills of $X_{1,t}, \ldots, X_{n_t,t}$. Suppose that the incompletely documented voyages yielded pelagic kills of $X_{n_{t}+1,t}, \ldots, X_{N_t,t}$. Suppose that the shore-based kill in year $t$ was $S_t$. We condition on the fixed values of $n_t$ and $N_t$ throughout our analysis. Let $D_t = \sum_{i=1}^{n_t} X_{i,t}$ and $U_t = \sum_{i=n_{t}+1}^{N_t} X_{i,t}$ be the total numbers of kills from documented and incompletely documented voyages in year $t$, respectively, and let the total number of whales killed be $K_t = S_t + D_t + U_t$.

Bockstoce and Botkin (1983) estimate the annual pelagic bowhead kill from 1848 through 1914 by dividing $D_t$, the total number of whales killed on voyages for which they found complete documentation, by the proportion of documented voyages in year $t$, $n_t/N_t$. Thus, their estimate for the total number of pelagic kills in year $t$ was $N_t D_t/n_t$. Breiwick et al. (1984) estimated the total number of whales killed to be $S_t + N_t D_t/n_t$. The entire kill record from 1848 to the present, based on the Breiwick et al. (1984) method, is shown in Fig. 1. The SC has previously treated this as the 'true record', known without error.

The goal of our analysis is to simulate alternative values for $K_t$ from a distribution with mean and variance which reflect our knowledge and uncertainty about the true total annual kill. Although the true number of whales killed from documented voyages and shore-based operations may be slightly different than $D_t$ and $S_t$, the dominant source of uncertainty about $K_t$ arises from failing to observe $X_{n_{t}+1,t}, \ldots, X_{N_t,t}$ or $U_t$. Bockstoce and Botkin (1983) are confident that their extensive research effort produced $D_t$ which are correct or very close to correct, and they assume that errors due to misreporting and to unknown outcome for struck and lost whales are a negligible source of uncertainty in these data, therefore, we condition on $D_t$. We also condition on the few shore-based kills, $S_t$. Thus, the randomness
in $K_t$ is confined to the $U_t$ component.

2.2 A Model for Uncertain Kills, 1848-1914

Suppose the $X_{i,t}$ for $i = 1, \ldots, N_t$ are independent random variables with mean $\mu_t$ and variance $\sigma_t^2$. In fact, the pelagic kills on different ships in the same year might be somewhat dependent, but we assume this covariance is small relative to the variability of a ship's catch and to other sources of variability described below. Bockstoce and Botkin’s (1983) estimate of annual pelagic kill is $D_t + (N_t - n_t)\hat{\mu}_t$ where $\hat{\mu}_t = D_t/n_t$. The Breiwick et al. (1984) estimate of annual total kill is $S_t + D_t + (N_t - n_t)\hat{\mu}_t$. Writing $K_t = S_t + D_t + \sum_{i=n_t+1}^{N_t} X_{i,t}$, it is clear that the conditional mean of $K_t$ given $(S_t, D_t)$ is $S_t + D_t + (N_t - n_t)\mu_t$ and the conditional variance of $K_t$ given $(S_t, D_t)$ is

$$\tau_t^2 = (N_t - n_t)\sigma_t^2$$

when $\sigma_t^2$ is known. We use the conditional distribution of $K_t$ given $(S_t, D_t)$ to represent the evidence and uncertainty about $K_t$. 

Figure 1: Annual bowhead kill, 1848-1993.
In order to simulate alternative kill records, we must estimate \( \tau_t^2 \), which reduces to the problem of estimating \( \sigma_t^2 \). If we observed some or all of the \( X_{i,t} \), we could estimate \( \sigma_t^2 \) by the sample variance of these \( X_{i,t} \). However, Bockstoce and Botkin (1983) report only \( N_t, n_t \), and \( D_t \) for each year; the \( X_{i,t} \) are unavailable. Therefore, the simple sample variance estimator for \( \sigma_t^2 \) is not available.

From the data of Bockstoce and Botkin (1983), we can also extract \( \omega_t \), a measure of the hunting conditions in year \( t \), which is the average number of ‘Type 1’ days for ships in year \( t \). A Type 1 day is a day of whaling effort by a single ship under weather conditions during which whales were caught during the 66-year period of commercial harvest. Days with a combination of ice, wind, and visibility conditions under which whales were never caught are not Type 1. For three years for which \( \omega_t \) was unavailable, we estimated \( \omega_t \) as the mean of the values for the two nearest years for which \( \omega_t \) was available. These three years also had no documented kills. In this case, the direct application of the model we describe below would assume zero kills on undocumented voyages also, and there would be no possible variability in the catch for these years. In order to permit some variability in these years, we replaced \( \hat{\mu}_t \) by \( ((D_{t-1}/n_{t-1}) + (D_{t+1}/n_{t+1}))/2 \), i.e. the mean of the numbers of kills per documented voyage for the two nearest years for which \( D_t \) was not zero. This retains zero kills for documented ships in year \( t \), but makes it possible for undocumented ships in the same year to have non-zero kills.

Our estimate of \( \sigma_t^2 \) is based on the variability of the \( D_t \). There are several reasons why the \( D_t \) vary from one year to the next. These include (i) different number of documented voyages, \( n_t \), between years, (ii) different hunting conditions, \( \omega_t \), between years, (iii) population depletion and dynamics, (iv) variation in whale behavior between years, (v) variation between years in the efficiency of documented ships relative to incompletely documented ships, and (vi) random variability in hunting success. To obtain an estimate of \( \sigma_t^2 \), we consider \( D_{t-1} \), \( D_t \), and \( D_{t+1} \). It is unlikely that substantial population dynamics (iii) could occur in three short years. Bockstoce and Botkin (1983, p. 114 and p. 117) found no evidence of variability from (v) in their analyses. Thus, we assume that over these three years, the variation due to (iii), (iv) and (v) is minimal compared to other sources of variability; however to the extent it exists, it should bias our variance estimator upwards, as discussed below.

Next, to adjust for (i) and (ii), we rescale the three documented pelagic kill totals for weather and number of documented voyages, relative to year \( t \), namely

\[
D'_{t-1} = \frac{\omega_t n_{t-1} D_{t-1}}{\omega_{t-1} n_{t-1}} \quad D'_t = D_t \quad D'_{t+1} = \frac{\omega_t n_{t+1} D_{t+1}}{\omega_{t+1} n_{t+1}}.
\]  

(2)
With this adjustment, \( D'_{t-1} \) and \( D'_{t+1} \) should be roughly equal to the numbers of whales that would have been documented pelagic kills in years \( t - 1 \) and \( t + 1 \) if the weather conditions and documentation rates for these years had been identical as for year \( t \). Furthermore, the variances of \( D'_{t-1} \), \( D'_{t} \), and \( D'_{t+1} \) should then be approximately equal, and the variance of each should be approximately \( n_t \sigma_t^2 \), since we have accounted for all variation listed above except (vi).

We obtain the estimate \( \hat{\sigma}_t^2 \) by rescaling the sample variance of \( D'_{t-1} \), \( D'_{t} \), and \( D'_{t+1} \). For 1848 and 1914, there was only one neighboring time point used in the sample variance calculation. Finally, we estimate

\[
\hat{\sigma}_t^2 = (N_t - n_t) \hat{\sigma}_t^2. \tag{3}
\]

The negative binomial distribution\(^1\) can be used to model the distribution of the unobserved \( U_t \), i.e.

\[
Pr[U_t = k] = \frac{\Gamma(\alpha_t + k)}{\Gamma(\alpha_t) \Gamma(k + 1)} \left( \frac{\beta_t}{\beta_t + 1} \right)^k \left( \frac{1}{\beta_t + 1} \right)^{\alpha_t} \tag{4}
\]

where \( \Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx \) is the gamma function. Letting \( U_t \sim NegBin(\alpha_t, \beta_t) \) with mean \( \alpha_t \beta_t \) and variance \( \alpha_t \beta_t (\beta_t + 1) \), we calculate the method of moments estimators \( \hat{\alpha}_t \) and \( \hat{\beta}_t \) by solving \( \alpha_t \beta_t = (N_t - n_t) \hat{\mu}_t \) and \( \alpha_t \beta_t (\beta_t + 1) = \hat{\sigma}_t^2 \). This permits us to simulate alternative \( K_t \) as \( S_t + D_t + Y \) where \( Y \sim NegBin(\hat{\alpha}_t, \hat{\beta}_t) \).

In several cases, we obtained \( \hat{\mu}_t > \hat{\sigma}_t^2 \), which does not yield a valid solution for the parameters of the negative binomial model. In these cases, we switched to a Poisson model. The largest adjustment was for 1867, in which 202 whales were documented killed, and 397 are expected to have been undocumented kills with a standard error which is effectively inflated from 9.3 to 19.9 by switching models. This approach can be justified by conceiving of the negative binomial model as a Poisson-gamma mixture, where, in our case, ships exhibit Poisson kill variation, and the gamma mixing distribution represents between-ship heterogeneity. In this case, our approach amounts to insisting on at least Poisson variation in every case.

Four desirable features of this method are:

1. The amount of uncertainty in the kill record is allowed to vary by year,

\(^1\)More formally, we may consider that \( X_{i,t}, i = n_t + 1, \ldots, N_t, \) are distributed as Poisson random variables with means \( \lambda_t \), with extra-Poisson variation represented by treating the \( \lambda_t \) as gamma random variables with shape parameters \( r_t \) and a common scale parameter \( c \). Then \( U_t \sim Poisson(\sum_i \lambda_t) \) where \( \sum_i \lambda_t \sim \text{gamma}(\sum_i r_t, c) \), i.e. \( U_t \) has a negative binomial distribution. The negative binomial model is a reasonable choice for simulation even when these modeling assumptions are not satisfied, since it is a distribution on the positive integers with mean and variance which are matched to the estimated values.
2. The uncertainty is skewed, since $K_i$ is bounded below by $S_t + D_t$ but not bounded above,

3. The annual uncertainty estimates are negatively correlated with the proportions of documented voyages, and positively correlated with the numbers of undocumented voyages, so years for which few voyages are documented have higher uncertainty, all else being equal, and

4. Underestimation of the true uncertainty is unlikely, since between-year heterogeneity (e.g. due to population dynamics, whale behavior, or efficiency misrepresentation) which has not been accounted for by the estimation procedure should tend to bias $\tau_i^2$ upwards.

In addition to the 3-year span running variance estimate described above, we tried a 5-year span estimator, which produced uncertain kill records which were qualitatively similar to those from the 3-year span analysis. This suggests that the assumption of negligible variability in kill due to population depletion and dynamics, (iii), is reasonable.

Fig. 2 shows the 95% probability interval for simulated values of $K_i$ for each year. This figure shows that the method we have described allows for a large amount of uncertainty in the historic kill record. The mean $CV_i = \bar{\tau}_i/(S_t + D_t + \dot{\alpha}_i \dot{\beta}_i)$ from 1848 through 1914 is 0.23, and the median is 0.14. The probability intervals shown in Fig. 2 are generally skewed; for example in 1850, the estimated total kill is 2,067 whales, and the 95% probability interval is (1,323, 3,022). The dominant feature of this graph is the large amount of uncertainty in the initial years of the hunt. This reflects the huge kill in these years, and the large and highly variable catch per unit effort in these years. Second, the result for 1870 seen in this graph is reassuring. In 1870, the estimated number of kills is clearly quite high, compared to the trend in neighboring years. Consequently, the result of our uncertainty model yields much greater variability for this year than for adjacent years, reflected in the greater length of the corresponding prediction interval. Finally, the years around 1880 also show unusually large variability. These years have unusually low documentation rates and atypical catch per unit effort, perhaps partially due to the advent of steam whaling. Bockstoce and Botkin (1983, p.124) note that these years have "relatively poor" catch per unit effort documentation. Thus, the results shown in Fig. 2 appear quite reasonable, and very consistent with what is known about the historic kill record.
3 MONTE CARLO INVESTIGATION

3.1 Method

We use the model for historic catch uncertainty presented above to examine the effect of such uncertainty on the results of the Bayesian synthesis bowhead assessment. We ran a complete Bayesian synthesis assessment, based on 200,000 trajectories drawn from the pre-model distributions of the inputs. However, for these 200,000 trajectories, we no longer treated the kill record as fixed. Instead, each trajectory was run with an alternative kill record drawn by simulating new \( K_t \) \((t = 1848, \ldots, 1914)\) from the \( \text{NegBin}(\hat{\alpha}_t, \hat{\beta}_t) \) distributions determined by our uncertainty model. The kills for subsequent years were unchanged. The trajectories were then compared to data on outputs, and reweighted to obtain final inference from the combined evidence. The conclusions from this analysis, therefore, reflect the additional source of uncertainty due to the historic kill record.

To carry out this analysis, we used the SC consensus pre-model distributions for all inputs and outputs of the BALEEN II model except for \( P_0 \). In the original analysis, the pre-model
distribution for $P_0$ was estimated from the (fixed) kill record. In order to fully reflect the uncertainty in our results due to kill record variability, we re-estimated the distribution for $P_0$ separately for each alternative kill record. The original SC consensus distribution for $P_0$ was based on the DeLury (1947) method, and was a fiducial distribution (Fieller, 1954) obtained for the ratio of the slope and intercept from a linear regression analysis based on the kill record (Givens et al., 1995b). It was beyond our computing capabilities to re-estimate this distribution 200,000 times. Therefore we switched to a different method.

Givens (1995) proposes an alternative distribution for $P_0$:

A simple upper bound for $P_0$ is the upper bound for the current population size plus the total of all historic kills. We use 1988 as the ‘current’ year, since this was the SC’s choice (IWC, 1995b) and since the results of the 1988 census (Raftery and Zeh, 1991; Raftery and Zeh, 1993) provide a higher upper bound for population than the results of the 1993 census (Raftery and Zeh, 1994; Zeh et al., 1995). This bound assumes that there is no density dependence in the direction opposite from that usually assumed, e.g. an Allee effect (Allee, 1931; Allee, 1941; Allee et al., 1949). See Fowler and Baker (1991) for a review of such dynamics. A simple lower bound is the maximum number of whales killed in any single year. Using this reasoning, $P_0$ must be between 2,709 and 31,158.

These bounds were conceived by Adrian Raftery and Greg Donovan, and conveyed via personal communication. Thus, for the original kill record, we assume $P_0 \sim \text{Unif}(2,709, 31,158)$.

The bounds of this uniform distribution can be quickly and efficiently recalculated for each of the 200,000 alternative kill records. Therefore, we take the pre-model distribution of $P_0$ to be the mixture distribution obtained by using the recalculated uniform distributions in combination with the varying historic kill record. Specifically, we may consider the historic kill record to be an additional (high-dimensional) parameter, say $R$, in the Bayesian synthesis analysis. Then the negative binomial model for kill record uncertainty constitutes the pre-model distribution for $R$, say $p(R)$. If $p(P_0 \mid R)$ is the pre-model distribution obtained by calculating the uniform bounds for a given value of $R$, then we use a joint pre-model distribution for $P_0$ and $R$ in our analysis, namely $p(P_0, R) \propto p(P_0 \mid R)p(R)$, and the marginal pre-model distribution for $P_0$ is proportional to $\int p(P_0, R) \, dR$. The marginal pre-model distribution for $P_0$ using this approach is very flat; a sample of 200,000 points from it is shown in Fig. 3.

### 3.2 Results

Our Bayesian synthesis results in this section and the rest of Part I are based on an initial sample size of 200,000 and a final sample size of 5,000.
Figure 3: Sample from the marginal pre-model distribution for $P_0$ determined by the negative binomial model for kill variability and the uniform bound calculations.

Fig. 4 shows post-model distributions of $RY$ and $Q$ resulting from three analyses. The quantity $Q$ is introduced by Givens et al. (1995c) as an implementation of the existing Aboriginal Management Scheme for subsistence whaling for bowheads. According to this implementation, $Q$ is an upper bound for the allowable bowhead subsistence take. The solid line shows the results of the SC consensus assessment. The short-dashed line shows the results when the pre-model distribution for $P_0$ is switched from the SC choice to a uniform distribution between 2,709 and 31,158. Recall that this switch was made out of necessity in order to have baseline results for direct comparison with the results obtained from implementing the negative binomial model for catch record uncertainty. The long-dashed line shows the results when the catch record uncertainty is included in the analysis. This figure shows that all three sets of results are nearly identical.

Fig. 5 shows post-model distributions for $MSY R_{(mat)}$, $P_0$, $ROI$, mature depletion, $s$, and % calves. The meaning of the different line types is the same as in Fig. 4, and, additionally, the SC consensus pre-model distribution is shown with the dotted line. Again, Fig. 5 shows that there are virtually no differences between the results of the three analyses. The same conclusion holds for inputs and outputs not pictured here.
Figure 4: Post-model distributions of RY and Q, for the SC consensus analysis (solid), SC analysis but with $P_0 \sim \text{Unif}(2,709, 31,158)$ (short dashes), and analysis with negative binomial model for uncertain catch record, uniform $P_0$ distributions, and the rest of the analysis as done by the SC (long dashes).

These results indicate that the Bayesian synthesis assessment is quite insensitive to potential inter-annual variability in the catch record. At first glance, this conclusion may seem weakened by our need to use the uniform distribution for $P_0$ in the analysis. While it would be ideal to recalculate the pre-model distribution for $P_0$ given an alternative kill record using the same model and estimation method used in the SC analysis, we feel the our reliance on uniform distributions is acceptable, given the results we obtained. The results show that the uniform distribution without kill record uncertainty produces results no different from the consensus SC results, and adding kill record uncertainty still does not change things. Thus, not only do we conclude that the SC results appear insensitive to kill record uncertainty, but also that using a broad uniform distribution for $P_0$ may be an adequate surrogate for directly incorporating kill record uncertainty.
Figure 5: Post-model distributions of $MSY_{(mat)}$, $P_0$, $ROI$, mature depletion, $s$, and % calves, for the SC consensus analysis (solid), SC analysis but with $P_0 \sim \text{Unif}(2,709, 31,158)$ (short dashes), and analysis with negative binomial model for uncertain catch record, uniform $P_0$ distributions, and the rest of the analysis as done by the SC (long dashes). The pre-model distributions are shown with the dotted lines.
4 ESTIMATING KILL RECORD BIAS

4.1 Method

The negative binomial method is just one way to examine kill record uncertainty. A second, simpler method involves treating the kill record as if there was systematic bias in the record for the years 1848-1914. We consider a constant bias, \( \alpha \), for all years 1848-1914, and form a new kill record by letting the newly hypothesized kill in year \( t \) be \( K_{t,\alpha} = \alpha K_t \) for \( t = 1848, \ldots, 1914 \). We can re-estimate a pre-model distribution for \( P_0 \) based on this record, for any given value of \( \alpha \).

We treat \( \alpha \) as an additional unknown parameter to be estimated using the Bayesian synthesis analysis. We let the pre-model distribution of \( \log \alpha \) be uniform between \( \log 0.5 \) and \( \log 2 \). The log transform is used so that, say, doubling and halving the kill record are treated equivalently, and so that the probability that \( \alpha < 1 \) and \( \alpha > 1 \) are equal. The pre-model distributions of all other quantities are the same as those used by the SC, except for \( P_0 \). The pre-model distribution of \( P_0 \) is the mixture distribution obtained by using the DeLury (1947) method from the original analysis (Givens et al., 1995b) in combination with the pre-model distribution for \( \alpha \). In other words, if \( p(P_0 \mid \alpha) \) is the pre-model distribution obtained from the DeLury analysis for a given value of \( \alpha \), then we use a joint pre-model distribution for \( P_0 \) and \( \alpha \) in our analysis, namely \( p(P_0, \alpha) \propto p(P_0 \mid \alpha)p(\alpha) \), and the marginal pre-model distribution for \( P_0 \) is proportional to \( \int p(P_0, \alpha) \, d\alpha \). Unlike the negative binomial case above, the original DeLury (1947) method for obtaining the pre-model distribution of \( P_0 \) (Givens et al., 1995b) can be easily re-applied for new kill records resulting from alternative values of \( \alpha \). The new pre-model distribution is simply a scaled version of the original distribution, where the scale factor is \( \alpha \). It is not necessary to resort to uniform distributions for \( P_0 \) in this study. Therefore, each alternative kill record is treated here in the exactly the same manner as it would have been in the original SC analysis.

Each trajectory considered by Bayesian synthesis uses a different random value of \( \alpha \) to adjust the kill record (and hence the pre-model distribution for \( P_0 \)). As before, trajectories are compared to data on outputs to determine which are most reasonable. The results include post-model distributions for all quantities of interest, including \( \alpha \) itself. With this approach, we can determine what values of \( \alpha \) receive the most support from the combined data, and which values are in conflict with the combined data.

4.2 Results
Figure 6: Post-model distribution for $\alpha$, the kill record bias parameter, is shown with the histogram. The pre-model distribution is shown with the dotted line.

Fig. 6 shows the results for estimating $\alpha$. The solid line shows the pre-model distribution we used for $\alpha$, and the histogram represents the post-model distribution. Where the histogram is above the line, these values of $\alpha$ are more consistent with the combined data. Where the histogram is below the line, the values of $\alpha$ do not agree well with the combined data. This figure shows that the post-model distribution assigns less probability to $\alpha > 1$ than the pre-model distribution. This is evidence against the possibility that the actual historic kills were much greater than originally assumed ($\alpha = 1$).

On the other hand, Fig. 6 seems to support the possibility that the actual kills were less than originally assumed. We believe this result should be interpreted with caution, since systematic kill record bias is confounded with other biological and population parameters because $P_0$ is re-estimated for each value of $\alpha$. Specifically, we believe the result can be explained in terms of MSY. Unlike $P_0$, MSY does not scale linearly with $\alpha$. Indeed, when the kill record is halved ($\alpha = .5$), MSY is only reduced by 30%. Therefore, when $\alpha = .5$, MSY is larger relative to the population and to kills than it is when $\alpha = 1$. In other words, when we halve the initial stock size and the kills, MSY is not as greatly reduced, the stock is more likely to survive than previously. Therefore, such trajectories are more likely to appear
Figure 7: Post-model distribution for $\alpha$, the kill record bias parameter, is shown with the histogram. For this graph, the pre-model distribution of $P_0$ was not re-estimated for each value of $\alpha$, but instead fixed at the SC consensus choice. The pre-model distribution for $\alpha$ is shown with the dotted line.

in the post-model sample. We believe this explains the left side of Fig. 6.

We can remove the confounding between $P_0$ and $\alpha$ by forcing $P_0$ and $\alpha$ to be pre-model independent. We allowed $\alpha$ to vary according to the same marginal pre-model distribution for $\alpha$ used above, but we retained the original SC consensus marginal pre-model distribution for $P_0$, which is the same for any $\alpha$. Fig. 7 shows the resulting post-model distribution for $\alpha$ from this analysis. These results show that, given the SC consensus distribution for $P_0$ which favors initial stock sizes in the 10,000–20,000 range, there is evidence against severe historic kill record bias in either direction. Values of $\alpha$ near 1 (no bias) are favored, and values far from 1 receive reduced post-model probability.
5 SENSITIVITY TO KILL RECORD BIAS

5.1 Method

Our third approach is not an estimation procedure but a systematic exploration of how Bayesian synthesis estimation results depend on the kill record. Here, we again suppose that the actual kill in year $t$, $K_{t,\alpha}$ is some fixed fraction of the estimated kill, $K_t$, across all years 1848-1914. In other words, we let $K_{t,\alpha} = \alpha K_t$ for several different values of $\alpha$. For each choice of $\alpha$, we generated the new kill record, estimated a pre-model distribution for $P_0$ based on these new data using the DeLury (1947) method used in the original analysis (Givens et al., 1995b), and ran a complete Bayesian synthesis analysis with the new kill data and the new distribution for $P_0$. Thus, the effect of changing the kill record is twofold: it changes the kills removed by the population dynamics model, and it changes the assumed distribution for $P_0$.

5.2 Results

We considered possibilities from halving to doubling the historic kill record, specifically $\alpha \in \{0.5, 0.667, 0.75, 0.8, 0.825, 0.85, 0.875, 0.9, 0.925, 0.95, 0.975, 1.0, 1.025, 1.05, 1.075, 1.1, 1.125, 1.15, 1.175, 1.2, 1.25, 1.333, 1.5, 1.667, 2.0\}$, and we display the results graphically.

Fig. 8 shows how the estimated median (solid line), 95% post-model probability interval (dotted lines), and post-model 5% point (dashed line) for RY depend on the value of $\alpha$. The tick marks on the horizontal axis correspond to the different values of $\alpha$ we examined. Only one Bayesian synthesis analysis was conducted for each value of $\alpha$, so the jagged nature of the lines is mostly due to Monte Carlo variation, while systematic trends in the lines are mostly due to altering the kill record. For example, it is clear that when $\alpha$ is very low, the Bayesian synthesis estimate of RY is also lowered beyond what can be explained by Monte Carlo variability. Generally, we see a moderate trend toward increasing RY as the true kills increase compared to the estimated historic kill record, but RY does not appear extremely sensitive to alternative historic catch records.

Fig. 8 also shows results for Q, a quantity which establishes a catch limit under the implementation of the existing Aboriginal Management Scheme for bowhead subsistence whaling offered by Givens et al., (1995c). The results for Q are very similar to those for RY. The estimate of Q increases moderately as the true kills increase relative to the estimated historic record, but generally Q is fairly robust to alternative historic kill records, and the 5th percentile of Q is particularly robust.

17
Figure 8: Sensitivity of RY and Q to $\alpha$. Shown are post-model medians (solid line), 95% probability intervals (dotted lines), and 5th percentiles (dashed line). The tick marks on the horizontal axes correspond to the different values of $\alpha$ we examined.

Fig. 9 shows similar graphs for $P_0$, $MSY R_{(mat)}$, ROI, mature depletion, MSY, and $MSY L_{(mat)}$. Not surprisingly, we find that the post-model estimates of $P_0$ and MSY, and the uncertainty of these estimates, increase as the historic kill record is inflated. Due to the effect on $P_0$, the estimate of depletion declines with increasing $\alpha$. Fig. 9 also shows that for very low $\alpha$, the analysis favors slightly higher values for $MSY R_{(mat)}$ and $MSY L_{(mat)}$, though these the post-model estimates of these quantities are very insensitive to nearly all changes in the kill record. The estimate of ROI is almost entirely unchanged by alterations to the kill record.

Fig. 10 shows the results for other biological and population parameters. We find that estimates of the age at sexual maturity ($m$), the % calves, and $P_{1988}$ do not vary when $\alpha$ is varied. Little influence is also seen for estimates of survival ($s$), % mature, and MCM, though increasing $\alpha$ appears to be associated with some reduction in MCM and % mature and some increase in $s$.

The dominant effect shown by these results is that variation in the kill record can influence the Bayesian synthesis post-model estimate of $P_0$, MSY, and depletion, but other quantities are relatively robust to moderate changes in the kill record.
Figure 9: Sensitivity of $P_0$, $MSY_{R(mat)}$, $ROI$, mature depletion, $MSY$, and $MSYL_{(mat)}$ to $\alpha$. Shown are post-model medians (solid line) and 95% probability intervals (dotted lines). The tick marks on the horizontal axes correspond to the different values of $\alpha$ we examined.
Figure 10: Sensitivity of $s$, $m$, $P_{1988}$, $MCM$, % calves, and % mature to $\alpha$. Shown are post-model medians (solid line) and 95% probability intervals (dotted lines). The tick marks on the horizontal axes correspond to the different values of $\alpha$ we examined.
6 CONCLUSIONS

The results of our investigations in Part I generally support the conclusion that potential variability in the historic catch record should not be a great concern for Bayesian synthesis assessment purposes. When kill record variability is incorporated into the assessment, the results are effectively unchanged.

In contrast, if a systematic bias existed in the kill record, some assessment results could be affected. Bayesian synthesis estimates of $P_0$, $MSY$, and mature depletion would be most affected. $RY$ is only moderately affected, and $Q$ even less so. Our results indicate that systematic downward bias in the kill record is unlikely to exist, i.e. the kills are unlikely to have been more than believed. This is fortunate, because kill bias in this this direction would be more likely to produce assessment results and management advice which endangered the stock. It is possible that the historic kills were less than believed, however support for this possibility is confounded with other factors, and one analysis shows evidence against both positive and negative kill record bias.

We believe it is unlikely that the existing data can be used to shed much more light on the question of what bias and uncertainties, if any, exist in the historic kill record.

Part II

HITTING WITH FIXED MSYR

7 HITTING AND PROJECTION

Butterworth and Punt (1995) have proposed an alternative method of fitting trajectories from the BALEEN II (de la Mare and Cooke, 1993) population dynamics model. This method, 'hitting with fixed MSYR', can be incorporated within the Bayesian synthesis framework and avoids the need to choose a pre-model distribution for the unexploited stock size, $P_0$. Butterworth and Punt (1995) have argued that the results of a Bayesian synthesis assessment of the bowhead stock can be sensitive to the choice of pre-model distribution for $P_0$, and that the particular choice used by the Scientific Committee in previous analyses (IWC, 1995; Givens et al., 1995b) is subject to unknown bias. Givens, Raftery, and Zeh (1995a) have argued that the sensitivity is not extreme, and that the amount of bias is probably small compared to the corresponding uncertainty. Application of the alternative hitting approach to Bayesian synthesis assessment would sidestep this debate.
The hitting approach to Bayesian synthesis assessment treats $P_0$ as a output of the modeling effort, rather than an input, and treats current population size as an input, rather than an output. A pre-model distribution is established for relevant biological and operational parameters and for current population size, say $P_{1988}$. Samples from this pre-model distribution are drawn. For each sample, a trajectory is fit using the sampled parameter values. This trajectory is fit using the 'hitting with fixed MSYR' component of the Hitter-Fitter population modeling and estimation package (de la Mare, 1989; Punt and Butterworth, 1991). The fit is achieved by solving for the value of $P_0$ which yields a trajectory which passes through the sampled value of $P_{1988}$, given the sampled value of MSYR and the other biological and operational parameters. As is done in the original 'projection' approach to Bayesian synthesis, the trajectories are then compared to available data on outputs, and a post-model distribution is calculated to summarize how likely each simulation is, given the combined evidence on inputs and outputs. In the hitting approach, $P_0$ is an output, so choosing a pre-model distribution for $P_0$ remains an option, though it is no longer required.

8 COMPARISON OF RESULTS

We first compare results from the hitting and projection approaches to Bayesian synthesis assessment using the consensus pre-model distributions of the Scientific Committee (IWC, 1995), except for the distribution for $P_0$, which is used only for the projection approach. We choose to ignore the distribution of $P_0$ for the hitting method because objections to it were the motivation behind proposal of the hitting method for Bayesian synthesis. Hitting was proposed to avoid use of an explicit distribution for $P_0$. In this and subsequent sections, the initial Bayesian synthesis sample sizes were 100,000 and 200,000, respectively, for hitting and projection, and the final sample sizes were 5,000 in both cases. A smaller initial sample size can be used for hitting because no trajectories are wasted by going extinct.

Fig. 11 shows the pointwise median and 95% probability regions for the population trajectory obtained using both methods. The shaded region in this figure represents the pointwise 95% post-model probability interval for the population trajectory using the hitting approach, with central dashed line corresponding to the pointwise median. The pointwise 95% post-model probability region for the population trajectory using the projection approach is shown by the hash lines, with central solid line corresponding to the pointwise median. This figure shows that the two methods produce qualitatively similar results. The most noticeable difference is that the hitting method produces more precise results than the projection method.
Figure 11: Pointwise 95% post-model probability regions, shaded (hitting) and hashed (projection), and pointwise post-model medians, dashed (hitting) and solid (projection), for the population trajectory.

for most of the trajectory. The results are nearly indistinguishable in the most recent 15 years. Also, the hitting method tends to support trajectories in the lower portion of the region favored by the projection method.

Fig. 12 shows the marginal post-model distributions of 1993 RY and Q obtained from the two methods. Higher values of RY are supported by the hitting method. The lower 5% point of these post-model distributions for RY are 162 and 104 for the hitting and projection methods, respectively. The corresponding 5% points for Q are 159 and 103.

Fig. 13 shows the marginal post-model distributions of several other important model inputs and outputs obtained from the two methods. This figure shows that the projection and hitting methods produce some noticeably different results. The hitting method favors much higher values of $MSY R_{(mat)}$, rate of population increase, and adult survival, and lower values of $P_0$. Table 1 shows post-model medians and 95% probability intervals for these four variables, RY, Q, MSY, and depletion of the mature population component. The marginal post-model distributions for model inputs and outputs not shown in Fig. 13 are very similar for both methods.

Although it is possible that the two methods could determine very different pairwise and high-dimensional relationships between variables while still resulting in similar post-model distributions, this did not occur in our analyses. Fig. 14 shows the relationships between $MSY L_{(mat)}$ and $MSY R_{(mat)}$ estimated from the samples from the post-model obtained with each method. The lines were obtained using the lowess nonparametric smoothing
Table 1: Comparison of results from hitting and projection methods. Post-model medians and 95% probability intervals are shown, as are 5th percentiles for $R_Y$ and $Q$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Projection</th>
<th>Hitting</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s$</td>
<td>.984</td>
<td>.990</td>
</tr>
<tr>
<td></td>
<td>(.967, .998)</td>
<td>(.972, .999)</td>
</tr>
<tr>
<td>$P_0$</td>
<td>12,866</td>
<td>11,732</td>
</tr>
<tr>
<td></td>
<td>(11,026, 18,046)</td>
<td>(10,598, 14,652)</td>
</tr>
<tr>
<td>$MSY R_{(mat)}$</td>
<td>4.16%</td>
<td>5.33%</td>
</tr>
<tr>
<td></td>
<td>(1.73%, 6.17%)</td>
<td>(3.01%, 6.89%)</td>
</tr>
<tr>
<td>ROI</td>
<td>2.21%</td>
<td>2.78%</td>
</tr>
<tr>
<td></td>
<td>(.78%, 3.38%)</td>
<td>(1.53%, 3.60%)</td>
</tr>
<tr>
<td>$R_Y$</td>
<td>197 (5%: 104)</td>
<td>250 (5%: 162)</td>
</tr>
<tr>
<td></td>
<td>(87, 303)</td>
<td>(144, 340)</td>
</tr>
<tr>
<td>$Q$</td>
<td>195 (5%: 103)</td>
<td>246 (5%: 159)</td>
</tr>
<tr>
<td></td>
<td>(86, 297)</td>
<td>(143, 336)</td>
</tr>
<tr>
<td>MSY</td>
<td>248</td>
<td>319</td>
</tr>
<tr>
<td></td>
<td>(108, 438)</td>
<td>(172, 523)</td>
</tr>
<tr>
<td>Depletion</td>
<td>.36</td>
<td>.36</td>
</tr>
<tr>
<td></td>
<td>(.28, .45)</td>
<td>(.29, .46)</td>
</tr>
</tbody>
</table>
method of Cleveland (1979) with a smoothing neighborhood spanning 66% of the data. This figure shows the most substantial difference that we were able to identify in the pairwise relationships of variables. In this figure, we see that, for $MSY_L^{(mat)}$ less than about .55, both methods produce results which associate a fixed increase in $MSY_L^{(mat)}$ with about the same amount of increase in $MSY_R^{(mat)}$, although the hitting method favors higher values of $MSY_R^{(mat)}$ overall. However, for $MSY_L^{(mat)}$ greater than .55, the hitting method associates increasing $MSY_L^{(mat)}$ with increasing $MSY_R^{(mat)}$, whereas the projection method supports little change in $MSY_R^{(mat)}$ when $MSY_L^{(mat)}$ increases.

Finally, we also ran results with the hitting method for the case when the pre-model evidence about $P_0$ agreed upon by the SC is also used. It has been suggested that the use of this evidence in the projection method but not in the hitting method is the essential reason for differences between the results from the two methods (Butterworth and Punt, 1995). We find that this is not the case for the SC analysis. If we compare the results of the hitting method with and without the SC distribution for $P_0$, we find that they are nearly identical in every case. Fig. 15 shows why. In this figure, we show a histogram of the post-model sample of $P_0$ from the analysis using the hitting method with no pre-model distribution for $P_0$. The solid line in this graph shows the density of $P_0$ from the SC consensus distribution over this range. We can imagine the results which would be obtained from hitting with the SC distribution for $P_0$ by considering reweighting the sample in the histogram with weights proportional to the density shown. This figure shows that the density is very flat over the range of $P_0$ supported by the hitting results. In fact, a majority of the sample would have
weights which differ by less than 2% from the weight of the median \( P_0 \), with 95% of the weights between -11% and +3%. Therefore, we conclude that adding the additional \( P_0 \) information to the results of the hitting analysis is completely inconsequential. Something about the hitting method itself produces results which are substantially different than the results from projection, regardless of evidence about \( P_0 \).

We do not know why the results differ, but Butterworth and Punt (1995) provide one potential explanation. They note that when the hitting method solves for \( P_0 \), the solution is not necessarily unique (Butterworth and Punt, 1995, footnote 1 and Fig. 1). If there are several potential solutions for \( P_0 \) in some cases, and if the hitting method consistently finds, say, the lowest value of \( P_0 \) which is a solution, then the results from the hitting method could produce results which differ from the projection results in the same manner we have shown in this paper. Clearly further investigation is required to identify the reasons for differences between the two methods.
Figure 13: Marginal post-model distributions for hitting (dashed line) and projection (solid line) Bayesian synthesis estimation of some model inputs and outputs. The pre-model distribution is shown with dotted lines.
Figure 14: Estimated relationship between $MSYL_{(mat)}$ and $MSYR_{(mat)}$ from the post-model distributions obtained from the hitting (dashed line) and projection (solid line) methods.

Figure 15: Histogram of post-model $P_0$ sample resulting from hitting method with no distribution for $P_0$. The SC consensus distribution for $P_0$ is superimposed. All of this sample would receive very similar weights if the SC consensus distribution were applied to the hitting results.
9 COMPARATIVE INFLUENCE OF EVIDENCE

Givens et al. (1995b) have noted that the chosen pre-model distribution of the model inputs defines an implicit probability distribution for the model outputs, since evidence about inputs can be transformed into evidence about outputs through application of the model. This implicit probability distribution was termed the *pre-weighted distribution* (Givens et al., 1995b) because a sample from it could be obtained by sampling input vectors from the pre-model distribution of inputs and running the model for each input vector, without reweighting the results according to the pre-model distribution of the outputs. Thus, the pre-weighted distribution isolates the evidence about the inputs and the evidence inherent in the model structure from the evidence about the outputs.

For the hitting method, the pre-weighted distribution will differ from the pre-model distribution due to the effect of the constraint that no trajectory is allowed to go extinct, regardless of the inputs used to produce it. For the projection method, the two distributions will differ due to the projection process of eliminating trajectories for which the corresponding model parameters yield extinction.

9.1 Results

In this section, we examine the graphs of the pre-model distribution, the pre-weighted distribution, the post-model distribution, and other intermediate results using various portions of the available evidence in an attempt to isolate and identify differences between the projection and hitting methods.

We begin by considering Fig. 16, which shows the results for $P_0$. This figure shows boxplots\(^2\) of several different distributions: (A) the pre-model distribution for $P_0$ agreed upon by the SC (IWC, 1995) ; (B) the pre-weighted distribution of $P_0$ for the projection method, obtained by sampling sets of inputs and eliminating those which lead to population extinction; (C) the pre-weighted distribution of $P_0$ for the projection method incorporating evidence about $P_{1988}$, obtained by reweighting the trajectories from (B) by the data-based SC consensus pre-model distribution for $P_{1988}$; (D) the post-model distribution of $P_0$ for the projection method; (E) the pre-weighted distribution of $P_0$ from the hitting method.

\(^2\)The top and bottom edges of the box indicate the 75th and 25th percentiles of the sample, and the line through the interior of the box indicates the median. The distance from the top edge to the bottom edge of the box is the interquartile range, so 50% of the sample falls between these bounds. The whiskers extend to the nearest sampled point within 1.5 times the interquartile range from the box edges. Some points (i.e. outliers and samples from extremely heavy-tailed distributions) may fall beyond the whiskers, but, for simplicity, they are not shown in these figures.
obtained by sampling sets of inputs for the hitting method from the SC consensus pre-model distributions and recording the \( P_0 \) 'hit' with each; (F) the pre-weighted distribution of \( P_0 \) for the hitting method incorporating evidence about \( P_0 \), obtained by reweighting the trajectories from (E) by the SC consensus pre-model distribution of \( P_0 \); (G) the post-model distribution of \( P_0 \) from the hitting method with no evidence about \( P_0 \); and (H) the post-model distribution of \( P_0 \) from the hitting method including SC consensus evidence about \( P_0 \).

A comparison of boxes A and B in Fig. 16 shows that the effect of eliminating trajectories which go extinct with the projection method is to shift the distribution towards higher values of \( P_0 \). Comparing boxes B and E shows that the effect of using the hitting method which forces every trajectory to avoid extinction, regardless of input parameters, is to narrow the distribution around low values for \( P_0 \), compared to the unconstrained projection method. Fig. 16 also shows that the evidence about outputs favors lower values of \( P_0 \) than are favored by the evidence about inputs; this is true for both hitting and projection. Butterworth and Punt (1995) criticize the rationale behind the derivation of the SC consensus distribution for \( P_0 \), and they propose the hitting method as an alternative which avoids all direct assumptions about \( P_0 \). Fig. 16 demonstrates that the hitting method still involves strong assumptions about \( P_0 \), but now these assumptions are buried in the hitting mechanism which includes the constraint that no trajectory can go extinct, regardless of the biological parameters used to produce it. Finally, a comparison of boxes G and H shows that the addition of the evidence about \( P_0 \) does not change the results of the hitting method.

Fig. 17 shows results for \( MSY R_{(mat)} \). The SC consensus pre-model distribution is identical to the hitting pre-weighted distribution, since \( MSY R_{(mat)} \) is an input and all trajectories, regardless of inputs, are forced to avoid extinction and become a candidate for reweighting and resampling at the final stage of Bayesian synthesis. Comparing boxes A, B, and D shows that the effect of including data about outputs is to make the distribution of \( MSY R_{(mat)} \) more precise for the projection method. In contrast, comparing boxes A, E, and G shows that the effect of the same data when hitting is used is to favor higher values of \( MSY R_{(mat)} \).

The results for 1993 RY are shown in Fig. 18. Box B shows that the assumptions implicit in the projection approach make virtually no constraint on permissible values of RY, whereas the assumptions implicit in the hitting approach effectively constrain RY to be between 0 and 400, given the available input evidence. Even when we compare boxes C and F, which show the use of exactly the same evidence for both methods, we see that the hitting method favors higher values of RY. The results for Q are similar.

Fig. 19 shows the results for 1978-1993 annual rate of total population increase, ROI.
The results for the projection method (boxes B and C) show that the assumptions about model inputs do not favor either increasing or decreasing population trends. In contrast, the constraints imposed by the hitting method (boxes E and F) favor increasing populations and are closer to the increase rates supported by the data. For both methods, the effect of the data on outputs is to further favor higher values of ROI.

The results for the 1993 depletion level of the mature component of the stock, relative to the pre-exploitation mature component, are shown in Fig. 20. No pre-model distribution was assumed for this quantity. This figure shows that the evidence about inputs is considered to be very uninformative by the projection method (box B), but it precisely determines depletion for the hitting method (box E). This is because $P_{1988}$ is an input for the hitting method. Indeed, comparison of boxes B, C, and F shows that $P_{1988}$ is the sole output which determines depletion for the hitting method; all other outputs have little influence, as does information about $P_0$. The fact that evidence about $P_0$ does not much influence estimated depletion for the hitting method can also be seen from Fig 16, where it is shown that the hitting method implies fairly strong constraints on $P_0$, compared to the pre-model evidence which might be used.

Figs 21 and 22 show the results for the recent percentages of calves and mature whales in the population. These figures are important because Butterworth and Punt (1995) have questioned whether data on the current population age structure have much influence on the results of a Bayesian synthesis analysis. These figures show that input and output evidence does provide information about the location and precision of age structure parameters. To the extent that this evidence is different than the pre-model distributions for these parameters, the choice of pre-model distributions can influence the results. The actual degree to which the age structure data influence the results from either method cannot be determined directly from this figure.

Butterworth and Punt (1995) also consider the influence of a pre-model distribution on MCM, the historic maximum ratio of calves to mature females. The results for MCM for each method are shown in Fig. 23. This figure shows that the two methods produce very similar distributions for MCM, despite their differing assumptions. This suggests that the influence of any pre-model distribution for MCM should be similar for both methods. Further, the evidence about inputs is not sufficient to eliminate unreasonable values of MCM for either method (boxes B, C, E, and F), therefore the MCM pre-model distribution adds information to the analysis.
Figure 16: Boxplots of pre-model distribution, pre-weighted and other intermediate distributions, and post-model distribution for $P_0$. 
Figure 17: Boxplots of pre-model distribution, pre-weighted and other intermediate distributions, and post-model distribution for $MSY R_{(mat)}$. 
Figure 18: Boxplots of pre-model distribution, pre-weighted and other intermediate distributions, and post-model distribution for 1993 RY. There was no pre-model distribution used for RY.
Figure 19: Boxplots of pre-model distribution, pre-weighted and other intermediate distributions, and post-model distribution for 1978–1993 annual rate of population increase, ROI.
Figure 20: Boxplots of pre-model distribution, pre-weighted and other intermediate distributions, and post-model distribution for 1993 depletion level of the mature population component. There was no pre-model distribution for mature depletion.
Figure 21: Boxplots of pre-model distribution, pre-weighted and other intermediate distributions, and post-model distribution for 1985-1992 average percent calves.
Figure 22: Boxplots of pre-model distribution, pre-weighted and other intermediate distributions, and post-model distribution for 1985-1992 average percent mature whales.
Figure 23: Boxplots of pre-model distribution, pre-weighted and other intermediate distributions, and post-model distribution for MCM, the historic maximum ratio of calves to mature females.
Figure 24: Sensitivity of $RY$ and $Q$ to $\alpha$ for the hitting method. Shown are post-model medians (solid line), 95\% probability intervals (dotted line), and 5\% percentiles (dashed line). The tick marks on the horizontal axes correspond to the different values of $\alpha$ we examined.

10 SENSITIVITY OF HITTING TO KILL RECORD BIAS

In Section 5 we examined how systematic bias in the historic kill record might affect the Bayesian synthesis assessment using the projection method. Here, we consider briefly the same question for the hitting method.

We considered scaling the historic kill record by $\alpha$ in exactly the same manner as in Section 5. However, with the hitting method, $P_0$ is an output, so no pre-model distribution is used for it.

The results are shown in Figs 24, 25, and 26 for the same variables examined in Section 5. A comparison of these figures with those from the earlier analysis in Section 5 shows that the results from hitting are very similar to those from projection. This indicates that both hitting and projection respond in the same way to potential kill record bias. Additionally, these figures show that the hitting method has the desirable property that it exhibits less Monte Carlo variability between runs.
Figure 25: Sensitivity of $P_0$, $MSYR_{(mat)}$, $ROI$, mature depletion, $MSY$, and $MSYL_{(mat)}$ to $\alpha$ for the hitting method. Shown are post-model medians (solid line) and 95% probability intervals (dotted line). The tick marks on the horizontal axes correspond to the different values of $\alpha$ we examined.
Figure 26: Sensitivity of $s$, $m$, $P_{1988}$, $MCM$, % calves, and % mature to $\alpha$ for the hitting method. Shown are post-model medians (solid line) and 95% probability intervals (dotted line). The tick marks on the horizontal axes correspond to the different values of $\alpha$ we examined.
11 CONCLUSIONS

The results of our investigations in Part II comparing the hitting and projections methods can be used to identify some important differences between methods. Section 8 shows that the hitting method produces higher estimates of \( R_Y \) and \( MSY R_{(mat)} \) compared to projection. The results of the two methods differ in other important ways, as well.

Section 9 shows that the hitting method constrains Bayesian synthesis results to a much more precise region, based on the input evidence, whereas the projection method considers the input evidence to be much less informative about reasonable trajectories. There are positive and negative aspects to this behavior. The extra constraints imposed by the hitting method mean that many more trajectories simulated on the basis of inputs produce reasonable output trajectories, therefore the hitting method uses Monte Carlo samples more efficiently. Indeed, although we ran the hitting method with only half the initial sample size of our projection method runs, the hitting method was able to produce a post-model sample with many more unique points than the projection method. However, since the hitting method requires nonlinear optimization for each trajectory, the total computing time for the hitting analyses were substantially longer than for projection, despite the difference in the number of trajectories run. The constraints imposed by the hitting method can also be seen in a negative light, compared to projection. Section 9 shows that the projection method is more strongly influenced by the evidence on outputs than is the hitting method. The data on outputs comes mostly from ice-based censuses and aerial photogrammetry, which we generally consider to be more reliable and precise than the evidence used for inputs. Therefore, we feel it is desirable for these output data to play a dominant role.

We also consider the role of the historic catch data for both methods. Section 10 shows that both hitting and projection Bayesian synthesis analyses are affected by systematic kill record bias in the same way.

ACKNOWLEDGEMENTS

This research was supported by the North Slope Borough (Alaska), the State of Alaska (through the Alaska Department of Community and Regional Affairs), and the National Oceanic and Atmospheric Administration (through the National Marine Mammal Laboratory to the Alaska Eskimo Whaling Commission). The authors thank Adrian Raftery and Greg Donovan for sharing the idea of alternative bounds for \( P_b \), and Judith Zeh and Adrian
Raftery for helpful comments regarding analyses preliminary to those presented here.

References


