Assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales using Bayesian model averaging

J.R. Brandon* and P.R. Wade#

Contact e-mail: jbrandon@u.washington.edu

ABSTRACT

Bayesian estimation methods are used to fit an age- and sex-structured population model to available data on abundance and stage-proportions (i.e. calves/mature animals in the population) for the Bering-Chukchi-Beaufort Seas stock of bowhead whales (*Balaena mysticetus*). The analyses consider three alternative population modelling approaches: (1) modelling the entire population trajectory from 1848, using the 'backwards' method where the trajectory is back-calculated based on assigning a prior distribution to recent abundance; (2) modelling only the recent population trajectory, using the 'forwards from recent abundance' method, where the population is projected forwards from a recent year and the abundance in that year is not assumed to be at carrying capacity; and (3) a version of (2) that ignores density-dependence. The 'backwards' method leads to more precise estimates of depletion level. In contrast, the 'forwards from recent abundance' method provides an alternative way of calculating catch-related quantities without having to assume that the catch record is known exactly from 1848 to the present, or having to assume that carrying capacity has not changed since 1848. Not only are all three models able to fit the abundance data well, but each is also able to remain consistent with available estimates of adult survival and age of sexual maturity. Sensitivity to the stage-proportion data and the prior distributions for the life history parameters indicates that use of the 1985 stage-proportion data has the greatest effect on the results, and that those data are less consistent with data on trends in abundance and age of sexual maturity. The analyses indicate that the population has approximately doubled in size since 1978, and the 'backwards' analyses suggest that the population may be approaching carrying capacity, although there is no obvious sign in the data that the population growth rate has slowed. Bayes factors are calculated to compare model fits to the data. However, there is no evidence for selecting one model over another, and furthermore, the models considered in this study result in different posterior distributions for quantities of interest to management. Posterior model probabilities are therefore calculated and used as weights to construct Bayesian model-averaged posterior distributions for outputs shared among models to take this ambiguity into account. This study represents the first attempt to explicitly quantify model uncertainty when conducting a stock assessment of bowhead whales.

KEYWORDS: ARCTIC; BOWHEAD WHALE; MODELLING; WHALING – ABORIGINAL; NORTHERN HEMISPHERE; BERING SEA; BEAUFORT SEA; CHUKCHI SEA

INTRODUCTION

Bowhead whales (*Balaena mysticetus*) have been caught by Alaskan Eskimos in an aboriginal fishery for hundreds of years or more and active whaling started along the arctic coast around 1,800-1,700 years before present (Braham, 1989; Dinesman and Savinetsky, 2003). A commercial fishery began takes from the Bering-Chukchi-Beaufort Seas (B-C-B) stock of bowhead whales in 1848, and the population is thought to have been severely depleted by the substantial commercial catches that occurred during the 1800s (Bockstoce and Botkin, 1983; Breiwick and Braham, 1990; Breiwick *et al.*, 1984). Although the commercial fishery almost completely collapsed early in the 1900s, aboriginal catches of B-C-B bowhead whales continue (e.g. Braham, 1995; George *et al.*, 1988; Suydam and George, 2004).

Most of the B-C-B bowhead whale stock migrates seasonally along the north slope of Alaska between wintering areas in the Bering Sea and summer feeding areas in the Beaufort Sea. Surveys have been conducted during the spring migration past Point Barrow during eleven years since 1978, to estimate the abundance and trends of this population (George *et al.*, 2004). The ice-based counts have resulted in abundance estimates substantiated by estimates using mark-recaptures of individually identifiable whales in aerial photographs (da Silva *et al.*, 2000). These abundance estimates have served as the primary basis for assessments

of the status of and management advice for the B-C-B bowhead whale stock by the Scientific Committee of the International Whaling Commission (IWC).

Given a history of catches, it was possible to conduct an initial assessment of the status of the population once the first abundance estimate was made in 1978. This assessment suggested that the population had experienced some recovery since its depletion in the 1800s, but that it was still below its carrying capacity (Breiwick *et al.*, 1984). Breiwick and Braham (1990) noted that estimates of carrying capacity and depletion level are sensitive to the current estimated population size. However, it is now possible to make more precise estimates of these quantities (conditional on the accuracy of the catch data and population dynamics model) because the number and precision of the abundance estimates has increased substantially.

Givens *et al.* (1993) and Raftery *et al.* (1995) introduced Bayesian methods (e.g. Press, 1989) to the assessment of bowhead whales. These methods can make use of multiple sources of data and fully characterise uncertainty. Bayesian assessments of the B-C-B bowhead whale stock (e.g. Givens *et al.*, 1995) using the *BALEEN II* model (de la Mare, 1989) have consequently been used extensively by the IWC Scientific Committee (Punt, 1999b). Although there was general agreement that using multiple sources of data in assessments was desirable, considerable debate ensued in the IWC Scientific Committee regarding the details of the statistical methods to be used in bowhead whale

* *University of Washington, School of Aquatic and Fisheries Sciences, Box 355020, Seattle, WA 98195, USA.*
# *National Marine Mammal Laboratory, Alaska Fisheries Science Center, NOAA Fisheries Service, 7600 Sand Point Way NE, Seattle, WA 98115, USA.*

assessments (e.g. Butterworth and Punt, 1995; Poole and Givens, 2001; Poole *et al.*, 1999; Punt and Butterworth, 1999; 2000; Schweder and Ianelli, 2000). One troublesome outcome of the results in the standard 'forwards from *K*' assessment (i.e. projecting the population dynamics model forwards from a prior distribution for the population size in 1848, which is assumed to be carrying capacity) was that the observed rate of increase of the population (ROI) differed substantially from the resulting posterior distribution for the ROI, even though the data on ROI were the main data source used in the analysis (e.g. Punt and Butterworth, 1999; Raftery and Poole, 1997). This and other related issues were determined to be due in part to a methodological issue associated with the forwards method and the specification of prior distributions. It was also determined that these issues could essentially be resolved by the use of the 'backwards' method, in which no prior is specified for the population size in 1848; instead, a prior distribution is specified for abundance in a recent year, and the population level in 1848 (assumed to be carrying capacity) is then back-calculated from that recent abundance (Butterworth and Punt, 1995; Poole and Raftery, 1998; Punt and Butterworth, 1999).

An important assumption made in bowhead whale assessments is that the catch history is known without error. It has been found that the catch record going back to 1842 for the eastern North Pacific stock of gray whales cannot be reconciled with the population's current dynamics (Punt and Butterworth, 2002), without assumptions that the catch record has been substantially under-estimated and/or that carrying capacity has changed since the mid 1800s. While it appears from past assessments that the bowhead whale catch record can be reconciled with the current dynamics, it is still appropriate to investigate an alternative assessment that does not make use of the historical catch record. This can be achieved using the method developed by Wade (2002a) for gray whales, where a model is projected forwards from a prior distribution for abundance in a recent year, with a separate prior distribution specified for carrying capacity. Such an assessment does not use the historic catch record prior to 1978, and does not need to assume that carrying capacity has remained constant since the mid 1800s; therefore, it should be robust to problems with these assumptions.

The 1998 assessment of the B-C-B bowhead whale stock (IWC, 1999; Punt, 1999a) used 'backwards' Bayesian estimation based on the *BALEEN II* model, and the then-available abundance estimates and the data on the proportion of calves and mature animals in the population during 1985-94 (the 'stage-proportion' data). New information available for this stock since the 1998 assessment includes: (1) a mark-recapture estimate of adult survival from aerial photographs (Zeh *et al.*, 2002); (2) an estimated age at sexual maturity and an estimate of maximum age (George *et al.*, 1999); (3) a recalculation of the stage-proportion data (Koski *et al.*, 2004); and (4) an estimate of abundance for 2001 (George *et al.*, 2004).

This paper examines the sensitivity of the results of the B-C-B bowhead whale assessment to modelling: (1) the entire population trajectory from 1848 (using the 'backwards' method); and (2) only the recent period (where the population is projected forwards from a recent year, and the abundance in that year is not assumed to be at carrying capacity). A third set of analyses is conducted using a Leslie matrix with no density-dependence. The population model used in this assessment is the density-dependent Leslie matrix model (Leslie, 1945; 1948) developed by Breiwick *et al.* (1984) except that reproduction and natural mortality are

assumed to occur before removal by catches. When parameterised in equivalent ways, the results of the *BALEEN II* model and this Leslie matrix model are almost identical (Punt and Butterworth, 2002; Wade, 2002a).

This paper also examines the sensitivity of the model outputs to the sources of data included in the assessment and the specifications for the prior distributions. Scenarios are specified to investigate the sensitivity of the results to data-based informative vs. uniform (less informative) prior distributions, as well as to how the stage-proportion data are constructed. The stage-proportion data are excluded altogether in some analyses, and varied in others based on whether the aerial photographic survey data from 1985[1] are included in their calculation or not. Within a scenario, identical prior distributions are assumed for the life-history parameters and matching data-sets are used in the likelihood function for each of the three modelling approaches to enable comparisons to be made among them.

Model uncertainty is accounted for within a Bayesian framework, and Bayes factors (Kass and Raftery, 1995) are calculated to quantify the evidence provided by the data in favour of the different modelling approaches. For the 'reference scenarios' there is no evidence for selecting one model over another, but there are important differences in the estimates of quantities that would be of interest to management. Therefore, we follow the philosophy outlined by a number of authors (e.g. Buckland *et al.*, 1997; Durban *et al.*, 2005; Hoeting *et al.*, 1999; Kass and Raftery, 1995; Patterson, 1999; Raftery *et al.*, 2005) and present quantities of interest as Bayesian model averages; weighting the output of contending models based upon their relative likelihoods, as opposed to selecting between them.

METHODS

Available data
The population dynamics models were fitted to three sources of data: (1) abundance estimates from ice-based surveys at Point Barrow, Alaska between 1978 and 2001 (Table 1); (2) proportion calves/mature animals in the population from 1985 to 1994 (Table 2); and (3) annual catches in individuals from 1848 to 2002 (Table 3).

All of these sources of data were used in the 1998 assessment but have been updated since. The first ice-based survey since 1993 was conducted in 2001 and has provided an abundance estimate for that year of 10,545 with a coefficient of variation of 0.128 (updated from George *et al.*, 2004 by Zeh and Punt, 2005). The catch data have been updated with the post-1998 catches and revisions to the catches for 1994-96. Two additional years of aerial photographic data have been analysed since the previous assessment and the stage-proportion data have been recalculated (Angliss *et al.*, 1995; Koski *et al.*, 2004).

Additional information available for this stock since the 1998 assessment includes a mark-recapture estimate of survival (Zeh *et al.*, 2002) and a recent estimate of age of sexual maturity and the maximum age (George *et al.*, 1999). This information was included in the analyses of this paper as data-based informative priors (see the section on Model parameters and prior distributions).

---

[1] 1985 appears to have been an anomalously late migration year; the aerial survey in 1985 is believed to have ended too early to have sampled the tail-end of the migration, typically when most cow/calf pairs are in the study area (Koski *et al.*, 2004).

Table 1

Estimates, coefficients of variance (CVs) (actually the standard errors of the log abundance estimates, which are approximately equal) and the correlation matrix for the indices of abundance for the B-C-B Seas stock of bowhead whales (Zeh and Punt, 2005).

Year	Estimate	CV	Correlation matrix										
1978	4,765	0.305	1.000										
1980	3,885	0.343	0.118	1.000									
1981	4,467	0.273	0.056	0.050	1.000								
1982	7,395	0.281	0.094	0.084	0.035	1.000							
1983	6,573	0.345	0.117	0.104	0.049	0.084	1.000						
1985	5,762	0.253	0.070	0.062	0.020	0.078	0.062	1.000					
1986	8,917	0.215	0.072	0.064	0.017	0.092	0.064	0.113	1.000				
1987	5,298	0.327	0.124	0.110	0.052	0.088	0.110	0.065	0.067	1.000			
1988	6,928	0.120	0.028	0.025	0.013	0.017	0.024	0.009	0.007	0.026	1.000		
1993	8,167	0.071	0.001	0.001	0.001	0.000	0.001	-0.001	-0.002	0.001	0.000	1.000	
2001	10,545	0.128	0.008	0.007	0.005	0.001	0.007	-0.004	-0.008	0.008	0.003	0.000	1.000

Table 2

The proportion of observed calves (p_c^{obs}) and mature (p_m^{obs}) animals with associated standard errors, over the years 1985-1994. Proportions are given based on including and ignoring the anomalous 1985 data set, as well as those used in the previous assessment (IWC, 1999; Koski *et al.*, 2004).

Scenario	p_c^{obs}	σ_{P_c}	p_m^{obs}	σ_{p_m}
Exclude 1985	0.0580	0.0062	0.4366	0.0106
Include 1985	0.0309	0.0034	0.4160	0.0096
1998 assessment	0.052	0.0164	0.411	0.0286

Table 3

Catches of B-C-B Seas bowhead whales, 1848-2002. Values in parenthesis are the catches used by Punt and Butterworth (1999) in the 1998 assessment where these catches differ from those used in the present analyses.

Year	Total kill	Year	Total kill	Year	Total kill	Year	Total kill
1848	18	1887	240	1926	35	1965	14
1849	573	1888	160	1927	14	1966	24
1850	2,067	1889	127	1928	30	1967	12
1851	898	1890	136	1929	30	1968	27
1852	2,709	1891	284	1930	17	1969	32
1853	807	1892	346	1931	32	1970	48
1854	166	1893	180	1932	27	1971	25
1855	2	1894	234	1933	21	1972	44
1856	0	1895	117	1934	21	1973	51
1857	78	1896	118	1935	15	1974	42
1858	461	1897	130	1936	24	1975	32
1859	372	1898	309	1937	53	1976	74
1860	221	1899	234	1938	36	1977	72
1861	306	1900	148	1939	18	1978	15
1862	157	1901	55	1940	20	1979	20
1863	303	1902	162	1941	38	1980	32
1864	434	1903	116	1942	26	1981	26
1865	590	1904	86	1943	14	1982	14
1866	554	1905	105	1944	8	1983	16
1867	599	1906	69	1945	23	1984	16
1868	516	1907	96	1946	20	1985	14
1869	382	1908	123	1947	21	1986	22
1870	637	1909	61	1948	8	1987	29
1871	138	1910	37	1949	11	1988	28
1872	200	1911	48	1950	23	1989	25
1873	147	1912	39	1951	23	1990	41
1874	95	1913	23	1952	11	1991	47
1875	200	1914	61	1953	41	1992	46
1876	76	1915	23	1954	9	1993	51
1877	270	1916	23	1955	36	1994	39 (38)
1878	80	1917	35	1956	11	1995	56 (57)
1879	266	1918	27	1957	5	1996	42 (45)
1880	480	1919	33	1958	5	1997	62
1881	435	1920	33	1959	2	1998	51
1882	242	1921	9	1960	33	1999	47
1883	42	1922	39	1961	17	2000	42
1884	160	1923	12	1962	20	2001	67
1885	377	1924	41	1963	15	2002	44
1886	168	1925	53	1964	24		

Population dynamics model

The underlying population model is a simplified age- and sex-structured Leslie matrix (Leslie, 1945; 1948) projected as:

$$\underline{n}_{t+1} = \left(\mathbf{A}_t \underline{n}_t \right) - \underline{h}_t \qquad (1)$$

where:

$\underline{n}_t$ is vector of population size in each age class at the start of year t (defined when calving and natural mortality occur);

$\mathbf{A}_t$ is the Leslie matrix for year t;

$\underline{h}_t$ is the vector of age-specific catches during year t.

The catches and birth rates are assumed to be equal for both males and females (i.e. the vectors above are divided equally by sex). The parameters that define the entries of the Leslie matrix are: (1) S_{juv}, the survival rate of immature whales (assumed identical for calves and juveniles); (2) a_T, the last age with survival rate S_{juv}; (3) S_a, the survival rate of mature whales; (4) a_m, the age at sexual maturity (the last age class with zero fecundity); (5) f_{max}, the maximum fecundity rate; and (6) a_{max}, the maximum age, after which survival becomes zero. Fecundity is assumed identical for all mature animals, and is calculated as the number of female calves per mature female. Recruitment to the fishery is assumed to be knife-edged and to occur at age 1, and the catch is distributed uniformly over all recruited age-classes (i.e. uniform selectivity across recruited age-classes).

The projections are initialised from a stable age distribution for the population in the year prior to that with the first catch (e.g. 1977 or 1847) based on the values for the parameters sampled from the prior. This population vector is then projected one year forward without catch, and the population vector re-scaled so that the 1+ population size in the year with the first catch equals that generated from the prior for 1+ abundance for that year.

Density dependence is assumed to affect fecundity according to:

$$f_t = f_0 + \left(f_{max} - f_0 \right) \left[1 - \left(\frac{N_t^{1+}}{K^{1+}} \right)^z \right] \qquad (2)$$

where:

f_t is the fecundity during year t;

f_{max} is the maximum fecundity (in the limit of zero population size);

N_t^{1+} is the (1+) population size at the start of year t;

K^{1+} is the pre-exploitation (1+) population size;

z is the shape parameter;

f_0 is the fecundity at carrying capacity.

Given values of life-history parameters in the model, the value for f_0 is determined from the characteristic equation of the Leslie matrix given equilibrium conditions:

$$f_0 = \frac{1 - S_a}{S_{juv}^{(a_T)} S_a^{(a_m - a_T)} \left[1 - S_a^{(a_{max} - a_m - 1)} \right]} \qquad (3)$$

Model parameters and prior distributions
Table 4 lists the parameters and their priors used in three 'reference' scenarios. The population trajectory is modelled in three ways: (1) a density-dependent model initialised in 1848 (abbreviated: '1848 DD'); (2) a density-dependent model initialised in 1978 ('1978 DD'); and (3) a density-independent model initialised in 1978 ('1978 NON DD'). The six life-history parameters of the Leslie model are included in each of three models, but the remaining parameters differ among models. The '1848 DD' model includes a parameter for the population size in 1993, N_{1993}^{1+}, and one for the maximum sustainable yield level, $MSYL^{1+}$. The '1978 DD' model also includes the parameter $MSYL^{1+}$ but instead of placing a prior on N_{1993}^{1+} places one on N_{1978}^{1+}. This model also includes an additional (explicit) prior on the carrying capacity, K^{1+}. The '1978 NON DD' model includes priors on N_{1978}^{1+} and the maximum population growth rate in the absence of density dependence, r. For this model, fecundity and population growth rates apply only to the specified period, and where the distinction is appropriate, they are referred to as f and r. However, when methods are consistent across models, these rates are referred to as f_{max} and r_{max} for the sake of simplicity.

'Data-based' prior distributions are assigned to adult survival rate and the age at sexual maturity and the maximum age of the Leslie matrix is determined from the results of recent research on ageing. The informative prior for S_a (Table 4) approximates the Bayesian posterior calculated for this parameter based on a mark-recapture analysis of photo-identification data (Zeh *et al.*, 2002). Information on age-at-maturity is taken from a study by George *et al.* (1999) that estimated ages of caught animals based on the chemistry of eye lenses. Those authors fitted a growth curve to these ages from known lengths and combined this relationship with previous data on length at sexual maturity to provide an estimate of the age-at-maturity. The ageing results also estimated some animals were older than the previously accepted maximum age, and this result is supported by the recent recovery of traditional whaling tools in five whales (reported in the same study). The maximum age in the Leslie matrix is therefore set to 200 to reflect this information.

Punt and Butterworth (1999) noted that due to the functional relationships among the life-history parameters in an age-structured population model, placing a prior on S_{juv}

would be an instance of Borel's Paradox (i.e. effectively placing two priors on the same parameter); instead the value of S_{juv} is solved for analytically in this study by rearranging the characteristic equation of the Leslie matrix given the values for the remaining five parameters and λ, the dominant eigenvalue of the Leslie matrix (i.e. $r_{max} + 1$) (Breiwick *et al.*, 1984):

$$S_{juv} = \left[\frac{\lambda^{(a_m + 1)} - S_a \lambda^{(a_m)}}{S_a^{(a_m - a_T)} f_{max} \left\{ 1 - \left(S_a / \lambda \right)^{(a_{max} - a_m - 1)} \right\}} \right]^{\frac{1}{a_T}} \qquad (4)$$

The value for S_{juv} is forced to be less than that of S_a. If necessary, values for f_{max} and S_a are re-sampled (see below), until this condition is met, or 1,000 re-samples occur. If this maximum is reached, a new value for r_{max} is re-sampled, and the process repeated until an acceptable sample from the prior occurs.

The priors for the remaining life history parameters are also based on available information. The prior distribution for f_{max} is based on an assumed range of a 2.5- to 4-year calving interval for large baleen whales (IWC, 1998). Note that f_{max} is specified in the standard Leslie matrix formulation as female calves per female per year (i.e. a fecundity rate of 0.125 implies a female calving interval of 8 years, and therefore a total calving interval of 4 years, assuming an equal sex ratio of calves). The age of transition from immature to adult survival is assigned a discrete uniform prior over the interval 1 to 9 years.

Output quantities
Posterior distributions are calculated for several output quantities that are functions of the parameters in Table 4. The maximum sustainable yield rate ($MSYR^{1+}$) is calculated as $\lambda - 1$ based on the f_t value associated with $MSYL^{1+}$. Maximum sustainable yield (MSY^{1+}) is calculated as the product, $MSYR^{1+} \cdot MSYL^{1+} \cdot K^{1+}$. Current replacement yield (RY^{1+}) is calculated as the difference between the number of 1+ animals in 2002 (prior to the removal of catches in that year) and the number of such animals at the end of 2001. The quantity Q_1^{1+}, designed to meet the intent of aboriginal whaling management objectives (Wade and Givens, 1997), is also calculated. This quantity has the property that the proportion of net production allocated to recovery increases at higher levels of stock depletion[2]. Specifically:

$$Q_1^{1+} = \min \left(MSYR^{1+} * N_{2002}^{1+}, 0.9 MSY^{1+} \right) \qquad (5)$$

The post-model-pre-data distribution is reported for the parameters. This distribution arises after conditioning the specified priors on the model (i.e. by eliminating combinations of parameters for which the juvenile survival rate implied by equation (4) exceeds the adult survival rate drawn from the joint prior distribution). Likewise, post-model-pre-data distributions for output quantities are calculated as the distributions for these quantities in the sampled joint prior space.

Parameter estimation
The Sampling-Importance-Resample (SIR) algorithm (Rubin, 1988; Smith and Gelfand, 1992; Wade, 2002a) is used to generate samples of parameter vectors (and output quantities of interest) from the posterior distribution. This

---

[2] This definition applies to a population above some minimum level, P_{min} (assumed here to be $0.1K^{1+}$), below which catches are set to zero.

Table 4

Prior distributions used for the reference scenarios for each model type. Dashes (-) represent prior distributions that are equal to those from the model in the column to the left (e.g. the 1978 NON DD model had the same prior on the size of the 1+ population in 1978 as the 1978 DD model). 'N/A' represents a prior that was not applicable to a certain model or models. Fecundity is defined as female calves per mature female. For the 1978 NON DD model r_{max} and f_{max} are not maxima, but constant values. Results from the reference scenario were used in all figures and when performing the Bayesian model averaging. The reference scenarios are based on the stage-proportion data set that ignores the data for 1985.

		Model type		
Parameter		1848 DD	1978 DD	1978 NON DD
S_a	Adult survival	N(0.990, 0.02), truncated at 0.940 and 0.995[a]	-	-
f_{max}	Maximum fecundity	U[0.125, 0.200][b]	-	-
a_T	Age-at-transition to adult survival	U[1, 9][c]	-	-
a_m	Age-at-sexual maturity	N(20.0, 3.0) truncated at 13.0 and 26.0[d]	-	-
$r_{max} = \lambda - 1$	Intrinsic population growth rate	U[0.005, 0.075][e]	-	-
N^{1+}_{1978}	1+ population size in 1978	N/A	U[3,000, 9,000][f]	-
N^{1+}_{1993}	1+ population size in 1993	N(7,800, 1,200)[g]	N/A	N/A
K^{1+}	1+ carrying capacity	N/A	U[8,000, 30,000][h]	N/A
$MSYL^{1+}$	MSYL in terms of the 1+ component	U[0.40, 0.80][i]	-	-

[a]Based on the posterior distribution for adult survival rate obtained by Zeh *et al.* (2002).
[b]The prior for the maximum number of calves (of both sexes) per mature female selected by the Scientific Committee of the International Whaling Commission was U[0.25, 0.4] (IWC, 1995). This is the corresponding prior given fecundity has been defined here as female calves per mature female per year.
[c]Selected by the Scientific Committee of the International Whaling Commission (IWC, 1995) although there is little information on the value of this parameter (Givens *et al.*, 1995).
[d]Based on a best estimate of 20 years and a lower confidence for the age-at-maturity of 14 years (IWC, 1995).
[e]Preliminary trials indicated there was no posterior probability outside this range, which was confirmed in the final analyses. This range was therefore selected to improve the efficiency of the numerical integration while not affecting the results.
[f]Selected to encompass a plausible range of values for 1+ population size in 1978.
[g]Selected by the Scientific Committee of the International Whaling Commission (IWC, 1995) based on the prior distribution assumed for the Bayes empirical Bayes estimate of abundance (Raftery and Zeh, 1991).
[h]Based on the range selected by the Scientific Committee of the International Whaling Commission (IWC, 1995).
[i]Selected to encompass the range of values commonly assumed when conducting assessments of cetacean populations.

algorithm involves randomly sampling a large number of parameter vectors $\underline{\theta}_i$ (draws) from the prior distribution. A population trajectory is then calculated for each vector of parameter values, and this trajectory is used to determine the likelihood of the data for each random draw. 10,000 draws (which form the numerical representation of the posterior distribution) are then selected by sampling (with replacement) from the initial samples from the prior, with probability proportional to the likelihood. Following Punt and Butterworth (1999) and Raftery *et al.* (1995), the SIR algorithm is considered to have converged if the number of unique parameter vectors in the sample from the posterior is fairly high (>5,000) and if the most frequently re-sampled parameter vector did not occur in the posterior sample more than ten times.

The total negative log-likelihood of a model trajectory, given a vector of parameters and the data, consists of contributions from four data sources: (1) the estimate of abundance for 1993; (2) the estimates of abundance for the remaining years; (3) the proportion of calves in the population; and (4) the proportion of mature animals in the population. The abundance estimates are assumed to be indices of the 1+ component of the population. The scientific surveys at Point Barrow are assumed to have occurred after the aboriginal catch, and the likelihood function is calculated accordingly (i.e. catches are removed before calculating the likelihood of the data for a given year). Model-predicted proportions are calculated over the period 1985 to 1994, as the actual stage proportions are based on data for these years.

The estimate of abundance for 1993 is assumed to be independent of the remaining estimates (Punt and Butterworth, 1999) and to have normally as opposed to log-normally distributed sampling error. The contribution of the abundance estimates to the negative of the log-likelihood function (ignoring constants independent of model parameters):

$$L_1 = 0.5 \frac{\left(\hat{N}_{1993} - 8293\right)^2}{626^2}$$

$$L_2 = 0.5 \sum_{t_1} \sum_{t_2} \left(\ell n \hat{N}_{t_1} - \ell n N_{t_1}^{obs}\right)^T \sum_{t_1, t_2}^{-1} \left(\ell n \hat{N}_{t_2} - \ell n N_{t_2}^{obs}\right)$$

(6)

where:

N_t^{obs} is the estimate of abundance for year t;
$\hat{N}_t$ is the model estimate of 1+ abundance for year t;
Σ is the variance-covariance matrix for the logarithms of the estimates of abundance (excluding 1993).

The estimates of abundance (Table 1) are based on combining the data from visual counts at Point Barrow, Alaska, and estimates of the proportion of animals which passed within visual range based on acoustic data. Eqn. 6 accounts for the correlation among the non-1993 estimates of abundance that arises because the proportion within visual range is treated as a random effect when constructing the estimates of abundance (Zeh and Punt, 2005).

The contribution of the proportion data to the likelihood function follows Punt (2006), i.e. given the bootstrapping approach adopted to calculate the length-frequency distributions from which the proportion data were calculated (Koski *et al.*, 2004), it was reasonable to assume that the estimates are normally distributed (ignoring constants):

$$L_3 = \frac{1}{2\left(\sigma_{p_c}\right)^2}\left(p_c - p_c^{obs}\right)^2 \qquad L_4 = \frac{1}{2\left(\sigma_{p_m}\right)^2}\left(p_m - p_m^{obs}\right)^2$$

(7)

where:

p_c^{obs} is the observed fraction of the population that consisted of calves between 1985 and 1994;
σ_{p_c} is the standard deviation of p_c^{obs};

p_c is the model-estimate of the fraction of the population that consisted of calves between 1985 and 1994;

p_m^{obs} is the observed fraction of the population that consisted of mature animals between 1985 and 1994;

σ_{p_m} is the standard deviation of p_m^{obs};

p_m is the model-estimate of the fraction of the population that consisted of mature animals between 1985 and 1994.

Model comparison

The three models considered in this paper are compared using Bayes factors (Kass and Raftery, 1995). The Bayes factor is calculated as the probability of observing the data given one hypothesis (model) divided by the probability of observing the same data given an alternative hypothesis, i.e.:

$$B_{12} = \frac{pr(D|H_1)}{pr(D|H_2)} \qquad (8)$$

In the context of model comparison, the hypotheses represent competing models and the Bayes factor is used as the evidence provided by the data in favour of one model over another. Although Eqn. (8) has the form of a likelihood ratio, if there are unknown parameters in either of the competing models, the probability densities must be found by integrating, as opposed to maximising, over the parameter space. Therefore, for a given model, the probability of the data is:

$$pr(D|H) = \int pr(D|H,\theta) pr(\theta|H) d\theta \qquad (9)$$

This integration is based on the sample from the prior using the equation:

$$\hat{pr}(D|H) = \frac{1}{n_1} \sum_{i=1}^{n_1} pr(D|\underline{\theta}_i, H) \qquad (10)$$

where $\underline{\theta}_i$ is the ith (of n_1) samples from the prior distribution.

Model averaging

Model uncertainty is accounted for by calculating the posterior probability of each model conditioned on the data and the priors, and then combining results across models as a weighted average of the posterior densities for a quantity of interest (Kass and Raftery, 1995). Hoeting *et al.* (1999) provide a convenient method of calculating the posterior probability of model H_k (where, $k=1, \ldots, K$ models are being considered) based on Bayes' theorem:

$$pr(H_k|D) = \frac{pr(D|H_k) pr(H_k)}{\sum_{i=1}^{K} pr(D|H_i) pr(H_i)} \qquad (11)$$

where $pr(H_k)$ is the prior probability that H_k is the true model and $pr(D|H_k)$ is the estimate of the probability of the data (Eqn. (10)). All of the probabilities are conditional on the set of models being considered (Hoeting *et al.*, 1999). For f_{max}, r_{max} and quantities related to carrying capacity, only the two models incorporating density dependence could be used to derive model averages. Under the set of models considered, these posterior model probabilities were used to determine model-averaged posterior probability distributions for the model outputs, θ:

$$pr(\theta|D) = \sum_{i=k}^{K} pr(\theta|H_k, D) pr(H_k|D) \qquad (12)$$

In the context of the SIR algorithm used here, Bayesian model averaging was accomplished by selecting a number of random draws from the posterior for each model and combining them to form a model-averaged posterior. This number was determined by the posterior probability for each model. All models were considered equally probable *a priori* (i.e. objective ignorance regarding the true model), so the posterior probability of a given model is determined using the values from Eqn. (10) normalised to sum to one over models.

RESULTS AND DISCUSSION

Fits to the data

Fig. 1 shows the fits to the abundance estimates for three reference scenarios and the Bayesian model average. All three models provide a relatively good fit to these data, although the abundance estimate for 2001 falls well above the upper 90% credibility limit from the '1848 DD' model. This occurs because the median of the posterior distribution for 1+ population size in 2002 for this model (9,496) is smaller than the posterior medians for the other two models (Fig. 2, Table 5).

The abundance estimates indicate the population has been increasing steadily over 1978-2001, and the data on adult survival, age of sexual maturity, and the stage-proportion data (excluding 1985) are all relatively consistent with this increase (i.e. the inclusion of those data in the analysis does not lead to the model being unable to mimic the abundance data).

Backwards to 1848 (density dependent model): 1848 DD

The upper left panel of Fig. 1 shows the posterior median time-trajectory of 1+ population size along with its 90% credibility interval for the reference scenario. The population size is estimated to have declined dramatically during the 1800s, being reduced to approximately half of its pre-exploitation level within five years of the start of the commercial fishery, and 10% of this size by the early 1900s. However, the population recovered steadily thereafter. The 90% credibility interval for the post-model-pre-data distribution for K^{1+} is [8,000-30,000], with lower values favoured (Fig. 2). In contrast, the 90% credibility interval for the posterior distribution of K^{1+} is [9,000, 14,000] (Table 5) indicating that the data update the prior distribution substantially. The 2002 population size is estimated to be above 50% of K^{1+}, and there is a high probability of it being above $MSYL^{1+}$ (Fig. 3, Table 5). The posterior distribution for replacement yield in 2002 has a mode around 200, with a lower 5th percentile of 61 (Fig. 4, Table 5). In contrast, the lower 5th percentile for Q_1^{1+} is 99 (Fig. 4, Table 5). These two quantities differ because the current population size is estimated to be larger than $MSYL^{1+}$, and approaching K^{1+}. Therefore, density dependence has slowed population growth and RY^{1+} has decreased. This is the same situation that led to the use of Q_1^{1+} as a more appropriate measure of sustainable catch (to achieve IWC management goals) for the eastern North Pacific stock of gray whales (Wade, 2002a).

The constraints imposed by the relationships among the life-history parameters constrained r_{max} to be less than about 0.07 (Fig. 3), although they also reduced the (prior) probability of values of r_{max} larger than 0.06. The posterior for r_{max} assigns most support to values larger than 0.03 (posterior median 0.041, Table 6). The posterior distributions for adult and immature survival favour higher
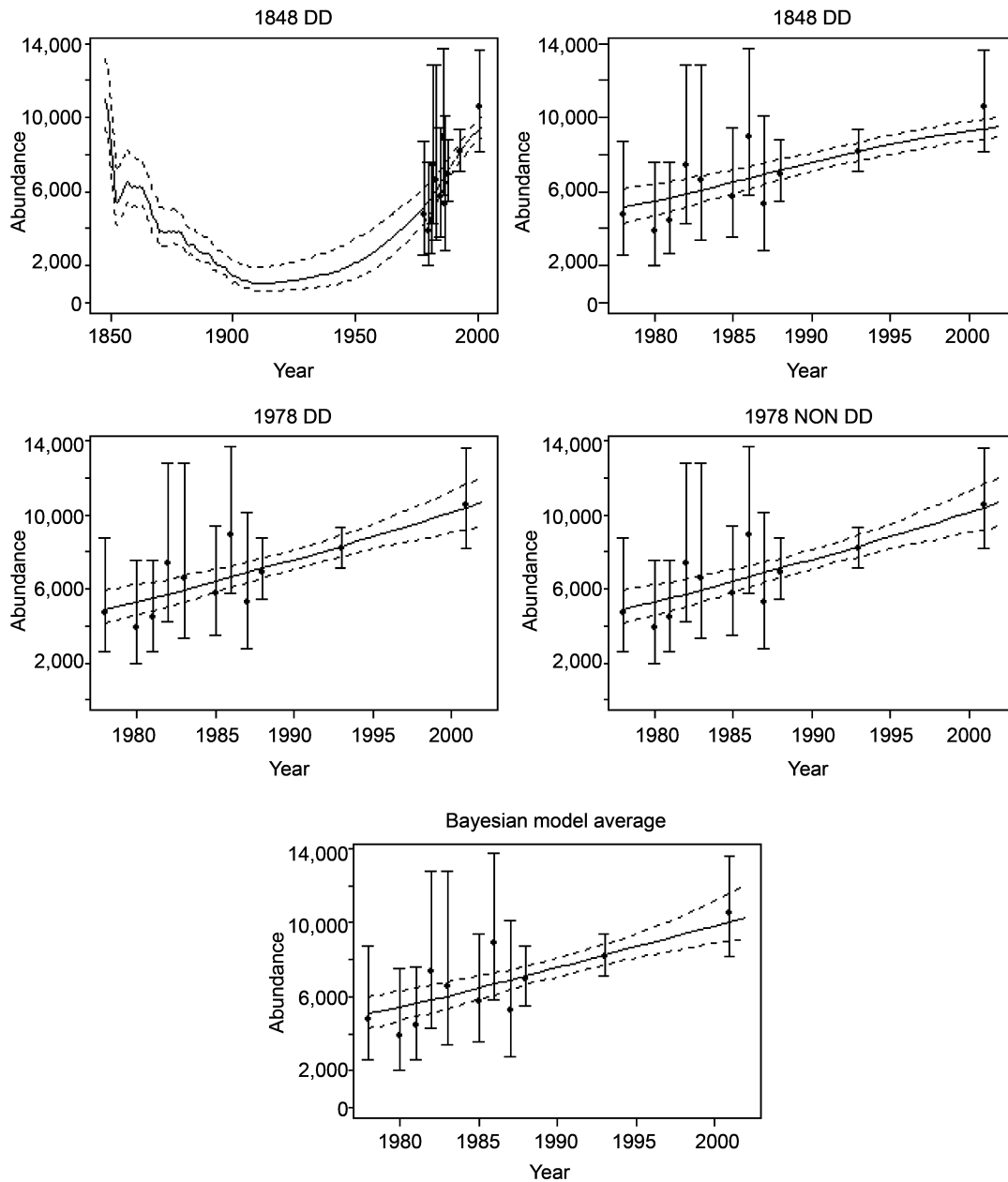
Fig. 1. Time trajectories (medians and 90% credibility intervals) for 1+ population size for the three reference scenarios and the Bayesian model average. The two uppermost plots are for the 1848 DD model, showing the entire trajectory from 1848, and only the recent trajectory from 1978 for comparison. Error bars represent 95% CIs, and are assumed to be log-normally distributed for all abundance estimates except 1993 (second to last), which is assumed to have a normally distributed error structure.

values than implied by the prior distributions for these parameters, and that for a_m values lower than implied by the prior distributions (Figs 5 and 6). The post-model-pre-data distribution for fecundity favours higher values (~0.20, Fig. 6), but the posterior median is 0.171, or a calving interval of approximately three years (Table 6). The results in Figs 3, 5 and 6 show that the data are clearly capable of updating the prior distributions for the life-history parameters.

The results for this model are not particularly sensitive to changing the prior distributions for S_a and a_m and to ignoring the proportion data (Table 5). The most noteworthy feature of these sensitivity tests are the changes to the catch-related outputs (RY^{1+}, Q_1^{1+} and MSY^{1+}). Results are not shown in Table 5 for the case in which the 1985 stage-proportion data are included in the analyses due to computational difficulties in achieving convergence. Preliminary analyses including these data indicated,

however, that they are inconsistent with what is known about bowhead whale life history and the time series of abundance estimates. This inconsistency was the cause of the inability to achieve convergence.

Forwards from 1978 (density dependent model): 1978 DD

The posterior for K^{1+} from this analysis is much more uncertain than that from the '1848 DD' analysis (Fig. 2). This is because this analysis ignores the information contained in the 1848-1977 catch record, and because the abundance estimates show no evidence for a reduction in trend (which would be expected as the population approaches carrying capacity) and, unlike the '1848 DD' model, the '1978 DD' model does not make the assumption that the population size in 1848 was K^{1+}. Therefore, although this model confirms that the population is
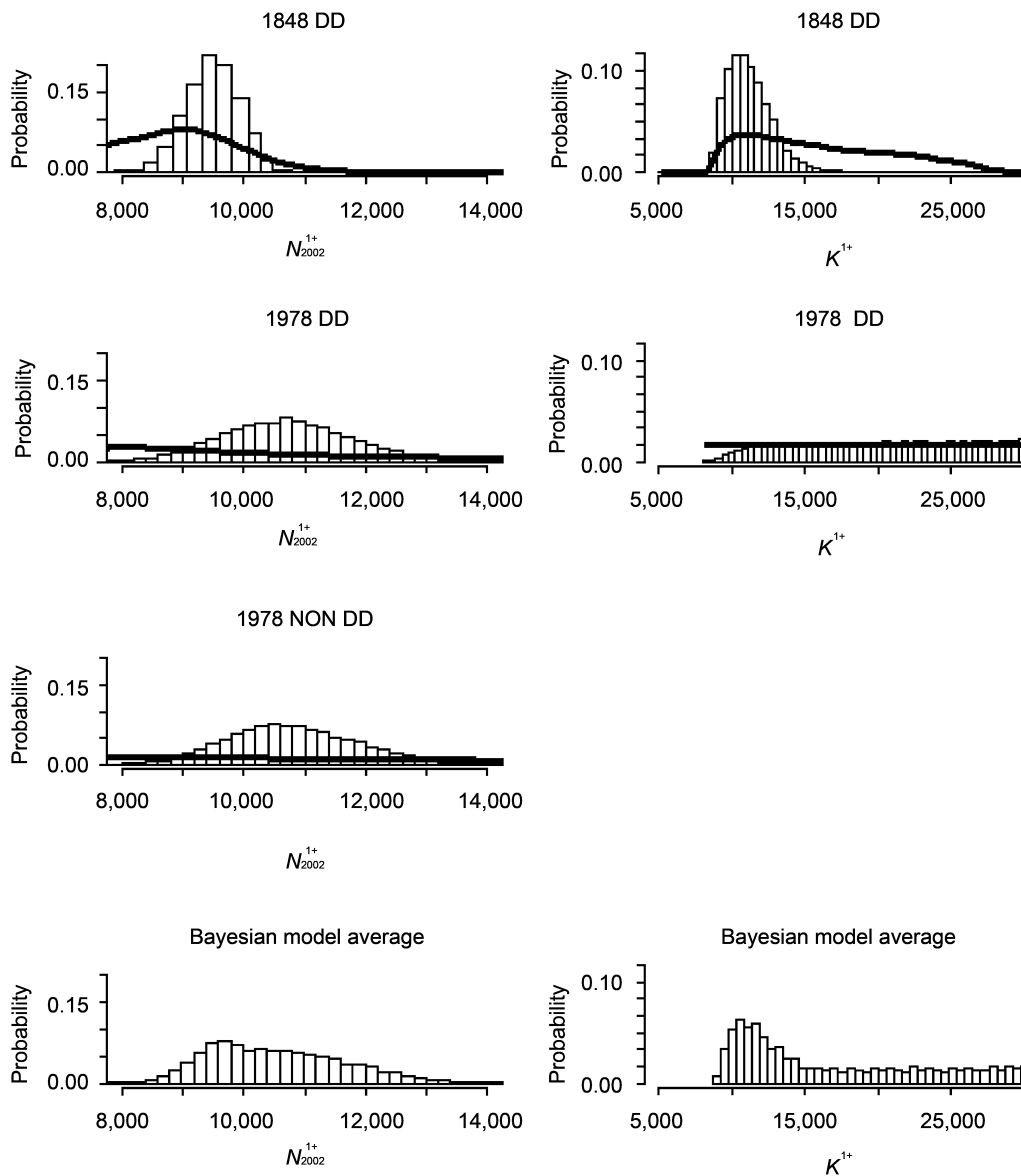
Fig. 2. Posterior (vertical bars) and post-model-pre-data (solid lines) distributions for 1+ population size in 2002, N_{2002}^{1+} (left panels) and 1+ carrying capacity, K^{1+} (right panels). Results are shown for only two of the three reference scenarios for K^{1+}, and the Bayesian model average for K^{1+} is based on the results of these two scenarios only.

increasing (Fig. 1), it infers that the population is currently at a much lower fraction of its (current) carrying capacity than the '1848 DD' model (Fig. 3; Table 5). The posterior for N_{2002}^{1+}/K^{1+} is strongly influenced by the prior distribution assumed for K^{1+} given the inability of the data to place an upper bound on K^{1+}. This is clearly evident from the results of the sensitivity test in which the upper limit of the prior for K^{1+} is increased from 30,000 to 100,000. The results for this sensitivity test imply an increase to the median of the posterior for K^{1+} of 165% and a reduction to the posterior median for N_{2002}^{1+}/K^{1+} of 62% (Table 5) as the upper bound for K^{1+} is increased by 233%.

Given that there is little independent information on which to base a prior distribution for K^{1+}, the choice of the prior for K^{1+} is essentially arbitrary, and it should be recognised that this 'forwards' analysis consequently does not provide robust estimates of quantities related to K^{1+} (such as N_{2002}^{1+}/K^{1+}). However, Punt and Butterworth (1999) noted that some key management-related quantities (e.g. RY^{1+} and Q_1^{1+}) are relatively insensitive to the prior assumed for K^{1+}, so this approach still has some value. The implicit (post-model-pre-data) distribution for RY^{1+} favours (is

skewed towards) values less than 200 (Fig. 4). Despite this, the posterior median is 324, with a lower 5th percentile of 147 (Table 5). A similar result is evident for Q_1^{1+}, with low values favoured by the post-model-pre-data distribution, but higher values supported by the data (median=295, lower 5th percentile=160) (Fig. 4, Table 5). In essence, the joint prior distribution for the parameter values, conditioned on the population dynamics model, is not neutral (non-informative) with respect to these catch-related quantities, but the data are influential enough to move the posterior distribution away from the mode of the prior distribution.

The posterior distributions for $r_{\max}$, $MSYR^{1+}$, and the life history parameters for this model are generally similar to those for the '1848 DD' model. The most noteworthy difference between the posterior distributions for the '1978 DD' and '1848 DD' models in Table 6 relates to the posterior median for S_{juv} which is larger for the '1978 DD' model. As was the case for the '1848 DD' model, there is again little sensitivity to changing the priors for S_a and for a_m, and ignoring the stage-proportion data (Tables 5 and 6). In contrast, inclusion of the 1985 survey data when calculating the stage-proportion data has a large impact on

Table 5

Posterior medians (5th, 95th percentiles) for eight management-related quantities. This table includes results for all models and scenarios, and Bayesian model-averaged results for the reference scenarios. Model averages for all quantities, except RY and N^{1+}_{2002} (i.e. the two quantities not dependent on K), could only be based on the 1848 and 1978 DD models. Where noted, additional scenarios are based on the reference scenarios in Table 4, but changed to examine the sensitivity of the results to different proportion data sets and less informative priors. The uniform prior for S_a is U[0.940, 0.995], and that for a_m is Discrete U[13, 26]. Unique draws and the maximum number of times an individual draw is included in the final SIR resample are listed in the rightmost two columns.

	N^{1+}_{2002}	K^{1+}	RY^{1+}	Q^{1+}_1	N^{1+}_{2002}/K^{1+}	$MSYL^{1+}$	MSY^{1+}	$MSYR^{1+}$	Unique draws	Max sampled draw (n)
1848 DD										
Reference scenario	9,496 (8,750, 10,180)	10,960 (9,190, 13,950)	171 (61, 233)	228 (149, 296)	0.888 (0.647, 0.985)	0.734 (0.639, 0.793)	253 (166, 329)	0.033 (0.019, 0.048)	7,399	9
No proportion data	9,380 (8,652, 10,070)	10,980 (9,245, 14,710)	160 (63, 222)	216 (132, 283)	0.874 (0.602, 0.983)	0.703 (0.555, 0.791)	240 (147, 315)	0.032 (0.017, 0.047)	9,456	4
Uniform priors on S_a and a_m	9,488 (8,782, 10,180)	10,580 (9,112, 13,610)	159 (59, 229)	241 (158, 301)	0.921 (0.674, 0.986)	0.740 (0.644, 0.795)	267 (176, 335)	0.035 (0.020, 0.049)	6,527	9
1978 DD										
Reference scenario	10,670 (9,042, 12,410)	20,510 (11,010, 29,120)	324 (147, 501)	295 (160, 439)	0.530 (0.356, 0.925)	0.651 (0.573, 0.779)	368 (206, 599)	0.030 (0.018, 0.043)	8,992	5
No proportion data	10,410 (8,740, 12,380)	20,350 (10,600, 29,070)	297 (119, 519)	270 (136, 457)	0.524 (0.348, 0.930)	0.661 (0.528, 0.785)	341 (169, 644)	0.029 (0.016, 0.043)	9,519	3
Include 1985 proportion data	9,294 (7,780, 10,720)	13,510 (8,110, 28,890)	166 (44, 280)	193 (74, 289)	0.705 (0.299, 0.990)	0.607 (0.459, 0.784)	225 (89, 333)	0.025 (0.009, 0.048)	8,034	8
Uniform priors on S_a and a_m	10,820 (9,124, 12,600)	19,870 (10,750, 29,090)	336 (138, 525)	309 (168, 459)	0.554 (0.362, 0.947)	0.654 (0.570, 0.781)	379 (215, 624)	0.032 (0.019, 0.045)	7,493	9
U[8,000, 100,000] prior on K	10,830 (9,160, 12,690)	54,430 (14,990, 95,500)	363 (188, 566)	267 (136, 426)	0.201 (0.110, 0.710)	0.606 (0.523, 0.777)	712 (282, 1,737)	0.025 (0.014, 0.038)	7,554	7
1978 NON DD										
Reference scenario	10,740 (9,130, 12,700)	N/A	366 (204, 588)	N/A	N/A	N/A	N/A	N/A	9,262	5
No proportion data	11,020 (9,199, 13,130)	N/A	402 (206, 654)	N/A	N/A	N/A	N/A	N/A	9,809	3
Include 1985 proportion data	10,280 (8,796, 12,090)	N/A	310 (166, 511)	N/A	N/A	N/A	N/A	N/A	6,264	9
Uniform priors on S_a and a_m	11,110 (9,265, 13,050)	N/A	414 (217, 644)	N/A	N/A	N/A	N/A	N/A	8,374	6
Bayesian model average	10,276 (8,907, 12,406)	13,854 (9,466, 28,475)	297 (92, 539)	257 (155, 412)	0.720 (0.372, 0.980)	0.695 (0.581, 0.788)	302 (180, 555)	0.031 (0.018, 0.046)	-	-

the results. Specifically, K^{1+} and $MSYR^{1+}$ are estimated to be lower, and N^{1+}_{2002}/K^{1+} higher, with the population estimated to be above $MSYL^{1+}$ with almost 100% probability. r_{max} is estimated to be higher (as the population is estimated to be closer to K^{1+} and therefore experiencing a growth rate much lower than r_{max}). The estimates of the catch-related quantities are considerably lower for this scenario (e.g. the posterior median and the lower 5% percentile for RY^{1+} are 166 and 44 respectively).

Forwards from 1978 (density independent model): 1978 NON DD

The posterior median for RY^{1+} for this model ranges from 310 to 414 across the scenarios (166-217 for the lower 5th percentile for RY^{1+}). The posterior for for the '1978 NON DD' model is centred on lower values than those for the other two models (Fig. 3). This is to be expected because the r for the '1978 NON DD' model' pertains to the current rate of increase rather than the increase rate in the limit of zero population size. The inclusion in the analyses of the 1985 stage-proportion data is again very influential. For example, the posterior distribution for fecundity for the 'include 1985 proportion data' sensitivity test does not overlap with that for the reference scenario.

Model comparison and Bayesian model averages

Bayes factors based on pair-wise comparisons of models range from 1.10 to 1.51, and indicate that there is no evidence for selecting one model over another (Table 7). Rather, these Bayes factors imply that the best approach to summarising the state of the B-C-B bowhead whale stock is to consider all three models, e.g. through Bayesian model averaging. Average likelihoods of draws from the initial sample range from 0.522 ('1848 DD') to 0.789 ('1978 NON DD'). The two models that involve forward projection from recent abundance ('1978 DD' and '1978 NON DD') have slightly higher average likelihoods, and hence posterior model probabilities, than the model which started the population projection in 1848, although differences are not large (Table 8).

The time-trajectory of 1+ population size (medians and 90% credibility intervals) from the Bayesian model-averaged posterior provides, as expected, a good fit to the abundance estimates (Fig. 1, lower panel). The fit to the estimate of abundance for 2001 for the model-averaged posterior is not quite as good as for the '1978 DD' and '1978 NON DD' models because of the impact of including the '1848 DD' model in the average. The model-averaged posterior distribution for K^{1+} (Fig. 2) has a mode close to
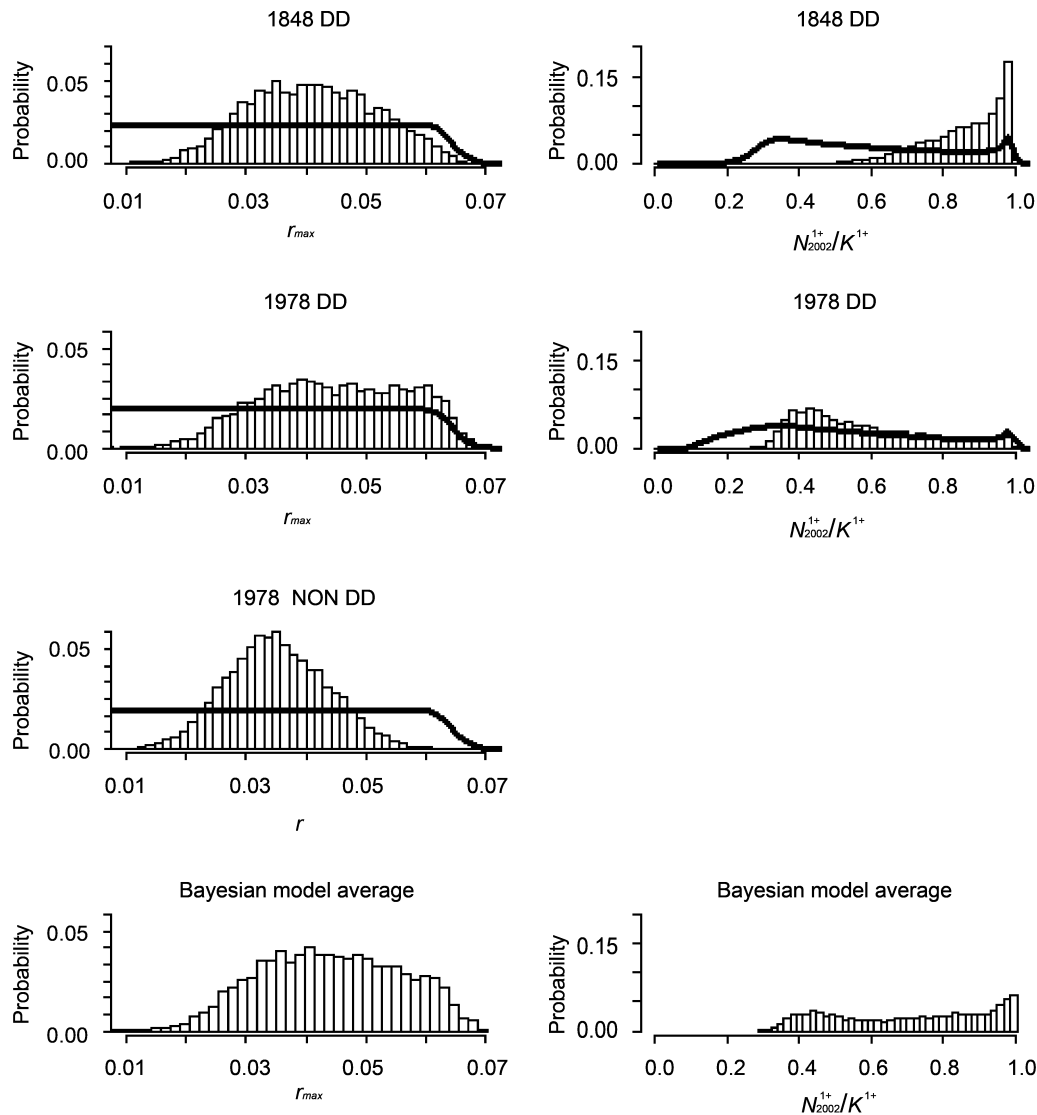
Fig. 3. Posterior (vertical bars) and post-model-pre-data (solid lines) distributions for intrinsic population growth rate, r_{max} and r (left panels) and recent depletion in terms of the 1+ component, N^{1+}_{2002}/K^{1+} (right panels). The Bayesian model average is based only on the two models with density dependence.

that of the posterior median for the '1848 DD' model, and a long tail caused by the uncertainty associated from the '1978 DD' model. The model-average posterior for N^{1+}_{2002} / K^{1+} (Fig. 3) is wide, but less so than that for the '1978 DD' model. The model-averaged posterior for RY^{1+} is slightly irregular because it consists of the combination of a bimodal posterior (for the '1848 DD' model) and a symmetric posterior (for the '1978 DD' model). In contrast to the model-average posterior for RY^{1+}, that for Q^{1+}_1 is actually quite symmetric (Fig. 4, Table 5).

The '1978 NON DD' model estimates only the recent fecundity and rate of increase for the population, whereas the two density-dependent models estimate the maximum fecundity and rate of increase. Therefore, the posterior for f_{max} and r_{max} is averaged across the two models with density dependence only. Maximum fecundity and population growth rate are relatively consistent across these two models and have a median of 0.171 and 4.3% respectively (Fig. 6 and Fig. 3, Table 6). Likewise, both adult and juvenile survival rates are consistent across models, with a median for adult survival of 0.990 and for juvenile survival of 0.932 (Fig. 5, Table 6).

It is straightforward to calculate model-averaged posterior probability distributions given different prior probability distributions for the models. For example, the models based on starting the projections in 1978 could be assigned probabilities of 0.25 and that which starts the population projections in 1848, a prior probability of 0.5 to indicate, for example, that the assumption that carrying capacity has not changed over the last 150 years is equally as likely as some shift in the equilibrium population size during this time. Alternatively, the models with density-dependence could be assigned prior probabilities of 0.25 each, and that which ignores density-dependence a prior probability of 0.5.

General discussion

The three models have shown good concurrence. However, use of the historic catch record leads to lower estimates of RY^{1+} because the analysis estimates the population to be close to carrying capacity and so the growth rate is reduced compared to that at low population size. However, there is no (visual) evidence in the abundance estimates for a reduction in trend. It therefore appears that it is the
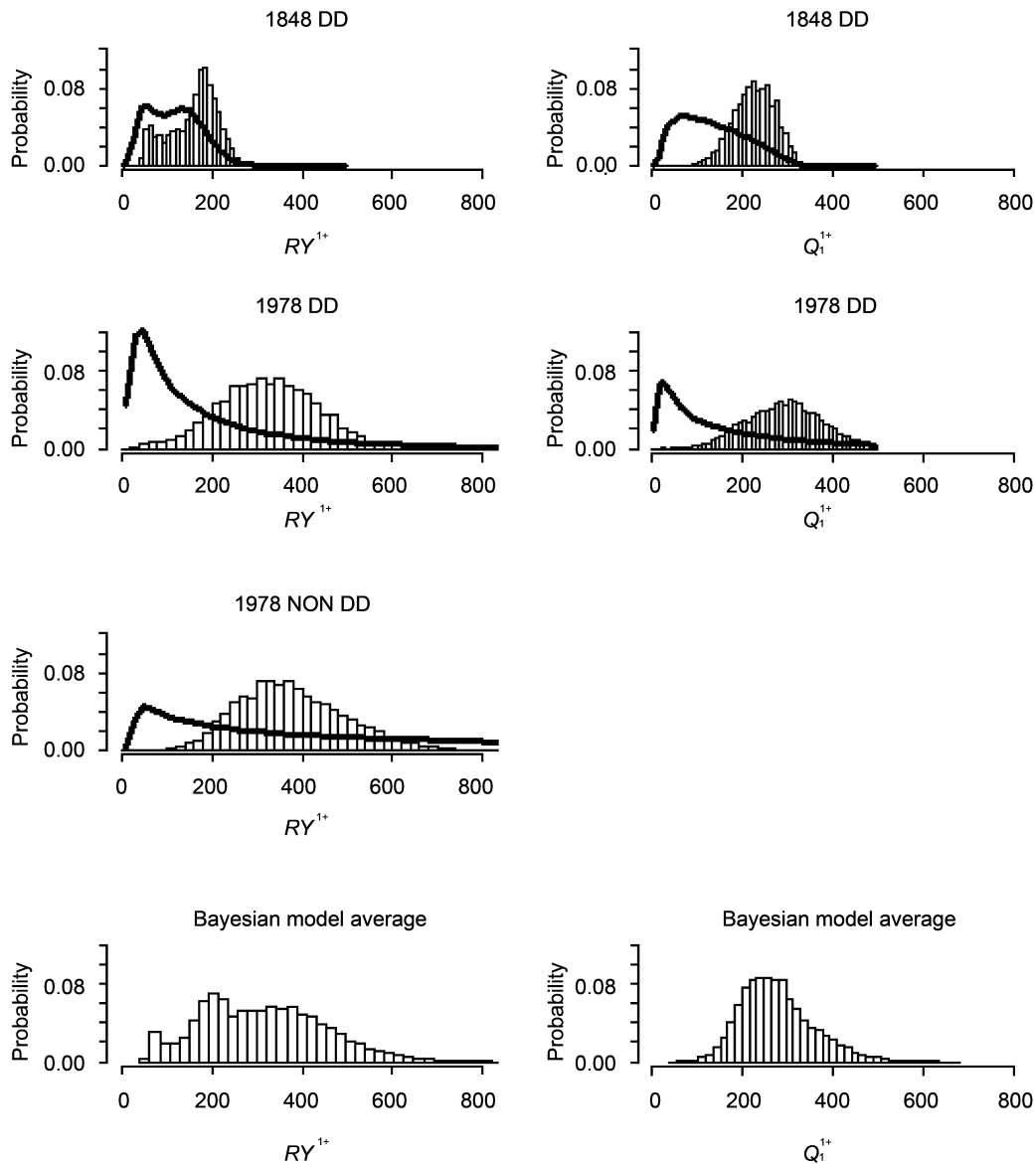
Fig. 4. Posterior (vertical bars) and post-model-pre-data (solid lines) distributions for replacement yield, RY^{1+} (left panels) and Q_1^{1+} (right panels). Results are shown for only two of the three reference scenarios for Q_1^{1+}, and the Bayesian model average for Q_1^{1+} is based on the results of these two scenarios only.

combination of the magnitude of the historical catches and the values for the biological parameters that determines the estimate of the carrying capacity. One implication of starting the population projection in 1848 is that the model under-predicts the 2001 estimate of abundance (probably because if the population is approaching carrying capacity, a near-linear growth in population size could not still be occurring). The '1848 DD' model is assigned less weight than the '1978 DD' and '1989 NON DD' models using Bayes factor, but the discrepancy between the predictions of the '1848 DD' model and the data remains sufficiently small that the '1848 DD' model cannot be rejected. It should be stressed that the results in this paper do not suggest that changes are required to the *Bowhead SLA* used to provide management advice on bowhead whales (e.g. IWC, 2003)

RY^{1+} is the catch that will keep a population at its current size. This quantity is less useful as the basis for management advice for the B-C-B bowhead whale stock now that at least some of the analyses suggest the recruited population may be approaching K^{1+}. Obviously RY^{1+} will be zero if the population stops increasing because it reaches carrying capacity. Q_1^{1+} is therefore a more appropriate catch-related

quantity to examine because it does not become zero at carrying capacity. Furthermore, this quantity represents a catch level that has been argued to meet the requirements of aboriginal subsistence management (Wade, 2002b; Wade and Givens, 1997). The fact that there is no evidence to select one model over the others and not all models result in similar estimates of catch quantities is a reason why model uncertainty is important to include when conducting assessments of marine renewable resources. The lower 5th percentile of the Bayesian model-averaged posterior of Q_1^{1+} is 155 whales, and represents our best estimate of the catch level that would meet the intent of aboriginal whaling management objectives, taking into account both parameter, and model uncertainty (to the extent that model uncertainty can be captured by the three models considered in this study). It should be noted, in light of recent discussions regarding stock structure (IWC, 2005), that the results presented here are based on the assumption that the B-C-B bowhead whales comprise a single stock.

The actual aboriginal catch quotas are driven by need, and have averaged 36 whales per year from 1978-2002. There appears to be little effect on population size due to this catch
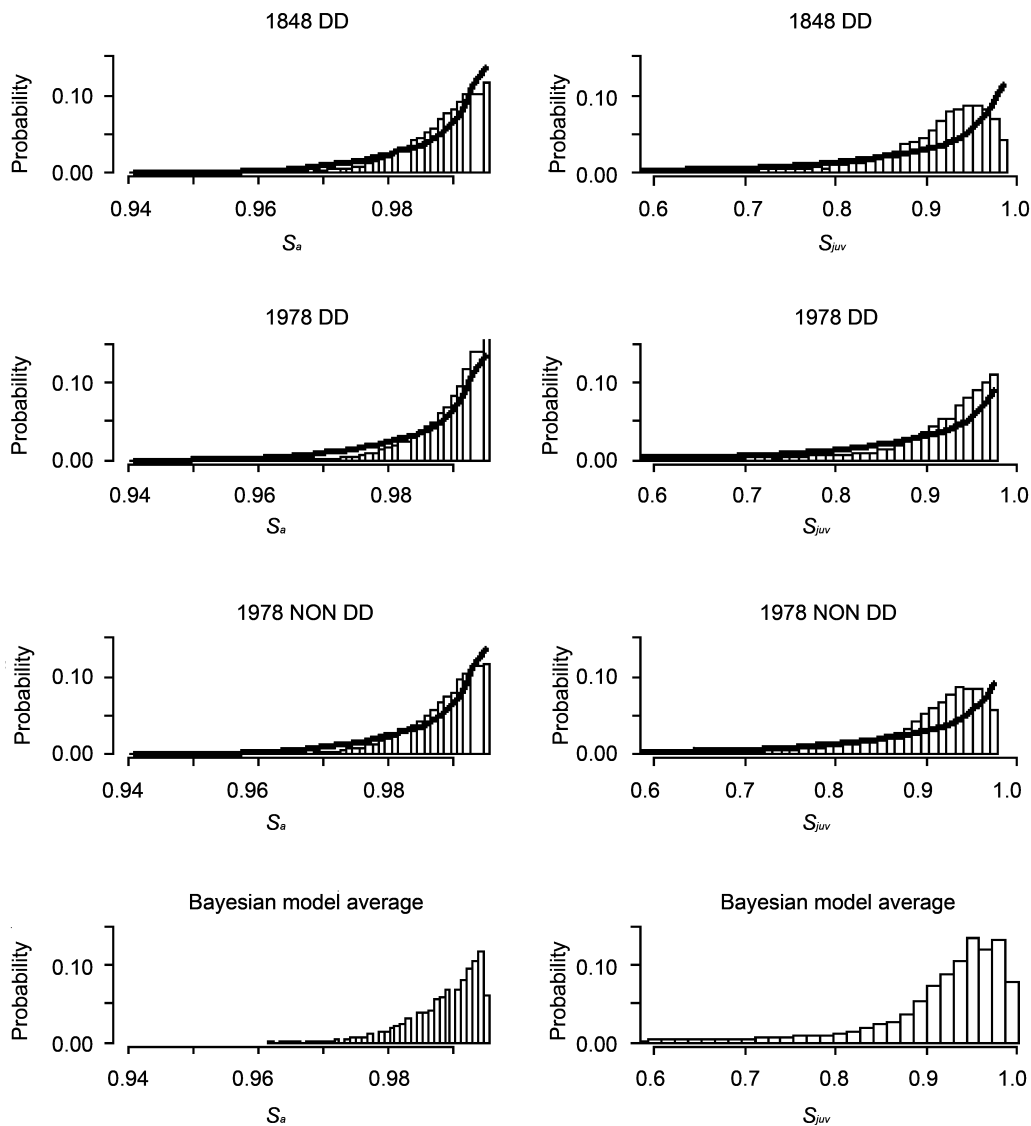
Fig. 5. Posterior (vertical bars) and post-model-pre-data (solid lines) distributions for adult survival rate, S_a (left panels) and calf and juvenile survival rate, S_{juv} (right panels). Results are shown for the three reference scenarios and for the Bayesian model average.

level. In fact, during this timeframe the B-C-B stock of bowhead whales is estimated to have more than doubled. Another way of putting an average take of 36 whales per year into perspective is to examine the annual net production over the last 25 years. If it is assumed that the population was increasing at a constant 3.5% per year (the median rate estimated from the density independent Leslie matrix analysis), the population increased by about 175 whales per year in 1978 (population size ~5,000), about 260 whales per year in 1990 (population size ~7,500), and about 350 whales per year in 2002 (population size ~10,000). Given that the population has increased from about 5,000 whales to about 10,000 whales, in this timeframe an average kill of 36 whales per year represents an annual catch rate between 0.35-0.70% of the total population size.

What is known about B-C-B bowhead whale life-history vital rates (survival, fecundity, etc.) appears consistent with the available data on trends in abundance and the proportion of the population in three stages (calves, immature and mature). Overall, the results support a value of r_{max} of between 0.03-0.05, a range often assumed for cetaceans, particularly species with delayed sexual maturity and a longer than 2-year calving interval (Reilly and Barlow,

1986; Wade, 1998; 2002a). In light of the reproductive life-history of this species, the results make clear that the observed population growth rates can only be supported by extremely high survival rates, as already suggested by the estimates of adult survival (0.990) in Zeh *et al.* (2002), and the observations of exceptionally old individuals (George *et al.*, 1999).

The analyses of this paper are based on the same types of data that were available for the 1998 assessment of this stock. Several other sources of data exist. For example, Schweder and Ianelli (2000) examined whether the data on the age-composition of the 1973-93 catches are consistent with the abundance and proportion data. Punt (2006) shows that it is possible to reconcile the abundance, proportion, length-frequency and age-composition data within a Bayesian framework. It would be straightforward conceptually (but perhaps computationally challenging) to use the approach outlined in this paper to compare models that utilise these additional data sources. Bringing in those additional data, as well as doing a full model comparison of a variety of models, was beyond the scope of this paper, but we agree this would be important future work, particularly in light of the methods now developed in Punt (2006).
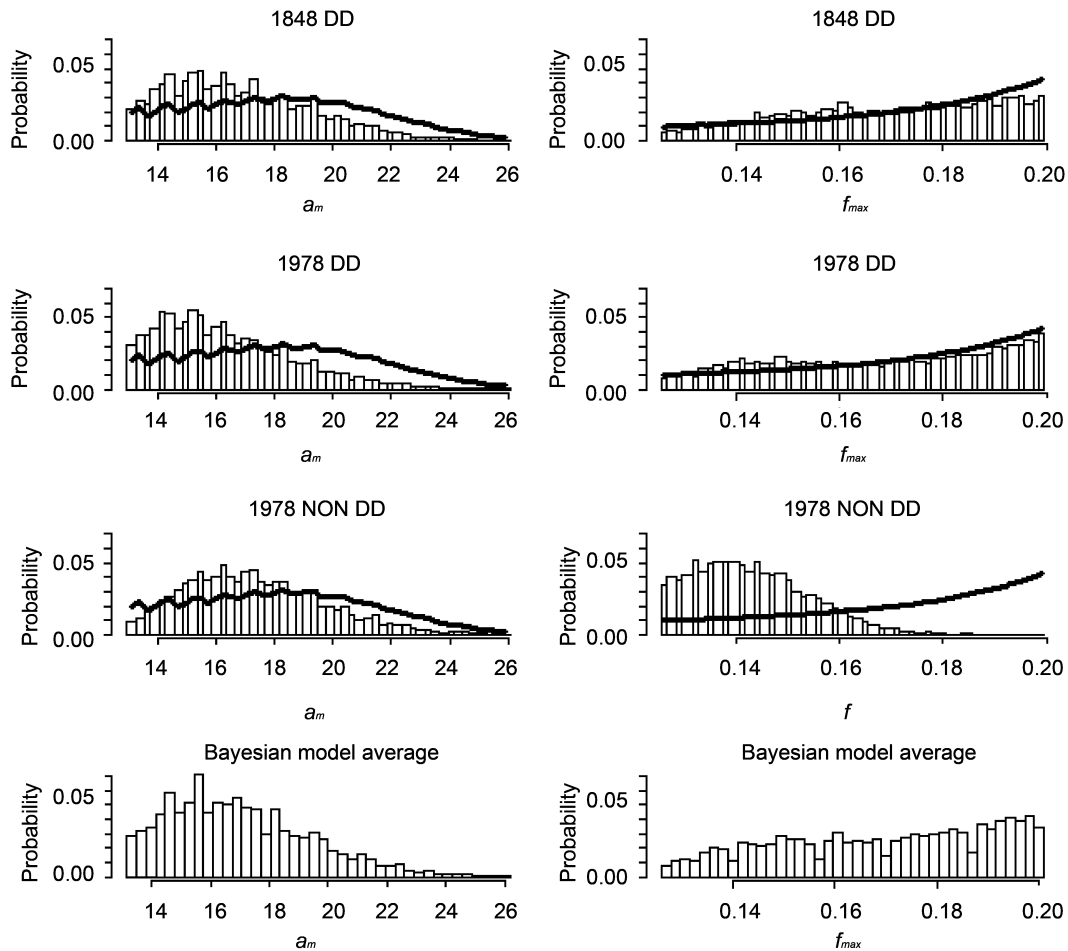
Fig. 6. Posterior (vertical bars) and post-model-pre-data (solid lines) distributions for the age-at-maturity, a_m (left panels) and fecundity, f_{max} or f (right panels). Results are shown for the three reference scenarios and for the Bayesian model average, which is based only on the two models with density dependence for fecundity.

Table 6

Posterior medians (5th, 95th percentiles) for eight life history-related quantities for all models and scenarios, and the Bayesian model-averaged results for the reference scenario. For the 1978 NON DD model, fecundity and population growth rate only apply to the specified period, and are referred to here as f and r. The model-averaged results for these rates are based only on the two models with density dependence, and refer to f_{max} and r_{max}.

	S_a	S_{juv}	f_{max}(or f)	r_{max}(or r)	a_m	a_T	P_c	P_m
1848 DD								
Reference scenario	0.989	0.926	0.171	0.041	16	5	0.055	0.436
	(0.977, 0.995)	(0.718, 0.980)	(0.135, 0.198)	(0.024, 0.059)	(14, 21)	(1, 9)	(0.046, 0.065)	(0.418, 0.453)
No proportion data	0.991	0.957	0.181	0.046	17	5	N/A	N/A
	(0.978, 0.995)	(0.739, 0.990)	(0.136, 0.199)	(0.024, 0.062)	(13, 23)	(1, 9)		
Uniform priors on S_a and a_m	0.988	0.943	0.174	0.044	15	5	0.056	0.437
	(0.972, 0.994)	(0.786, 0.981)	(0.138, 0.198)	(0.026, 0.059)	(13, 20)	(1, 9)	(0.048, 0.065)	(0.420, 0.454)
1978 DD								
Reference scenario	0.991	0.945	0.171	0.045	16	5	0.057	0.435
	(0.979, 0.995)	(0.765, 0.988)	(0.133, 0.198)	(0.025, 0.063)	(13, 21)	(1, 9)	(0.048, 0.066)	(0.418, 0.452)
No proportion data	0.991	0.957	0.181	0.045	17	5	N/A	N/A
	(0.978, 0.995)	(0.753, 0.990)	(0.136, 0.199)	(0.025, 0.062)	(13, 23)	(1, 9)		
Include 1985 proportion data	0.993	0.981	0.182	0.049	20	5	0.034	0.424
	(0.979, 0.995)	(0.750, 0.993)	(0.126, 0.199)	(0.027, 0.065)	(13, 25)	(1, 9)	(0.025, 0.040)	(0.394, 0.663)
Uniform priors on S_a and a_m	0.989	0.956	0.171	0.047	15	5	0.058	0.436
	(0.974, 0.995)	(0.816, 0.988)	(0.135, 0.198)	(0.027, 0.064)	(13, 19)	(1, 9)	(0.050, 0.066)	(0.419, 0.453)
U[8,000, 100,000] prior on K	0.990	0.940	0.162	0.042	16	5	0.057	0.435
	(0.978, 0.995)	(0.752, 0.987)	(0.130, 0.197)	(0.025, 0.061)	(14, 21)	(1, 9)	(0.049, 0.067)	(0.417, 0.453)
1978 NON DD								
Reference scenario	0.990	0.923	0.141	0.035	17	5	0.060	0.434
	(0.977, 0.995)	(0.717, 0.978)	(0.127, 0.163)	(0.022, 0.049)	(14, 22)	(1, 9)	(0.054, 0.068)	(0.417, 0.450)
No proportion data	0.989	0.936	0.173	0.038	19	5	N/A	N/A
	(0.975, 0.995)	(0.684, 0.985)	(0.132, 0.198)	(0.022, 0.053)	(14, 23)	(1, 9)		
Include 1985 proportion data	0.989	0.823	0.197	0.031	18	4	0.081	0.424
	(0.976, 0.995)	(0.477, 0.899)	(0.191, 0.200)	(0.018, 0.044)	(14, 22)	(1, 9)	(0.080, 0.084)	(0.414, 0.437)
Uniform priors on S_a and a_m	0.988	0.943	0.144	0.038	15	5	0.061	0.435
	(0.973, 0.995)	(0.783, 0.982)	(0.128, 0.166)	(0.023, 0.052)	(13, 21)	(1, 9)	(0.054, 0.069)	(0.418, 0.452)
Bayesian model average	0.990	0.932	0.171	0.043	16	5	0.058	0.435
	(0.978, 0.995)	(0.733, 0.984)	(0.133, 0.198)	(0.025, 0.062)	(14, 21)	(1, 9)	(0.049, 0.067)	(0.417, 0.452)

Table 7

Bayes factors for comparison of paired models. Evidence categories are modifications of the original categories of Jefferys (1961), as presented by Kass and Raftery (1995) and used by Wade (2002a) in an assessment of the eastern North Pacific gray whales: >150 is decisive evidence, 12-150 is strong evidence, 3-12 is positive evidence, and 1-3 is not worth more than a bare mention. All comparisons are based on the results of the reference scenarios.

Models	Bayes factor	Evidence for the first model
1978 DD *vs* 1848 DD	1.37	Not worth more than a bare mention.
1978 NON DD *vs* 1848 DD	1.51	”
1978 NON DD *vs* 1978 DD	1.10	”

Table 8

The average likelihood (Eqn. (10)) and posterior model probabilities used in the Bayesian model averaging. It was only possible to consider the two DD models for averaging quantities related to carrying capacity. All models were considered equally likely *a priori*.

Model	Average likelihood	Posterior model probabilities for models considered	
		All three	1848 DD and 1978 DD
1848 DD	0.522	0.258	0.422
1978 DD	0.715	0.353	0.578
1978 NON DD	0.789	0.389	N/A

This study represents the first attempt to quantify model uncertainty when conducting assessments of the B-C-B bowhead whale stock. The analyses consider three alternative models and take model uncertainty into account by weighting alternative models based on their posterior model probabilities and by calculating a Bayesian model-averaged posterior. The only previous attempt to consider model uncertainty when conducting assessments of whale stocks was by Wade (2002a), who compared models for the Eastern North Pacific stock of gray whales with and without additional variance about the abundance estimates. In that case, one model received almost all the weight making model-averaging redundant. In contrast, in this study all three models were assigned non-negligible weight and this led to different estimates of quantities of interest (e.g. carrying capacity and related measures). The Bayesian model-averaged posterior distribution clearly represents our best efforts to incorporate all levels of uncertainty in the estimates of these quantities.

ACKNOWLEDGMENTS

REFERENCES

Angliss, R.P., Rugh, D.J., Withrow, D.E. and Hobbs, R.C. 1995. Evaluations of aerial photogrammetric length measurements of the Bering-Chukchi-Beaufort Seas stock of bowhead whales (*Balaena mysticetus*). *Rep. int. Whal. Commn* 45:313-24.

Bockstoce, J.R. and Botkin, D.B. 1983. The historical status and reduction of the western Arctic bowhead whale (*Balaena mysticetus*) population by the pelagic whaling industry, 1848-1914. *Rep. int. Whal. Commn (special issue)* 5:107-41.

Braham, H.W. 1989. Eskimos, yankees, and bowheads. *Oceanus* 32(1):54-62.

Braham, H.W. 1995. Sex and size composition of bowhead whales landed by Alaskan Eskimo whalers. pp.281-313. *In*: McCartney, A.P. (ed.), *Hunting the Largest Animals: Native Whaling in the Western Arctic and Subarctic. Occasional Pub. No. 36*. The Canadian Circumpolar Institute, Edmonton, Alberta. 355pp.

Breiwick, J.M. and Braham, H.W. 1990. Historical population estimates of bowhead whales: sensitivity to current population size. *Rep. int. Whal. Commn* 40:423-26.

Breiwick, J.M., Eberhardt, L.L. and Braham, H.W. 1984. Population dynamics of western Arctic bowhead whales (*Balaena mysticetus*). *Can. J. Fish. Aquat. Sci.* 41(3):484-96.

Buckland, S.T., Burnham, K.P. and Augustin, N.H. 1997. Model selection: an integral part of inference. *Biometrics* 53:603-18.

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

da Silva, C.Q., Zeh, J., Madigan, D., Laake, J., Rugh, D., Baraff, L., Koski, W. and Miller, G. 2000. Capture-recapture estimation of bowhead whale population size using photo-identification data. *J. Cetacean Res. Manage.* 2(1):45-61.

de la Mare, W.K. 1989. Report of the Scientific Committee, Annex L. The model used in the HITTER and FITTER programs (Program:FITTER.SC40). *Rep. int. Whal. Commn* 39:150-51.

Dinesman, L.G. and Savinetsky, S.B. 2003. Secular dynamics of the prehistoric catch and population size of baleen whales off the Chukchi Peninsula, Siberia. pp.137-66. *In*: McCartney, A.P. (eds), *Indiginous Ways to the Present: Native Whaling in the Western Arctic.* Canadian Circumpolar Institute Press, Edmonton.

Durban, J.W., Elston, D.A., Ellifrit, D.K., Dickson, E., Hammond, P.S. and Thompson, P.M. 2005. Multisite mark-recapture for cetaceans: population estimates with Bayesian model averaging. *Mar. Mammal Sci.* 21(1):80-92.

George, J.C., Bada, J., Zeh, J., Scott, L., Brown, S.E., O'Hara, T. and Suydam, R. 1999. Age and growth estimates of bowhead whales (*Balaena mysticetus*) via aspartic racemization. *Can. J. Zool.* 77:571-80.

George, J.C., Philo, L.M., Carroll, G.M. and Albert, T.F. 1988. 1987 subsistence harvest of bowhead whales, *Balaena mysticetus*, by Alaskan Eskimos. *Rep. int. Whal. Commn* 38:389-92.

George, J.C., Zeh, J., Suydam, R. and Clark, C. 2004. Abundance and population trend (1978-2001) of western Arctic bowhead whales surveyed near Barrow, Alaska. *Mar. Mammal Sci.* 20(4):755-73.

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. *Rep. int. Whal. Commn* 43:495-500.

Givens, G.H., Zeh, J.E. and Raftery, A.E. 1995. Assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales using the BALEEN II model in a Bayesian synthesis framework. *Rep. int. Whal. Commn* 45:345-64.

Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T. 1999. Bayesian Model Averaging: a tutorial. *Statistical Sciences* 14(4):382-417.

International Whaling Commission. 1995. Report of the Scientific Committee, Annex F. Report of the sub-committee on aboriginal subsistence whaling. *Rep. int. Whal. Commn* 45:142-64.

International Whaling Commission. 1998. Report of the Scientific Committee. Annex J. Report of the Sub-Committee on Aboriginal Subsistence Whaling. *Rep. int. Whal. Commn* 48:237-48.

International Whaling Commission. 1999. Report of the Scientific Committee. Annex G. Report of the Sub-Committee on Aboriginal Subsistence Whaling. *J. Cetacean Res. Manage. (Suppl.)* 1:179-94.

International Whaling Commission. 2003. Report of the Scientific Committee. Annex E. Report of the Standing Working Group on the Development of an Aboriginal Subsistence Whaling Management Procedure (AWMP). *J. Cetacean Res. Manage. (Suppl.)* 5:154-255.

International Whaling Commission. 2005. Report of the Scientific Committee. Annex F. Report of the Sub-Committee on Bowhead, Right and Gray Whales. *J. Cetacean Res. Manage. (Suppl.)* 7:189-208.

Jeffreys, H. 1961. *Theory of Probability*. 3rd ed. Oxford University Press, London. i-vii+411pp.

Kass, R.E. and Raftery, A.E. 1995. Bayes factors. *J. Am. Stat. Assoc.* 90:773-95.

Koski, W.R., Rugh, D.J., Punt, A.E. and Zeh, J. 2004. A new approach to estimating the length-frequency distribution of the Bering-Chukchi-Beaufort bowhead whale population using aerial photogrammetry and a summary of other life-history parameters estimated from photo-identification/photogrammetry data. Paper SC/56/BRG2 presented to the IWC Scientific Committee, July 2004, Sorrento, Italy (unpublished). 25pp. [Paper available from the Office of this Journal].

Leslie, P.H. 1945. On the use of matrices in certain population mathematics. *Biometrika* 33:183-212.

Leslie, P.H. 1948. Some further notes on the use of matrices in population mathematics. *Biometrika* 35:213-45.

Patterson, K.R. 1999. Evaluating uncertainty in harvest control law catches using Bayesian Markov chain Monte Carlo virtual population analysis with adaptive rejection sampling and including structural uncertainty. *Can. J. Fish. Aquat. Sci.* 56:208-21.

Poole, D. and Givens, G.H. 2001. An exploratory assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales using a stochastic population model. *J. Cetacean Res. Manage.* 3(1):1-6.

Poole, D., Givens, G.H. and Raftery, A.E. 1999. A proposed stock assessment method and its application to bowhead whales, *Balaena mysticetus*. *Fish. Bull.* 97:144-52.

Poole, D. and Raftery, A.E. 1998. Assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales using Bayesian and full pooling Bayesian synthesis methods. Paper SC/50/AS6 presented to the IWC Scientific Committee, April 1998 (unpublished). 34pp. [Paper available from the Office of this Journal].

Press, S.J. 1989. *Bayesian Statistics: Principles, Models and Applications*. John Wiley, New York. 237pp.

Punt, A.E. 1999a. Report of the Scientific Committee, Annex G. Appendix 4. Results for two selected assessments of the Bering-Chukchi-Beaufort Seas stock of bowhead whales. *J. Cetacean Res. Manage. (Suppl.)* 1:192-94.

Punt, A.E. 1999b. Report of the Scientific Committee. Annex R. A full description of the standard BALEEN II model and some variants thereof. *J. Cetacean Res. Manage. (Suppl.)* 1:267-76.

Punt, A.E. 2006. Assessing the Bering-Chukchi-Beaufort Seas stock of bowhead whales using abundance data together with data on length or age. *J. Cetacean Res. Manage.* 8(2):127-37.

Punt, A.E. and Butterworth, D.S. 1999. On assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales (*Balaena mysticetus*) using a Bayesian approach. *J. Cetacean Res. Manage.* 1(1):53-71.

Punt, A.E. and Butterworth, D.S. 2000. Why do Bayesian and Maximum likelihood assessments of the Bering-Chukchi-Beaufort Seas stock of bowhead whales differ? *J. Cetacean Res. Manage.* 2(2):125-33.

Punt, A.E. and Butterworth, D.S. 2002. An examination of certain of the assumptions made in the Bayesian approach used to assess the eastern North Pacific stock of gray whales (*Eschrichtius robustus*). *J. Cetacean Res. Manage.* 4(1):99-110.

Raftery, A., Givens, G.H. and Zeh, J.E. 1995. Inference from a deterministic population dynamics model for bowhead whales. *J. Am. Stat. Assoc.* 90(430):402-30.

Raftery, A.E., Gneiting, T., Balabdaoui, F. and Polakowski, M. 2005. Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather Rev.* 133(5):1,155-74.

Raftery, A.E. and Poole, D. 1997. Bayesian Synthesis assessment methodology for Bowhead whales. Paper SC/49/AS5 presented to the IWC Scientific Committee, September 1997, Bournemouth (unpublished). 40pp. [Paper available from the Office of this Journal].

Raftery, A.E. and Zeh, J.E. 1991. Bayes empirical Bayes estimation of bowhead whale population size based on the visual and acoustic census near Barrow, Alaska, in 1986 and 1988. Paper SC/43/PS8 presented to the IWC Scientific Committee, May 1991 (unpublished). 51pp. [Paper available from the Office of this Journal].

Raftery, A.E. and Zeh, J.E. 1993. Estimation of bowhead whale, *Balaena mysticetus*, population size. pp.163-240. *In*: Gatsonis, C., Hodges, J.S., Kass, R.E. and Singpurwalla, N.D. (eds), *Case Studies in Bayesian Statistics*. Springer-Verlag, New York.

Reilly, S.B. and Barlow, J. 1986. Rates of increase in dolphin population size. *Fish. Bull.* 84(3):527-33.

Rubin, D.B. 1988. Using the SIR algorithm to simulate posterior distributions. pp.395-402. *In*: Bernardo, J.M., DeGroot, M.H., Lindley, D.V. and Smith, A.F.M. (eds), *Bayesian Statistics 3: Proceedings of the Third Valencia International Meeting, June 1-5, 1987*. Clarendon Press, Oxford. 805pp.

Schweder, T. and Ianelli, J.N. 2000. Assessing the Bering-Chukchi-Beaufort Seas stock of bowhead whales from survey data, age-readings and photo-identifications using frequentist methods. Paper SC/52/AS13 presented to the IWC Scientific Committee, June 2000, Adelaide, Australia (unpublished). 16pp. [Paper available from the Office of this Journal].

Smith, A.F.M. and Gelfand, A.E. 1992. Bayesian statistics without tears: a sampling-resampling perspective. *Am. Stat.* 46(2):84-88.

Suydam, R.S. and George, J.C. 2004. Subsistence harvest of bowhead whales (*Balaena mysticetus*) by Alaskan Eskimos, 1974 to 2003. Paper SC/56/BRG12 presented to the IWC Scientific Committee, July 2004, Sorrento, Italy (unpublished). 12pp. [Paper available from the Office of this Journal].

Wade, P.R. 1998. Calculating limits to the allowable human-caused mortality of cetaceans and pinnipeds. *Mar. Mammal Sci.* 14(1):1-37.

Wade, P.R. 2002a. A Bayesian stock assessment of the eastern Pacific gray whale using abundance and harvest data from 1967-1996. *J. Cetacean Res. Manage.* 4(1):85-98.

Wade, P.R. 2002b. Population dynamics. pp.974-79. *In*: Perrin, W.F., Wursig, B. and Thewissen, J. (eds), *Encyclopedia of Marine Mammals*. Academic Press, San Diego. 1,414pp.

Wade, P.R. and Givens, G.H. 1997. Designing catch control laws that reflect the intent of aboriginal subsistence management principles. *Rep. int. Whal. Commn* 47:871-74.

Zeh, J., Poole, D., Miller, G., Koski, W., Baraff, L. and Rugh, D. 2002. Survival of bowhead whales, *Balaena mysticetus*, estimated from 1981-98 photo-identification data. *Biometrics* 58:832-40.

Zeh, J.E. and Punt, A. 2005. Updated 1978-2001 abundance estimates and their correlations for the Bering-Chuckchi-Beaufort Seas stock of bowhead whales. *J. Cetacean Res. Manage.* 7(2):169-75.