A Bayesian stock assessment of the eastern Pacific gray whale using abundance and harvest data from 1967-1996[1]

PAUL R. WADE

Office of Protected Resources, National Marine Fisheries Service, NOAA, c/o National Marine Mammal Laboratory, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, WA 98115-0070, USA
Contact e-mail: paul.wade@noaa.gov

ABSTRACT

Abundance and harvest data since 1966/67 were used to assess the eastern Pacific stock of gray whales. A Bayesian statistical method was used to estimate probability distributions for the parameters of both a simple and an age and sex structured population dynamics model, as well as output quantities of interest. Model comparisons using the Bayes Factor provided conclusive evidence that an additional parameter should be used to account for unexplained variation in the abundance time series. Incorporating the additional variance parameter decreased the precision of the estimates of the other parameters. Point estimates of carrying capacity ranged from 24,640-31,840 for the different models, but the posterior distributions from the selected models were very broad and excluded few values. The current depletion level (population size as a fraction of carrying capacity) was estimated to be about 0.75, with a lower 2.5th percentile of 0.36. The probability that the population was still below one-half of its carrying capacity was estimated to be 0.21, with a corresponding probability of 0.28 that the population was still below its maximum sustainable yield level. Quantities from which catch limits could potentially be calculated were estimated, including current replacement yield, maximum sustainable yield and the quantity Q_1 (described in Wade and Givens, 1997).

KEYWORDS: GRAY WHALE; PACIFIC OCEAN; MODELLING

INTRODUCTION

The eastern Pacific stock of gray whales (*Eschrichtius robustus*) was harvested commercially from the mid-1800s to recent times and has been harvested by aboriginals since at least the 1600s. Although the commercial harvest no longer occurs, an aboriginal harvest still exists (e.g. IWC, 1998, p.243). The gray whale is one of the best studied populations of whales. The population was surveyed on its southbound migration in 16 out of 21 years from 1966/67 to 1987/88. Three additional surveys were carried out in 1992/93 and 1993/94 (Laake *et al.*, 1994) and in 1995/96 (Hobbs *et al.*, 1996), making possible a reassessment of the stock. This paper reports the results from fitting two density-dependent models to the data. The status of the stock relative to its equilibrium population size was estimated, along with quantities of interest for setting catch limits for the harvest.

Two different population dynamics models were used: (1) generalised-logistic; and (2) density-dependent Leslie matrix. Model 1 is a simple model while model 2 is both age- and sex-structured. Bayesian statistical methods (e.g. Press, 1989) were used to estimate the model parameters by fitting the models to the abundance data. The models were compared using Bayes factors (Kass and Raftery, 1995) to evaluate which model best fitted the data.

Attempts to fit models to the available time series of gray whale abundance estimates have indicated that there is a lack of fit of the data to the model not accounted for by the estimated variance of the abundance estimates. This is readily apparent when the abundance estimates are examined, as several adjacent estimates have non-overlapping 95% confidence limits, indicating significant differences from one year to the next (Reilly, 1992). For example, the significant increase of greater than 30% from the 1992/93 estimate to the 1993/94 estimate (Table 1) is biologically implausible for gray whales and suggests that the estimated CVs do not account for all of the

Table 1

Abundance estimates used in the analysis, taken from Buckland and Breiwick (2002) for 1967/68 to 1987/88, from Laake *et al.* (1994) for 1992/93 and 1993/94 and from Hobbs *et al.* (1996) for 1995/96. Standard errors (SE) and coefficients of variation (CV) are also shown.

Year	Abundance	SE	CV
1967/68	13,012	893	0.069
1968/69	12,244	484	0.040
1969/70	12,777	525	0.041
1970/71	11,170	806	0.072
1971/72	9,841	442	0.045
1972/73	16,962	660	0.039
1973/74	14,817	592	0.040
1974/75	13,134	540	0.041
1975/76	14,811	690	0.047
1976/77	15,950	524	0.033
1977/78	17,127	966	0.056
1978/79	13,300	501	0.038
1979/80	16,581	668	0.040
1984/85	21,942	994	0.045
1985/86	20,450	727	0.036
1987/88	21,113	688	0.033
1992/93	17,674	1,029	0.058
1993/94	23,109	1,262	0.055
1995/96	22,571	1,174	0.052

variance associated with the estimates of the number of southbound migrating gray whales in each year. It is unknown whether this additional variance is due to variance in components of the estimation technique that are not accounted for, or to variance in the proportion of the population that migrates past the California monitoring site. Therefore, a method was developed to account for this unexplained variance by estimating a new parameter that represents additional variance in the abundance estimates. Adding the additional variance term to each of the two models resulted in a total of four models being considered.

An unresolved issue regarding the eastern Pacific gray whale is that it has not been possible to reconcile the catch history from the 1800s with the recent time series of abundance data in a simple way. Several attempts have been made to project population models forwards from the 1800s

assuming the population was at carrying capacity prior to the start of commercial harvests in 1846, but such projections cannot produce a trend that agrees with the recent abundance estimates, which indicate the population roughly doubled between 1967 and 1988 (Reilly, 1981; Lankester and Beddington, 1986; Butterworth *et al.*, 2002). The catch history and current trend can only be reconciled through fairly dramatic assumptions, such as an increase in the carrying capacity from 1846-1988 of at least 2.5 times, an underestimation of the historic commercial catch from 1846-1900 of at least 60%, or annual aboriginal catch levels prior to 1846 of at least three times the level previously thought (Butterworth *et al.*, 2002). Although all of these scenarios are plausible, there is little concrete evidence for any of them. Any stock assessment based on projections from a pre-1900 carrying capacity would involve making an untestable assumption such as that the commercial harvest was twice the level previously thought.

A stock assessment based on data collected since 1967, while ignoring some important information regarding the catch history, will not be subject to these potential problems. Therefore, the status of the eastern Pacific stock of gray whales has been assessed here using only the harvest data collected since 1967. The analysis starts projecting the population in 1967 and does not make any assumptions about what level the population was at relative to carrying capacity at that time.

METHODS

Available data
Abundance surveys for the eastern Pacific stock of gray whales take place from December to February, so they are referred to by two years (e.g. a survey from December 1995 to February 1996 is called the 1995/96 survey). Abundance estimates (Table 1) for 1967/68 to 1987/88 were from Buckland and Breiwick (2002). These were the same estimates as used in Butterworth *et al.* (2002). Preliminary estimates for 1992/93 and 1993/94 were from Laake *et al.* (1994), while that for 1995/96 is from Hobbs *et al.* (1996). The catch history prior to 1993 was obtained directly from the IWC (Table 2), while that for 1994 was obtained from Blokhin (1995).

Population dynamics models
Two different models were used:

Generalised-Logistic

$$N_{t+1} = N_t \left[1 + r_{max} \left(\frac{N_t}{N_{eq}} \right)^z \right] - H_t \qquad (1)$$

where:

N_t the population size in year t;
r_{max} the maximum net recruitment rate;
N_{eq} the equilibrium population size (or 'carrying capacity');
z the density-dependent exponent which sets the maximum sustainable yield level (the relative population size at which the maximum sustainable yield is obtained); and
H_t the number of animals harvested in year t.

Table 2

Harvest data used in analysis, obtained directly from the IWC. The proportion of females was assumed to be 67% in 1991. The observed proportion of females in 1994 was 0.666 (Blokhin, 1995). The 1995 catch data* (Borodin, 1996) were not yet available at the time of the analysis, and thus have not been subtracted in projecting the population to 1996.

Year	Male	Female	Total
1966	123	97	220
1967	94	156	250
1968	67	134	201
1969	59	155	214
1970	26	125	151
1971	51	102	153
1972	22	160	182
1973	97	81	178
1974	94	90	184
1975	58	113	171
1976	69	96	165
1977	86	101	187
1978	94	90	184
1979	57	126	183
1980	53	128	181
1981	36	100	136
1982	56	112	168
1983	46	125	171
1984	59	110	169
1985	55	115	170
1986	46	125	171
1987	47	112	159
1988	43	108	151
1989	61	119	180
1990	67	96	163
1991	57	113	170
1992	0	0	0
1993	0	0	0
1994	15	29	44
1995	44*	41*	85*

Density-Dependent Leslie Matrix
The population was projected as:

$$n_{t+1} = A_t n_t - h_t \qquad (2)$$

where:

n_t vector of population size of each age class at time t;
A_t the Leslie matrix in year t;
h_t vector of age-specific harvests at time t.

A simplified Leslie matrix (Leslie, 1945; 1948) was used with a total of four variable parameters. Three were elements of the matrix: (1) s_j, the survival rate of unrecruited age classes ('juvenile' survival rate); (2) s, the survival rate of recruited age classes ('adult survival rate'); (3) f_t, the fecundity rate at time t (assumed identical for all mature age classes); and (4) *ASM*, the age of sexual maturity (with the first non-zero fecundity in the subsequent age class). The maximum age was fixed at 60 years.

This model was identical to a usual Leslie matrix model, except that the fecundity term was density-dependent with a form similar to the generalised-logistic:

$$f_t = f_0 + (f_{max} - f_0) \left[1 - \left(\frac{N_t}{N_{eq}} \right)^z \right] \qquad (3)$$

where:

f_t the realised fecundity in year t;
f_{max} the maximum fecundity rate; and
f_0 fecundity at a net recruitment of zero, which can be solved directly from the other parameters.

This model is thus identical to the model in Breiwick *et al.* (1984). The population growth rate (λ) associated with a Leslie matrix using a fecundity value of f_t was referred to as λ_t and the population growth rate associated with $f_t = f_{max}$ was referred to as λ_{max}. Estimates of λ_{max} were expressed as λ_{max}-1 so they would be comparable to the parameter r_{max} of the generalised-logistic model.

The harvest kills were subtracted after reproduction because in recent decades they have occurred in the summer, after the calving season in the winter. In each year, the kills were distributed to each recruited age-class according to the age distribution in that year. Recruitment was assumed to be knife-edge and to occur at age five. Each trajectory was initiated with the stable age distribution associated with that population size. In other words, the starting population size for a trajectory was used in equation (3) to find the value of f_t associated with that population size and the particular values for f_{max}, N_{eq} and z used on that trajectory. Then the stable age distribution was found for the Leslie matrix composed of s, s_j, ASM and that value of f_t.

As the sex-ratio of the kill was not equal, two vectors of population size were projected, one for each sex. The same survival rates were used in the Leslie matrix to project each vector, but, obviously, the fecundity terms were set to zero when projecting the male population vector. The sex-ratio of calves was assumed to be 50:50; therefore the number of males in age-class 1 was set to be equal to the number of age-class 1 females at each time step. The population was assumed to have a 50:50 sex ratio at the beginning of 1968.

Additional variance term

Two additional models were specified by adding a parameter, CV_{add}, to models 1 and 2 that represented additional variance in the abundance estimates. This was similar to the method employed by Butterworth *et al.* (1993). Here, the additional variance is thought of as the unexplained variance between annual estimates of abundance not accounted for by the estimated variance of those abundance estimates; note that the mechanism that causes the process error is unknown and is not explicitly modelled here. CV_{add} was parameterised as a coefficient of variation and was considered constant across years. This implies that in any year there was the same possibility of additional variance not accounted for by the estimated variance of the abundance estimates. CV_{add} was incorporated into the likelihood function in each year as an additive variance term to the abundance estimates, with the assumption that this additional variance has a Gaussian distribution. In other words, in any year, a new total CV was calculated as the square root of the sum of the squares of $CV(t)$ and CV_{add}

$$CV_{tot}(t) = \sqrt{CV^2(t) + CV_{add}^2} \qquad (4)$$

where $CV(t) = S(t)/N(t)$. The likelihood component from that year's abundance estimate was calculated as usual with the new total CV term (i.e. $S(t) = N(t)CV_{tot}(t)$).

Statistical methods

A Bayesian statistical method (e.g. Press, 1989) was used to estimate the parameters of the models and other output quantities. The same techniques were used to investigate the population dynamics of eastern tropical Pacific dolphins (Wade, 1994) and spectacled eiders (Taylor *et al.*, 1996). The method is somewhat similar in approach to Bayesian synthesis analyses of bowhead whales, *Balaena mysticetus* (Givens *et al.*, 1993; 1995; Raftery *et al.*, 1995).

Any Bayesian analysis involves integrating the product of a prior distribution for a parameter and a likelihood function that links the probability of the observed data to the specification of particular values for the parameter. The likelihood function for the parameters in a population model, given a time-series of abundance estimates, was calculated according to the methods reported in de la Mare (1986). In any single year, the likelihood of an observed abundance estimate $N(t)$ given a specified model population size N_t is straight forward; it is the likelihood function defined by the assumed sampling distribution of the abundance estimate. The sampling distribution of the abundance estimates is assumed to be a Gaussian distribution with estimated mean $N(t)$ and standard error $S(t)$ and thus the likelihood is:

$$L\left(N_t \mid N(t), S(t)\right) = \frac{1}{\sqrt{2\Pi}S(t)} e^{-\frac{1}{2}\left(\frac{N(t) - N_t}{S(t)}\right)^2} \qquad (5)$$

Although N_t is not an explicit parameter of the model, the model parameters uniquely determine a population trajectory $N_{68/69}, N_{69/70}, \ldots N_{95/96}$. Therefore, the total likelihood given the data is the product series of all the individual likelihoods of the N_ts (the model trajectory) given the $N(t)$s (the time-series of abundance estimates).

$$L\left(\theta \mid \bar{N}(t), \bar{S}(t)\right) = \prod_{t=68/69}^{95/96} \left[L(N_t \mid N(t), S(t)) \right] \qquad (6)$$

Note that the 1967/68 abundance estimate is not used in the likelihood calculation because it is used as a prior distribution for the population size in the beginning of 1968 (see below).

The necessary integration was approximated by using the Sampling-Importance-Resampling routine of Rubin (1988), which Smith and Gelfand (1992) advocate as a particularly useful and simple integration technique for Bayesian statistics. In this method, values for the parameters are randomly selected from their joint prior distribution to form a sample set of parameter values, called θ_i. The likelihood of the data given this particular θ_i is calculated and stored. This is repeated, generating an initial sample of n_1 θ_is with associated likelihoods. This initial sample of n_1 θ_is is then re-sampled n_2 times with replacement with probability equal to weight q_i, where:

$$q_i = \frac{L(\theta_i \mid x)}{\sum_{j=1}^{n_1} L(\theta_j \mid x)} \qquad (7)$$

This process forms a second sample, called the re-sample. Rubin (1988) showed that the re-sample is a random sample from the joint posterior distribution of size n_2. Because the initial sample is sampled with replacement, the re-sample can contain more than one occurrence of the same set of values. When properly implemented, this feature of the method improves its efficiency and makes the re-sample a better approximation of the posterior distribution than would a random sample from the joint prior distribution of similar size. However, if the initial sample is not large enough, the re-sample may be overly influenced by just a few sets of values that occur many times. Such problems can be avoided by inspecting some characteristics of the re-sample, such as the number of unique θ_is and the maximum number of occurrences of a single θ_i and increasing the size of the initial sample if necessary. In this study, values for n_1 were set to values large enough to avoid those problems and thus yield sufficiently smooth posterior distributions. The actual value

used for n_1 varied in the different analyses depending upon the number and kind of parameters estimated. The value of n_2 was set to 5,000 for each analysis.

The surveys occurred in the beginning of each year and were assumed to apply to the 1+ population, as in previous studies such as Butterworth *et al.* (2002). The likelihood at any time-step t was calculated prior to reproduction and harvest in the model (i.e. the population size at the end of year t-1) and therefore the model population size used in equations (1) and (2) was the sum of all males and females in all age classes.

Prior distributions

$N_{67/68}$ was the initial population size (at the beginning of 1968) in each model. The prior distribution for $N_{67/68}$ was based on the abundance estimate in that year (1967/68) and was thus a Gaussian distribution with mean 13,012 and

standard error 893 (Table 1). For the analyses with the additional variance term, the prior distribution for $N_{67/68}$ is dependent on the value of CV_{add} and was thus a Gaussian with mean 13,012 and standard error of $CV_{tot}(67/68)*13,012$. In other words, for any particular set of parameter values drawn from the joint prior, a value for CV_{add} would be chosen first and then that value used to draw a value from the prior for $N_{67/68}$. The resulting prior distribution for $N_{67/68}$ thus represented a combination of normal distributions with the same mean but different variances.

The prior distributions of the population growth parameters in the different models (e.g. r_{max} and λ_{max}-1) were uniform distributions from 0.01 to 0.13. Preliminary trials of the analyses indicated there was little posterior probability below this range, which was confirmed in the final analyses. There was some posterior probability above
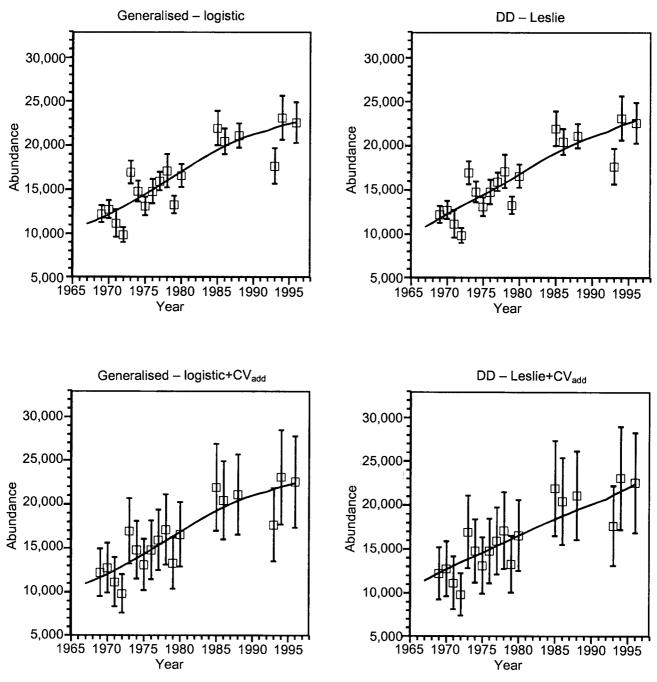

Fig. 1. Population trajectories with the greatest probability for all four model combinations. Also shown are abundance estimates with 95% confidence limits. Note that for the models with the additional variance term CV_{add}, the confidence limits have been re-calculated using the estimated total CV for each abundance estimate.

this range for some of the models, but the prior distributions on the life history parameters of the Leslie matrix make a value of 0.13 very unlikely for λ_{max}-1 and so that value was set as the upper limit.

The ranges for the life-history parameters in the age-structured model were set to be wide enough to encompass what are thought to be reasonable values (IWC, 1993). The prior distribution for f_{max} was a uniform distribution from 0.15 to 0.30 (parameterised as female calves per adult female). Age of sexual maturity (*ASM*), with the first non-zero fecundity in the next year, was a discrete uniform from 5-9 years. The adult survival rate, s, was a uniform distribution from 0.95 to 0.999. The only restriction on the juvenile survival rate, s_j, was that it be less than s. The prior distributions for these parameters did not remain uniform on these ranges. As mentioned above, a uniform distribution was set for λ_{max}-1. Then, values for f_{max}, *ASM* and s were drawn from the prior distributions described above. From these values, s_j can be calculated (Breiwick *et al.*, 1984). If $s_j < s$, this set of values was used. If s_j was $> s$, then f_{max}, *ASM* and s were re-drawn from their uniform distributions but retaining the original value for λ_{max}-1. This resulted in non-uniform realised prior distributions for these parameters, which were stored and were plotted along with

the posterior distributions, but retained the uniform prior distribution for λ_{max}-1. An explicit prior was not set for s_j because this would have resulted in two different prior distributions being established for λ_{max}-1 and because little information exists regarding s_j.

The prior distribution for N_{eq} was set as uniform from 17,000 to 70,000. The lower bound was found, through preliminary analyses, to have very little probability in any of the posterior distributions. Therefore, the lower bound for the prior distribution is uninformative in the sense that any lower value could have been used instead without influencing the results (although computation time would increase because the value of n_1 would need to be increased). The upper bound was set to a value thought to be greater than the greatest possible value, but it was necessarily a somewhat arbitrary value. The historic catch information has been used to estimate historic population size by back-calculating from a recent abundance estimate. Between the start of commercial whaling and 1900, approximately 15,000 whales were estimated to have been harvested. Using this information, Henderson (1972) concluded that the population did not exceed 15,000-20,000 at the start of commercial whaling. However, as discussed above, Reilly (1981) and Butterworth *et al.* (2002) have shown that it is
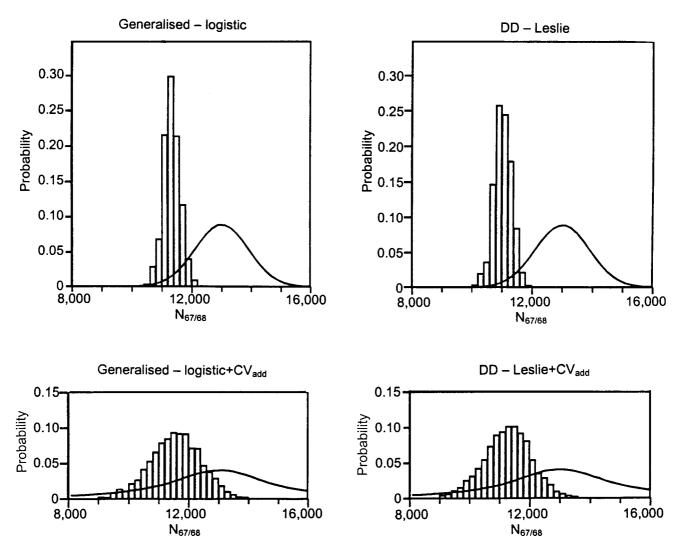

Fig. 2. Posterior probability distributions for the population size in 1967/68 ($N_{95/96}$), as approximated by the posterior sample. The prior distributions are represented as lines. For the models without the additional variance term CV_{add}, the prior was based on the abundance estimate from 1967/68 and was a normal distribution with a mean of 13,012 and standard error of 893. For the models with the additional variance term CV_{add}, the prior was a combination of normal distributions, all with a mean of 13,012 but with a variable standard error that depended upon the prior value of CV_{add}, that was used.

impossible to project back to a historic population size and have a trajectory consistent with the recent abundance trend without making a major untestable assumption, such as that commercial harvests were greater than estimated. Therefore, the previous estimates of historic population size may be questionable. By making certain assumptions, Reilly (1981) was able to construct some sensible population trajectories and concluded that a carrying capacity of 24,000 was in best agreement with the available information in his study. Butterworth *et al.* (2002) investigated a broader range of plausible scenarios to also construct a variety of sensible population trajectories and they found that historic population sizes greater than 30,000 were possible. However, the only simulations in their study which approach a historic population size of 60,000 either assume that commercial harvests were five times greater than estimated, or produce trajectories that do not substantially increase from 1967 to 1988. Therefore, a carrying capacity of greater than 70,000 seems unlikely.

Other lines of evidence are consistent with the idea that gray whales are currently close to their carrying capacity. For example, Stoker (1990) concluded that the recent decline of amphipods in one of the major feeding areas of the gray whale could have been caused by gray whale predation. Reilly (1992) described a recent decline in gray whale pregnancy rates in the aboriginal catch data, although he cautions that sampling bias could have produced this result because of the known differences in pregnancy rates in different areas.

The prior distribution for the maximum sustainable yield level (MSYL) was a uniform from $0.5N_{eq}$ to $0.7N_{eq}$. Values were drawn from this distribution and then transformed into the appropriate value for z. This creates a non-uniform prior for z, but MSYL was the parameter of interest and so it was most appropriate to set a uniform distribution for it.

The prior distribution for CV_{add} in each model was a uniform distribution from 0.0 to 0.35. In preliminary analyses this range was found to span the region of posterior probability in all of the analyses. Again, this makes the prior for CV_{add} uninformative in the sense that the specific limits of this prior distribution do not affect the results. Any value lower than 0.0 could have been used for the lower bound and any value higher than 0.35 could have been used for the upper bound without influencing the results. The value of 0.0 seems a sensible lower bound, as it represents the case where there is no additional unexplained variance. Choosing an upper bound to be a value less than 0.35 would influence the resulting posterior distribution. However, there was no *a priori* reason for choosing any particular upper bound, so there was no justifiable reason for choosing any particular upper bound lower than 0.35. Choosing an upper bound greater than 0.35 would not have changed the results, but would have increased the computation time.

Posterior distributions
The re-sampling approach to integrating the solution makes it easy to form posterior distributions for other quantities that are functions of the model parameters. Because the
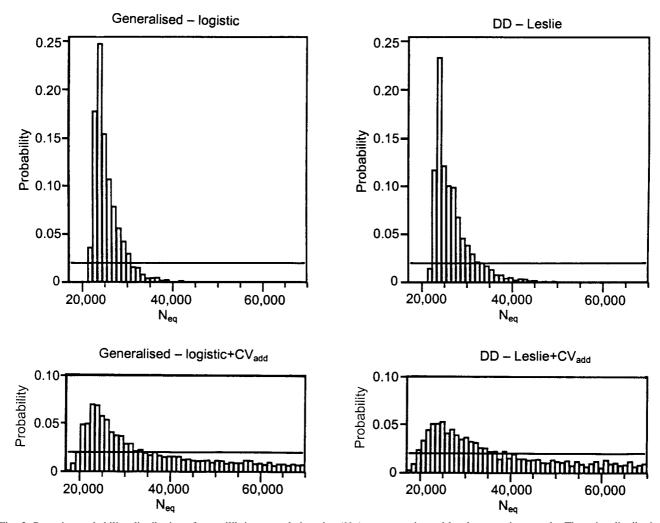

Fig. 3. Posterior probability distributions for equilibrium population size (N_{eq}), as approximated by the posterior sample. The prior distributions (uniform(17000,70000)) are represented as lines.

Generalised – logistic

DD – Leslie

Generalised – logistic+CV$_{add}$
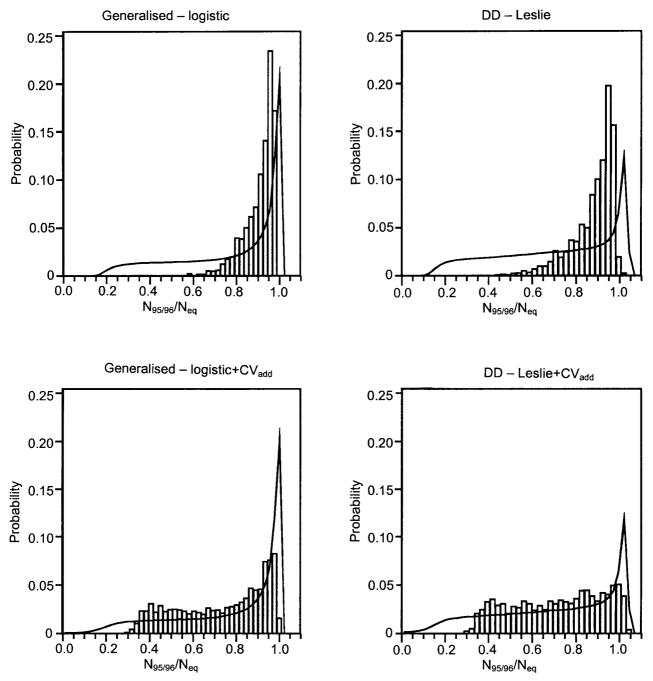

DD – Leslie+CV$_{add}$

Fig. 4. Posterior probability distributions for the ratio of population size in 1996 ($N_{95/96}$) to the equilibrium population size (N_{eq}), as approximated by the posterior sample. The implicit prior distributions are represented as lines.

re-sample of size n_2 is a random sample from the joint posterior distribution, it automatically has imbedded in it any covariances between the parameter estimates. A probability distribution for any function of the parameters is easily approximated from that function calculated from the n_2 re-samples. Therefore, for example, it was possible to directly assess the population status by forming the posterior probability distribution for the ratio of the current population size to equilibrium population size.

Posterior distributions were calculated for several output quantities of interest that were functions of the other parameters. The maximum sustainable yield rate (MSYR) was calculated as the λ_t-1 value associated with the MSYL, defined in terms of the 1+ population. The maximum sustainable yield (MSY) was calculated as the product of MSYR, MSYL and N_{eq} (because MSYL was parameterised as a fraction of N_{eq}). Current replacement yield (RY) was

calculated directly as the model population size in 1996 minus the model population size in 1995. Another catch statistic was calculated, based on the catch control law quantity Q_1 described by Wade and Givens (1997) that was designed to meet the intent of aboriginal whaling management objectives. Q_1 was calculated as 0.9MSY for populations above the MSYL, as the minimum of 0.9MSY and the product N_t*MSYR for populations below the MSYL and as zero for populations below P_{min}, the population size below which no aboriginal catches are allowed. P_{min} was assumed to be 0.1*N_{eq}.

Prior distributions were also calculated for the output quantities of interest. They were simply the distributions of these parameters in the initial sample n_1. They represent the implied prior distributions for these parameters that result from the prior distributions specified for the parameters of the population dynamics models.
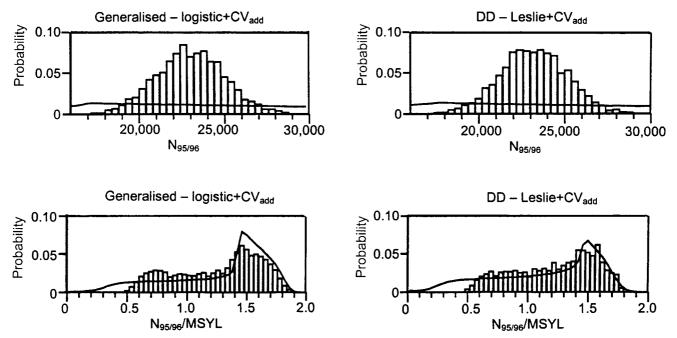
Fig. 5. Upper panel: posterior probability distributions for the population size in 1995/96 ($N_{95/96}$), as approximated by the posterior sample. Note that although there was an abundance estimate in 1995/96, these posterior distributions are different because they are conditioned on the full time series of abundance estimates. Lower panel: posterior probability distributions for the ratio of population size in 1996 ($N_{95/96}$) to the maximum sustainable yield level (MSYL), as approximated by the posterior sample. The implicit prior distributions are represented as lines.

Model comparison

The models were compared through the use of the Bayes factor (Kass and Raftery, 1995). The Bayes factor is defined as the ratio of the probability of the data given by one hypothesis to the probability of the data given by a second hypothesis:

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

This has the form of the ratio of two likelihood functions, but the densities in equation (8) are obtained by integrating, not maximising, across the parameter space and so the Bayes factor is not equivalent to a likelihood ratio test unless there are no unknown parameters. In comparing models, the different hypotheses represent different models and so the Bayes factor is the ratio of the probability of the data given one model to the probability of the data given a different model. The Bayes factor is only dependent upon the evidence provided by the data and is thus not influenced by the prior probability of the hypotheses. When the different hypotheses are given equal prior probability, the Bayes factor is equal to the posterior odds, or the posterior probability of H_1, $pr(H_1|D)$, divided by $1-pr(H_1|D)$.

The Bayes factor is found by calculating the probability of the data given a hypothesis, H:

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

Under a Monte Carlo numerical integration method, such as Sampling-Importance-Resampling, this is estimated as:

$$\tilde{pr}(D|H) = \frac{1}{n_1} \sum_{j=1}^{n_1} pr(D|\theta_i^{(H)}) \qquad (10)$$

where the $\theta_i^{(H)}$s are the n_1 initial samples from the prior distribution. This represents the average likelihood of the sampled parameter values (Kass and Raftery, 1995).

RESULTS

Initial samples in the numerical integration technique ranged from 250,000 to 4,000,000 (Table 3). A re-sample of 5,000 points was drawn in each case from the initial sample, resulting in from 1,275 to 3,905 unique points in the re-sample. The models which did not include the additional variance term required a relatively greater number of iterations because the posterior distributions from those analyses were narrower and thus a random point from the joint prior distribution was less likely to be a point that had any significant probability in the posterior distribution. The initial sample sizes used for the age structured models were large enough to result in sufficiently well determined posterior distributions. Sample sizes larger than necessary were used for the generalised-logistic models because the computational speed of the simpler model allowed it.

The use of the additional variance term was decisively supported by the Bayes factor comparisons (Table 4). A visual explanation for this can be seen in that only one of the confidence limits re-calculated with CV_{add} lies outside the model trajectory (in 1972), while without the use of CV_{add} from 6 to 7 of the confidence limits on the abundance estimate lie outside the model trajectory (Fig. 1). The comparison between the simple model and the age-and-sex-structured model without CV_{add} resulted in

Table 3

Sample sizes and other information from the integration method, the Sampling-Importance-Re-sampling routine. 'n_1' is the size of the initial sample, 'n_2' is the size of the re-sample, 'Unique' is the number of unique values in the re-sample and 'Max' is the maximum number of occurrences of a unique value in the re-sample.

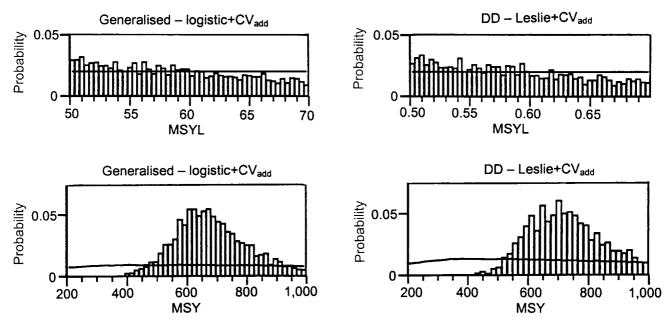|	n_1	n_2	Unique	Max
Generalised-logistic	4,000,000	5,000	1,810	47
Generalised-logistic+CV_{add}	1,000,000	5,000	3,905	6
Density-dependent Leslie	2,000,000	5,000	1,275	172
Density-dependent Leslie+CV	500,000	5,000	3,386	8

Fig. 6. Upper panel: posterior probability distributions for MSYL, the maximum sustainable yield level, as approximated by the posterior sample. The prior distributions (uniform(0.50,0.70)) are represented by lines. Lower panel: posterior probability distributions for MSY, the maximum sustainable yield, as approximated by the posterior sample. The implicit prior distributions are represented as lines.

Table 4

Calculated Bayes factors for pair-wise model comparisons. Evidence categories are modifications of the original categories of Jeffreys (1961), as presented in Kass and Raftery (1995): >150 is decisive evidence, 12-150 is strong evidence, 3-12 is positive evidence and 1-3 is evidence not worth more than a bare mention.

Generalised-logistic with CV_{add} *vs* w/out	0.504×10^{23}	Decisive
DD-Leslie with CV_{add} *vs* w/out	0.186×10^{24}	Decisive
DD-Leslie *vs* Generalised-logistic	3.72	Positive
DD-Leslie with CV_{add} *vs* Generalised-logistic with CV_{add}	1.01	Not worth more than a bare mention

marginal evidence in favour of the simple model (Table 4). However, when CV_{add} was used, there was no difference between the fit of the simple and age-structured models, as the Bayes factor was close to 1.0.

One of the effects of the additional variance term can be seen in the prior distribution for $N_{67/68}$, which was much broader and flatter with CV_{add} than without (Fig. 2). The posterior distribution was broader also, which gave an indication of how much less certain the trajectory of the population was when CV_{add} was used.

Both analyses without CV_{add} estimate N_{eq} to be within the range 22,000-39,000 and it can be seen that the posterior distribution falls off to zero probability quickly outside this range (Fig. 3). However, the analyses with CV_{add} provide a lower bound for N_{eq} but not an upper bound. The posterior distributions still have some posterior probability even at the highest value specified in the prior, 70,000. In spite of this large difference in the width of the posterior distributions, the values of N_{eq} with the highest probability (the modes of the distributions) are all fairly similar, in the neighbourhood of 23,000 (Fig. 3). Even the point estimates do not differ too greatly, as the posterior medians of approximately 25,000 increase to about 30,000-32,000 (Table 5).

The point estimates of current depletion level ($N_{95/96}/N_{eq}$) were all greater than 70% of equilibrium population size, indicating the population is close to its equilibrium level (Table 5b). However, the lower tail of the posterior

distribution extends as low as 30% of N_{eq} when CV_{add} was used in the model (Fig. 4), so there is still some probability that the population is not yet above 50% of N_{eq} (e.g. 0.21 for the density-dependent Leslie matrix model). The specified models do not allow for the model populations to 'overshoot' the equilibrium value by very much given the specified parameter values, which explains the peak at a value of 1.0.

The 2.5th percentiles for the posterior distributions for $N_{95/96}$ ranged from about 19,000-27,000 with CV_{add} (Table 5b, Fig. 5). This range was narrower than the re-calculated confidence limits (approximately 17,000-29,000) on the abundance estimate for 1995/96 (Fig. 1), meaning that conditioning on the entire time-series improved the estimate of current abundance, but not by a large amount.

Without CV_{add} there was a probability of 1.0 that the population was above MSYL, but under the models using CV_{add} there was substantial probability that the population was still below MSYL (Fig. 5). For example, the probability that the population was still below its maximum sustainable yield level was estimated to be 0.28 for the density-dependent Leslie matrix model. As in the previous plot, there was a peak in each distribution around the values representing the ratio of population sizes close to N_{eq} to MSYL.

The posterior distribution for MSYL was nearly identical to the prior distribution (Fig. 6). The posterior distributions for the quantities MSY (Fig. 6) and Q_1 (Fig. 7) were similar and fairly normally distributed, with point estimates ranging from 592-728 (Table 5c). In each case, the distributions were much broader and dispersed for the models with CV_{add}. The 2.5th percentiles of the posterior distributions were 407-564 for Q_1 and 468-627 for MSY (Table 5).

The widest posterior distributions for RY range from about 50 to about 1,000 animals (Fig. 7). The low numbers were a result of trajectories that closely approach equilibrium, as the population growth rate approaches zero at this level and thus RY approaches zero. The high numbers represent trajectories far from equilibrium that thus have growth rates on the order of 3-4% applied to a current population size (e.g. 0.35*23,000 = 805).
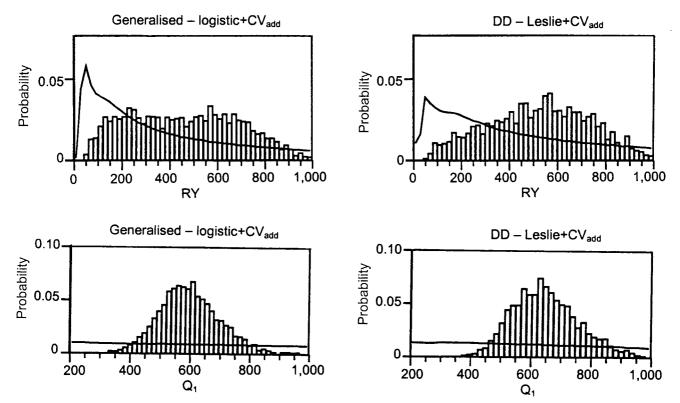
Fig. 7. Upper panel: posterior probability distributions for RY, the current replacement yield, as approximated by the posterior sample. Lower panel: posterior probability distributions for Q_1, as defined in Wade and Givens (1997), as approximated by the posterior sample. The implicit prior distributions are represented as lines.
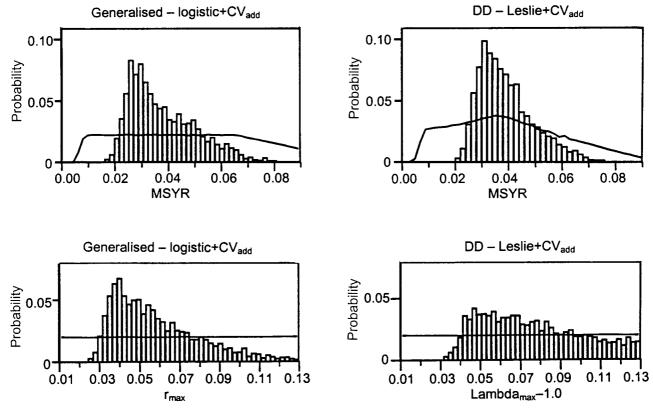

Fig. 8. Upper panel: posterior probability distribution for MSYR, the maximum sustainable yield rate, as approximated by the posterior sample. The implicit prior distributions are represented as lines. Lower panel: posterior probability distribution for r_{max} and λ_{max}-1, the population growth parameters, as approximated by the posterior sample. The prior distributions (uniform(0.01,0.13)) are represented as lines.
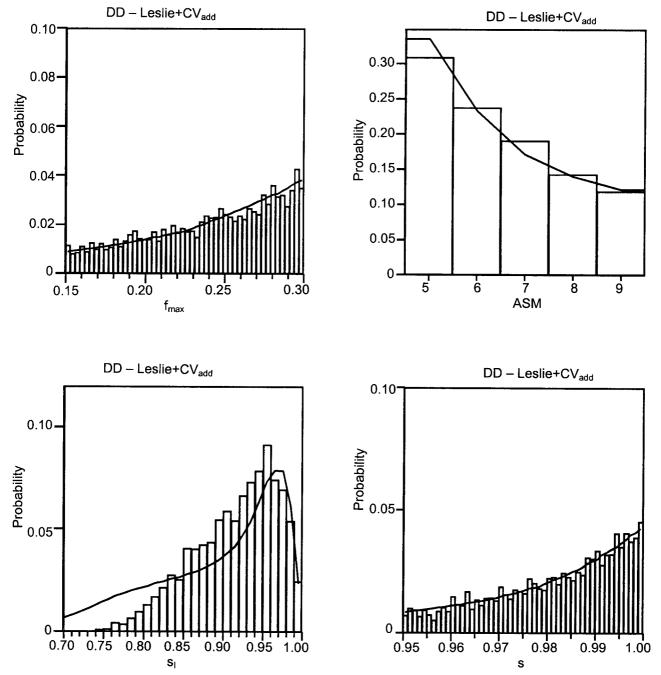
Fig. 9. Posterior probability distributions for four of the parameters of the density-dependent Leslie matrix model with the additional variance term CV_{add}. f_{max} is the maximum fecundity, ASM is the age of sexual maturity, s_j is the juvenile survival rate and s is the adult survival rate. The prior distributions for f_{max}, ASM and s were initially set to uniform distributions over the ranges shown, while the only prior restriction put on s_j was that it be less than s. The lines represent the prior distributions that resulted from forcing the Leslie matrix parameters to be consistent with a uniform prior from 0.1 to 0.13 for λ_{max}-1.

The posterior distributions for MSYR were all greater than 0.02 (Fig. 8), with point estimates ranging from 0.035 to 0.047 (Table 5). This resulted because of the increase in population size over the last 28 years in combination with the low probability that the population is at a small fraction of its equilibrium population size. The point estimate for r_{max} was 0.053 and for λ_{max}-1 was 0.072 (Table 5), with 2.5[th] percentiles of 0.031 and 0.039 respectively. The posterior distributions for r_{max} and λ_{max}-1 both had long tails on the right (Fig. 8), which caused the median values to be greater than values of the modes of the distributions.

The posterior distributions for the four life-history parameters were virtually identical to their prior distributions. The sloped prior distributions with higher probability at values that lead to higher growth rates were due to the fact that λ_{max}-1 itself was forced to have a uniform prior distribution.

DISCUSSION

The model comparisons through use of the Bayes factor led to the conclusion that an additional variance term should be included in population dynamics models fitted to the gray whale abundance data. In other words, it is clear that not all of the variance associated with the abundance estimate has

Table 5

A summary of the posterior distributions of the output quantities, including the posterior median and the 2.5th and 97.5th percentiles of the posterior sample.

(a) Population model parameters. CV_{add} is the additional variance term, r_{max} and λ_{max} are the maximum net recruitment rates, MSYR is the net recruitment rate at the maximum sustained yield level, f_{max} is the maximum fecundity rate, ASM is the age of sexual maturity, s is the adult survival rate, s_j is the juvenile survival rate. All quantities are for the 1+ population.

Model	Parameter	Posterior median	2.5th percentile	97.5th percentile
Generalised-logistic+CV_{add}	CV_{add}	0.140	0.095	0.217
Density-dependent Leslie+CV_{add}	CV_{add}	0.142	0.098	0.214
Generalised-logistic	r_{max}	0.055	0.045	0.087
Generalised-logistic+CV_{add}	r_{max}	0.053	0.031	0.113
Density-dependent Leslie	λ_{max-1}	0.072	0.053	0.128
Density-dependent Leslie+CV_{add}	λ_{max-1}	0.072	0.039	0.126
Generalised-logistic	MSYR	0.042	0.033	0.050
Density-dependent Leslie	MSYR	0.045	0.033	0.054
Generalised-logistic+CV_{add}	MSYR	0.035	0.021	0.068
Density-dependent Leslie+CV_{add}	MSYR	0.037	0.025	0.066
Density-dependent Leslie	f_{max}	0.244	0.155	0.297
Density-dependent Leslie+CV_{add}	f_{max}	0.248	0.158	0.298
Density-dependent Leslie	ASM	6	5	9
Density-dependent Leslie+CV_{add}	ASM	6	5	9
Density-dependent Leslie	s	0.982	0.952	0.999
Density-dependent Leslie+CV_{add}	s	0.985	0.953	0.999
Density-dependent Leslie	s_j	0.950	0.845	0.992
Density-dependent Leslie+CV_{add}	s_j	0.925	0.800	0.990

(b) Population size and depletion level. $N_{95/96}$ is the population size in the beginning of 1996, N_{eq} is the equilibrium population size, $N_{95/96}/N_{eq}$ is the population size in 1996 relative to the equilibrium population size, N_{95-96}/MSYL is the population size in 1996 relative to the maximum sustained yield level. All quantities are for the 1+ population.

Model	Parameter	Posterior median	2.5th percentile	97.5th percentile
Generalised-logistic	$N_{95/96}$	22,950	21,560	24,730
Density-dependent Leslie	$N_{95/96}$	23,290	21,870	24,990
Generalised-logistic+CV_{add}	$N_{95/96}$	22,990	19,080	27,070
Density-dependent Leslie+CV_{add}	$N_{95/96}$	23,200	19,280	27,090
Generalised-logistic	N_{eq}	24,640	22,160	34,250
Density-dependent Leslie	N_{eq}	25,590	22,630	41,140
Generalised-logistic+CV_{add}	N_{eq}	29,880	19,980	66,720
Density-dependent Leslie+CV_{add}	N_{eq}	31,840	19,890	66,840
Generalised-logistic	$N_{95/96}/N_{eq}$	0.933	0.711	0.981
Density-dependent Leslie	$N_{95/96}/N_{eq}$	0.912	0.606	0.980
Generalised-logistic+CV_{add}	$N_{95/96}/N_{eq}$	0.781	0.359	0.988
Density-dependent Leslie+CV_{add}	$N_{95/96}/N_{eq}$	0.734	0.356	1.018
Generalised-logistic	N_{95-96}/MSYL	1.437	1.291	1.660
Density-dependent Leslie	N_{95-96}/MSYL	1.434	1.122	1.706
Generalised-logistic+CV_{add}	N_{95-96}/MSYL	1.361	0.599	1.776
Density-dependent Leslie+CV_{add}	N_{95-96}/MSYL	1.293	0.600	1.725

(c) Catch limit quantities and related information. Q_1, is defined in Wade and Givens (1997), RY is the replacement yield in 1996 and MSY is the maximum sustained yield. All quantities are for the 1+ population.

Model	Parameter	Posterior median	2.5th percentile	97.5th percentile
Generalised-logistic	Q_1	592	532	660
Density-dependent Leslie	Q_1	647	563	732
Generalised-logistic+CV_{add}	Q_1	593	407	841
Density-dependent Leslie+CV_{add}	Q_1	637	460	907
Generalised-logistic	RY	285	136	567
Density-dependent Leslie	RY	391	219	671
Generalised-logistic+CV_{add}	RY	461	91	910
Density-dependent Leslie+CV_{add}	RY	533	118	947
Generalised-logistic	MSY	658	591	733
Density-dependent Leslie	MSY	719	626	813
Generalised-logistic+CV_{add}	MSY	678	468	1,122
Density-dependent Leslie+CV_{add}	MSY	728	524	1,284

been included in previous estimates. A similar conclusion was reached by the IWC Scientific Committee in 1994, where it was noted that 'sampling variability was either under-estimated or was not the only source of variation in the estimates' (IWC, 1995). Results are reported here for all four combinations of population model and use or not of CV_{add}, but the model comparisons clearly indicate that the two models that incorporate the additional variance term, CV_{add}, provide a better fit to the data. A robust management strategy should therefore be based on a model assessment including this term. Including the term CV_{add}, has a large influence on the results, as without it the evidence is very strong that the population is at a large fraction of its equilibrium and is no longer depleted. In general the same conclusion is suggested from the analyses with CV_{add}, but those analyses leave some possibility that the population is still currently depleted.

The lower percentiles of Q_1 were lower in the analyses including CV_{add}, but these lower values were clearly supported by the data and thus represent the preferred quantities on which to base management. It can be seen that the estimated distributions for Q_1 were very close to what a distribution of 0.9MSY would be. This resulted because of the high probability that the population was above MSYL, with the only difference due to the small fraction of trajectories which were still below MSY in 1996. Q_1 represents a harvest level that would allow a depleted population to increase and allows it to increase faster the more depleted it is (Wade and Givens, 1997). An estimate of about 600 is thus consistent with the evidence that the population has increased substantially under a harvest of approximately 174 per year. Note that because there is no probability that the current depletion level is less than 0.25 (Fig. 4), a value for P_{min} as high as 0.25 could have been used without changing the estimates of Q_1.

Another way of putting a potential harvest of 600 animals into context is to consider the available time-series of abundance estimates, which indicates that the population increased by approximately 3% per year from 1967/68 to 1987/88 (Reilly, 1992) while an annual harvest of approximately 174 whales took place. If the population was at approximately 10,000 animals in 1968 and increased at 3% per year for 20 years, it would have increased by an average of about 400 animals per year. Given that the harvest was about 174 per year, this indicates that the population, on average, produced about 575 more whales each year than would die from natural mortality. This can give some indication of what the population could currently sustain given the evidence of the last 28 years.

It should be noted that a harvest level of MSY (or even 0.9MSY) will cause a population well above MSYL (such as one close to carrying capacity) to decline. Because the gray whale is estimated to be well above MSYL, this suggests that a current harvest as high as the point estimates for Q_1 (593-646 using CV_{add}) would cause the population to decline. If continued, in theory such a harvest would cause the population to decline to a level just above the MSYL. Such a decline would be consistent with the specified aboriginal management principles.

When the gray whale population was clearly thought to be increasing (such as from 1968-1988), RY was a useful quantity to represent the maximum harvest that the population might currently sustain. Now that the recent abundance estimates indicate there is some likelihood that the population growth rate is slowing as the population approaches its equilibrium, the quantity RY becomes less useful for management. Obviously, if the population stabilises, RY will go to zero. Therefore, it now becomes important to use other quantities when assessing whether a specified harvest level would meet the requirement of aboriginal subsistence management.

As suggested in Kass and Raftery (1995), the posterior distributions for the quantities of interest could be combined across models, using the probability of each model as a weight for each posterior distribution. In the case here, this would probably change the results little, as such a combined distribution would be dominated by the models including CV_{add}, which give fairly similar results. The nearly equal probability of the generalised-logistic+ CV_{add}, model and the density-dependent Leslie matrix+ CV_{add}, model indicates that a combined posterior would be approximately an equal mixture of those two posterior distributions.

In conclusion, this method for accounting for additional variance results in an analysis that reconciles the contradiction inherent in sequential abundance estimates that differ more than expected according to their estimated variances. The evidence, as expressed by the Bayes factor, supports the use of a parameter representing unexplained variance for a robust stock assessment of the eastern Pacific gray whale population.

REFERENCES

Blokhin, S.A. 1995. Results of research on gray whales caught off the Chukotka Peninsula in 1994. Paper SC/47/AS20 presented to the IWC Scientific Committee, May 1995 (unpublished). 4pp. [Paper available from the Office of this Journal].

Borodin, R.G. 1996. Some data on Russian whaling for gray whales in 1995. Paper SC/48/AS22 presented to IWC Scientific Committee, June 1996 (unpublished). 2pp. [Paper available from the Office of this Journal].

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. and Breiwick, J.M. 2002. Estimated trends in abundance of eastern Pacific gray whales from shore counts (1967/68 to 1995/96). *J. Cetacean Res. Manage.* 4(1):41-8.

Butterworth, D.S., De Oliveira, J.A.A. and Cochrane, K.L. 1993. Current initiatives in refining the management procedure for the South African anchovy resource. pp. 439-11. *In:* G. Kruse, D.M. Eggers, R.J. Marasco, C. Pautzke and T.J. Quinn II (eds.) *Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations.* University of Alaska, Fairbanks. Alaska Sea Grant College program Report No. 93-02. [Available from University of Alaska, Anchorage Library, 3211 Providence Drive, Anchorage, AK99508].

Butterworth, D.S., Korrûbel, J.L. and Punt, A.E. 2002. What is needed to make a simple density-dependent response population model consistent with data for eastern North Pacific gray whales? *J. Cetacean Res. Manage.* 4(1):63-76.

de la Mare, W.K. 1986. Fitting population models to time series of abundance data. *Rep. int. Whal. Commn* 36:399-418.

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.

Henderson, D.A. 1972. *Men and Whales at Scammon's Lagoon.* Baja California Travels Series No. 29. Dawson's Book Shop, Los Angeles. 313pp.

Hobbs, R.C., Rugh, D.J., Waite, J.M., Breiwick, J.M. and DeMaster, D.P. 1996. Preliminary estimate of the abundance of gray whales in the 1995/96 southbound migration. Paper SC/48/AS9 presented to the IWC Scientific Committee, June 1996, Aberdeen (unpublished). [Paper available from the Office of this Journal].

International Whaling Commission. 1993. Report of the Special Meeting of the Scientific Committee on the Assessment of Gray Whales, Seattle, 23-27 April 1990. *Rep. int. Whal. Commn* 43:241-59.

International Whaling Commission. 1995. Report of the Scientific Committee. *Rep. int. Whal. Commn* 45:53-103.

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.

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

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

Laake, J.L., Rugh, D.J., Lerczak, J.A. and Buckland, S.T. 1994. Preliminary estimates of population size of gray whales from the 1992/93 and 1993/94 shore-based surveys. Paper SC/46/AS7 presented to the IWC Scientific Committee, May 1994 (unpublished). 13pp. [Available from the Office of this Journal].

Lankester, K. and Beddington, J.R. 1986. An age structured population model applied to the gray whale (*Eschrichtius robustus*). *Rep. int. Whal. Commn* 36:353-8.

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.

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

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.

Reilly, S.B. 1981. Population assessment and population dynamics of the California gray whale (*Eschrichtius robustus*). Ph.D. Thesis, University of Washington. 265pp.

Reilly, S.B. 1992. Population biology and status of eastern Pacific gray whales: recent developments. pp. 1,062-74. *In:* D.R. McCullough and R.H. Barrett (eds.) *Wildlife 2001: Populations.* Elsevier Applied Science Publishers, London. xv+1,163pp.

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

Smith, A.F.M. and Gelfand, A.E. 1992. Bayesian statistics without tears: a sampling-resampling perspective. *Am. Stat.* 46(2):84-8.

Stoker, S.W. 1990. Distribution and carrying capacity of gray whale food resources in the northern Bering and Chukchi Seas. Paper SC/A90/G13 presented to the IWC Scientific Committee Special Meeting on the Assessment of Gray Whales, Seattle, April 1990 (unpublished). 10pp. [Paper available from the Office of this Journal].

Taylor, B., Wade, P.R., Stehn, R. and Cochrane, J. 1996. A Bayesian approach to classification criteria for spectacled eiders. *Ecol. Appl.* 6(4):1077-89.

Wade, P.R. 1994. Abundance and population dynamics of two eastern Pacific dolphins, *Stenella attenuata* and *Stenella longirostris orientalis.* Ph.D. Thesis, Scripps Institution of Oceanography, University of California, San Diego. xvi+255pp.

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* 47:871-4.