Capture-recapture estimation of bowhead whale population size using photo-identification data

Cibele Q. da Silva[*], Judith Zeh[+], David Madigan[†], Jeff Laake[‡], Dave Rugh[‡], Lisa Baraff[‡], William Koski[#] and Gary Miller[#]

Contact e-mail: zeh@stat.washington.edu

ABSTRACT

Statistical models and maximum likelihood methods are developed for estimating bowhead whale population size from photo-identification data. These are tested on both simulated data and actual data from 1985 and 1986 photographic studies. Initially a multinomial model that accounts for unmarked whales is used. Variance is estimated using the parametric bootstrap. In the cases considered, the variance estimators perform similarly to previously used delta method based estimators in terms of confidence interval coverage, as long as log-normal rather than symmetric confidence intervals are used for the latter. Further models are developed to account for heterogeneity in capture probabilities (highly marked whales are more likely to be captured than moderately marked) and non-random sampling caused by age segregation. These models, particularly the latter, perform better than the multinomial model on simulated data that incorporate these violations of standard capture-recapture assumptions. All three models are applied to actual bowhead whale data. The resulting estimates of the 1+ population size (animals 1 year old or older) in 1985-86 range from 4,719 (using the non-random sampling model on the small dataset in which lengths are available for all whales so that age class can be determined) to 7,331 (using the heterogeneity model on the full dataset). Standard errors are comparable to those obtained from the ice-based census in years with sub-optimal environmental conditions. All confidence intervals include the ice-based census estimates for 1985 and 1986, as well as the corresponding values of 1+ population size in the most likely trajectory from a Bayesian synthesis analysis. These most likely values – 6,649 and 6,820 – incorporate the ice-based census estimates and additional data on bowhead whale population dynamics.

KEYWORDS: ABUNDANCE ESTIMATE; ARCTIC; BOWHEAD WHALE; MARK-RECAPTURE; PHOTO-ID

1. INTRODUCTION

Most estimates of the size of the Bering-Chukchi-Beaufort Seas stock of bowhead whales (*Balaena mysticetus*) have been based on the ice-based visual and acoustic counts of whales conducted at Point Barrow during the spring migration. These population estimates have formed the basis of management advice by the International Whaling Commission (IWC) Scientific Committee (e.g. IWC, 2000). However, at least since Rugh (1990) made the first attempt to compute a population estimate from photo-identification data, researchers involved in aerial photography of bowhead whales have been interested in obtaining an independent population size estimate from such data.

This paper describes three capture-recapture models developed to permit bowhead population size estimation from aerial photographs and presents results obtained using those models on real and simulated bowhead data. The models address problems caused by unmarked whales, heterogeneity in capture probabilities (highly marked whales are more likely to be captured in good photographs than whales that are only moderately marked) and non-random sampling resulting from age segregation. The study arose out of difficulties encountered when applying existing approaches to bowhead photo-identification data, and thus the data themselves are described below.

The actual bowhead data

Aerial photographs of bowhead whales suitable for identification of individuals using their natural markings have been collected in the Bering, Chukchi and Beaufort Seas since 1976. Most of the photographs have been collected by LGL Ltd. (LGL), the National Marine Mammal Laboratory (NMML) and Cascadia Research Collective (CRC). The collections are housed at LGL and NMML. Rugh *et al.* (1992) described how the photographs are taken, summarised the methods used initially for organising and scoring the collection for photo-identification and explained how individuals are identified. The use of capture-recapture techniques to estimate whale population parameters from these data was not envisaged in the early years of the studies, so a single score that combined quality and identifiability was assigned to each photograph.

Inadequate quality screening of photographs can lead to violations of assumptions required for capture-recapture estimation. For example, Hammond (1986), Hammond *et al.* (1990) and subsequent authors have recognised that if photographs are included in a sample on the basis of identifiability rather than photographic quality, heterogeneity of capture probabilities is inevitable, with well-marked whales more likely to be included in a sample. This violates a basic assumption of many capture-recapture analyses. The assumption that all marks are reported on recovery is a particular problem with aerial photographs of bowhead whales, because not all of the whale (from the tip of the rostrum to the tip of the tail) is visible in the majority of such photographs. This is because of water depth as a whale dives, splash, whale motion, sea ice, glare or mud on the whale, each potentially obscuring parts of a whale. Marks on a particular part of a whale cannot be reported on recovery of the whale in a subsequent photograph if the part of the

[*] *UFMG-ICEx-Dep. Estatistica, Caixa Postal 702, 30161-970 – Belo Horizonte-MG, Brazil.*
[+] *University of Washington, Department of Statistics, Box 354322, Seattle, WA 98195-4322, USA.*
[†] *Soliloquy, Inc., 255 Park Avenue S., New York, NY 10010, USA.*
[‡] *National Marine Mammal Laboratory, Alaska Fisheries Science Center, NMFS, NOAA, 7600 Sand Pt Way NE, Bin C15700, Seattle, WA 98115-0070, USA.*
[#] *LGL Limited, Environmental Research Associates, 22 Fisher St, PO Box 280, King City, Ontario, Canada L7B 1A6.*

whale that has the marks is not visible. In addition, it is possible that a whale would not be recognised because of changes in marks over time. Based on examinations of photographs of the same whales taken many years apart (Miller *et al.*, 1992), it is considered unlikely that large scars disappear. However, smaller marks may be disguised by new marks, and they are also more likely than large marks or groups of marks to be obscured in a photograph.

Another problem of greater importance for the bowhead whale than for several other whale species is the large number of unmarked whales in the population. Many young bowhead whales, in particular, are a uniform black colour and have not yet acquired distinctive scars that would permit them to be re-identified in aerial photographs. Because the majority of bowhead whales are not well enough marked to be re-identified, they cannot be considered 'marked' animals and therefore cannot be 'recaptured' (i.e. re-identified) even if they are photographed at a later time. Rugh *et al.* (1998) recognised that unmarked bowhead whales must be accounted for in any attempt to estimate population size using photo-identification data and capture-recapture techniques.

Rugh *et al.* (1998) developed a revised scoring system for aerial photographs of bowhead whales that addressed these problems. Because the largest available samples of photographs were collected in 1985, they suggested basing a population estimate on the 1985 data, with the photographs collected near Point Barrow during the spring migration providing the initial captures and the photographs taken a few months later in the Beaufort Sea providing the recaptures. However, only a few whales identified in the spring 1985 sample were recaptured in the summer sample, so it was clear that two years of data would need to be used to obtain a reliable estimate. Rugh (1990) and Whitcher *et al.* (1996) both noted that the 1986 samples also provided usable data. The 1985 and 1986 photographs, re-scored using the revised system, provided the data for the estimates of Section 5.

Only photographs in which the mid-back region was of good quality were used, so that whales with identifying marks in that region would be recognised when they were photographed on more than one sampling occasion. A whale had to be at least moderately marked on the mid-back to be treated as marked in the analyses; others were treated as unmarked even if they had identifying marks on other parts of their bodies. The scoring system has sufficiently stringent requirements for categorising a whale as moderately marked to ensure that a whale photographed on one occasion will be recognised if photographed again on a subsequent occasion, even if some changes in markings occur in between, provided that the photographs are of good quality. Three of us (WK, GM, DR) had to agree that two photographs were of the same whale to call the second a recapture, virtually eliminating the possibility of false recaptures.

Simulated dataset

Simulated data were used to develop and test the models and methods. A great deal is known about the bowhead whale population, making realistic simulations possible.

A total of four sampling occasions were considered in the simulation; two intra-year occasions (spring and summer) in 1985 and 1986. The output of the simulation includes: the capture history of the marked whales; the total number of photographs of both marked and unmarked whales at each sampling occasion; and the number of captured marked whales at each occasion. For the intra-year occasions the population is considered closed. However, inter-year additions and deletions were allowed for. Two kinds of deletions were considered: those caused by aboriginal hunting and natural mortalities.

Values of parameters used in the simulations were gathered from several sources, including papers and monographs describing the photographic surveys conducted by NMML and LGL, results of the scoring test described by Rugh *et al.* (1998) and papers on bowhead population size and dynamics. Important sources of information on characteristics (e.g. age composition) of the bowhead population are summarised by Givens (1993) and Raftery *et al.* (1995), who present a Bayesian synthesis approach for making inferences about characteristics of interest given different sources of information which are linked by a deterministic population dynamics model. A number of the parameters used here came from the most likely population trajectory in a Bayesian synthesis analysis carried out by Givens (pers. comm.).

According to the most likely trajectory, the age structure of the population (average 1978-1992) is 42% mature, 53% immature (aged 1-17 years) and 5% calves. Calves are not considered here because they have no identifying markings and are thus never part of the marked population. They are also excluded from the real data. Thus the population size estimates are for the 1+ population, i.e. age ≥ 1 year, the population generally considered in population dynamics models used by the IWC Scientific Committee(e.g. IWC, 2000). According to the most likely trajectory, the 1+ population sizes for the years 1985 and 1986 were 6,649 and 6,820 individuals respectively. Survival rates were based on the input parameters that produced the most likely trajectory: 0.9445 for the youngest 20% of the immature whales and 0.9853 for the rest. Natural deaths were determined by these rates, and 19 individuals killed in the subsistence harvest between the summer 1985 and spring 1986 samples were accounted for. Additions to the population, assumed to occur among the unmarked immature whales, were determined as the number needed to obtain the 1986 population size given the number of deaths.

Information about proportions of marked and unmarked bowhead whales was extracted from the datasets used in the evaluation of the new scoring system (Rugh *et al.*, 1998). The datasets contained information about photograph quality and identifiability for several regions of the whale, including rostrum, mid-back, lower-back and fluke. Identifiability is scored as **H**+, **H**-, **M**+, **M**-, **U**+, **U**-, and **X** and constituted the information used to estimate the proportions of highly (**H**) and moderately marked (**M**) bowhead whales as well as the proportion of the unmarked whales (**U**) in the population. The notation **X** stands for the photographs whose quality is so poor that it is impossible to determine whether the whale is marked. Quality is scored on a five-point scale (1+, 1-, 2+, 2-, 3), indicating how much of the area is visible: 1+ represents the highest and 3 the lowest quality. Only the mid-back region is considered here since Rugh *et al.* (1998) found that it had the largest number of marked whales in photographs of good quality (2- or better). The values used to generate the simulated data were 72.8% unmarked, 18.6% moderately marked and 8.6% highly marked whales in the 76.7% of the photographs of the mid-back region assumed to be of good quality. Because the photographs used by Rugh *et al.* (1998) did not include those from 1985 and 1986, the older dataset used by Whitcher *et al.* (1996) was used to determine the number of individuals photographed at each sampling occasion for the simulations. In the case of bowhead whales, it is not possible to know how many individuals were photographed, since the unmarked individuals cannot be recognised. For simulation purposes, these numbers were determined as the total number of good photographs taken at each occasion divided by the number of good quality photographs per marked individual. The resulting number of individuals photographed ranged from a low of 401 in spring 1986 to a high of 1,069 in summer 1985.

The simulated data for some of the models used here must specify the number of good photographs taken of each marked whale captured. These numbers were generated from a zero truncated Poisson distribution; see da Silva (1999) for details.

2. ACCOUNTING FOR UNMARKED WHALES: MULTINOMIAL MODEL

Background

Since the majority of bowhead whales are unmarked and therefore un-catchable using photo-identification techniques, it is essential to account for unmarked whales in estimating population size. Some previous work has been done on estimating population size when only part of the population is catchable. Seber (1982, p.72) gave an estimate

$$\hat{N} = \hat{N}^m / \hat{p}^* \tag{1}$$

where $\hat{N}^m$ is the estimated number of individuals in the catchable population and $\hat{p}^*$ is the estimated proportion of the population that is catchable. Using the delta method, he derived a variance expression under the assumption that $\hat{N}^m$ and $\hat{p}^*$ are statistically independent.

Williams *et al*. (1993), working with bottlenosed dolphin photo-identification data, used equation (1) with $\hat{N}^m$ the estimated number of marked individuals in the population and $\hat{p}^*$ the proportion of the photographs that were of marked individuals. Their estimated variance expression

$$V(\hat{N}) = \hat{N}^2 \left(\frac{V(\hat{N}^m)}{(\hat{N}^m)^2} + \frac{1-\hat{p}^*}{n\hat{p}^*} \right) \tag{2}$$

matches that given by Seber (1982) when binomial sampling is used to obtain $\hat{p}^*$; n is the number of photographs on which the estimate $\hat{p}^*$ is based, and $\hat{N}^m$ and $\hat{p}^*$ are used to approximate their expected values. Williams *et al*. (1993) obtained 95% confidence intervals by multiplying the square root of the variance estimate (equation 2) by 1.96.

This approach is simple and intuitively appealing. However, it can be criticised on several grounds. First, Williams *et al*. (1993) used photographs from the same studies to obtain $\hat{N}^m$ and $\hat{p}^*$, so the assumption of statistical independence of these estimates on which the delta method variance is based does not hold. Covariance between the estimates is not taken into account in equation (2). Second, numerous authors (e.g. Burnham *et al*., 1987; Garthwaite and Buckland, 1990; Cormack, 1992) have noted that capture-recapture estimates of population size have a skewed distribution, so the symmetric intervals used by Williams *et al*. (1993) are unsatisfactory. This paper develops alternative interval estimates of population size from photo-identification data when the population includes unmarked animals and this approach is compared with that of Williams *et al*. (1993) using simulated bowhead data.

Darroch's multiple recapture model for closed populations

Here we generalise the multiple recapture model of Darroch (1958) for closed populations. The notation and assumptions of the model used to define our likelihood functions are introduced below.

(1) N is the population size;
(2) t is the number of sampling occasions;
(3) u_i is the number of individuals caught in the i^{th} sample but not otherwise, u_{ij} is the number caught in the i^{th} and j^{th} samples but not otherwise, etc.;

Let w be a subset of the integers 1, …, t and

$$r = \sum_w u_w = \sum_i u_i + \sum_{i<j} u_{ij} + \ldots + u_{12\ldots t}$$

be the total number of different individuals caught in the complete experiment. Let $n_i = \sum_{w \supset i} u_w$ be the size of the i^{th} sample. For example, $n_2 = u_2 + u_{12} + u_{23} + u_{24} + u_{123} + u_{124} + u_{234} + u_{1234}$ if $t = 4$.

The probability distribution of $\{u_w\}$ assumed by Darroch (1958) is a multinomial distribution with parameters N and P_w, where P_w is the probability of an individual with capture history w being caught. Let $p_i = 1\text{-}q_i$ be the probability that any individual is caught in the i^{th} sample. The probability of any individual escaping capture throughout the experiment is $\prod_i q_i = Q$. The probability of being caught in samples i, …, 1 and no others is $\frac{p_i}{q_i} \ldots \frac{p_l}{q_l} Q = P_{i\ldots l}$. Therefore, the probability density of $\{u_w\}$ is multinomial, i.e.

$$P(\{u_w\}) = \frac{N!}{(N-r)! \prod_w u_w!} Q^{N-r} \prod_w P_w^{u_w} \tag{3}$$

where $0 \le u_w \le N$ subject to $\le \sum_w u_w \le N$. Darroch (1958) shows that equation (3) can also be written as:

$$P(\{u_w\}) = \frac{N!}{(N-r)! \prod_w u_w!} \prod_i p_i^{n_i} q_i^{N-n_i} \tag{4}$$

The above development requires several assumptions:

(1) the population is closed, i.e. it remains constant throughout the experiment;
(2) all individuals are equally likely to be members of any given sample, regardless of their previous capture history or of what other individuals are in the sample, although capture probabilities may differ between samples;
(3) all captured animals are marked and are correctly identified on recapture.

Generalisation of Darroch's model to bowhead whales

Let us now consider how to generalise the above model to a situation in which animals are captured and recaptured photographically, natural markings are used to identify individuals on recapture, and some individuals in the population lack identifying markings. Although this section focuses on bowhead whales, the same considerations apply to similar situations involving any species.

As noted in Section 1, it is necessary to use data from two different years for bowhead whales. It is therefore clear that the closed population assumption does not strictly apply since whales are born and die between samples. However, bowhead whales have high survival rates (Whitcher *et al.*, 1996; George *et al.*, 1999) and relatively low fecundity rates (Miller *et al.*, 1992). Therefore, rather than generalising to an open population model, the closed population assumption is retained and simulated data are used to determine whether its failure is problematic in this case.

For bowhead photo-identification, capture probabilities differ between samples because of differences in photographic effort between sampling occasions. There should be no behavioural response to capture; the whales are not trapped, handled, harmed or treated in any sense. Since in the case of the bowhead whales a *capture* means that a good quality photo of a whale was taken, the only source of some behavioural response could be if the animal felt annoyed or threatened by the noise of the aeroplane flying over it. However, no systematic divings have been observed during the photographic sessions. It therefore seems reasonable to assume that an individual's previous capture history should not affect its capture probability on a given sampling occasion. In this section, it is assumed that by restricting consideration to photographs of adequate quality, heterogeneity in capture probabilities between highly and moderately marked whales on a given occasion is avoided. It is further assumed that the capture of a particular individual does not make capture of a different individual more or less likely, since no long-term affiliations have been observed among bowhead whales photographed during the 1981-94 studies (Koski *et al.*, 1988; Richardson *et al.*, 1995). In short, it appears reasonable to assume that the second assumption in the previous section holds.

However, the third assumption clearly does not hold for bowhead whales. Whales that lack natural markings remain unmarked even when they are captured in a photograph and they cannot be identified on recapture. Restricting consideration to photographs of adequate quality makes it possible to assume that all marked animals are correctly identified on recapture. Thus it is reasonable to assume that the assumptions of Darroch's model hold for marked whales, but unmarked whales, the majority of bowhead whales, must be accounted for outside of that model. As in Williams *et al.* (1993), the photographs of the unmarked whales are used to do this.

Let:

- X_i^u equal the number of good photographs of unmarked whales taken at time i;
- X_i^m equal the number of good photographs of marked whales taken at time i;
- X_i equal the total number of good photographs taken at time i;
- n_i^m equal the number of individual marked whales captured at time i;
- r equal the total number of individual marked whales captured over the study; and
- $\{u_w\}$ equal a set which includes the number of individuals with capture history w.

The following relationship is observed:

$$X_i = X_i^u + X_i^m$$

The parameters in the model are:

- $N = N^m + N^u$, the total number of individuals in the population;
- N^u, the total number of unmarked individuals in the population;
- N^m, the total number of marked individuals in the population; and
- p_i, the probability that a given whale is photographed at sampling occasion i.

The distribution of X_i^m is assumed to be binomial with parameters

$$\left(X_i, \frac{N^m}{N} \right)$$

and the distribution of $\{u_w\}$ is multinomial given by Darroch's model (equation 3). The joint distribution of these variables is

$$P(\{u_w\},\{X_i^m\}) = P(\{X_i^m\}|\{u_w\})P(\{u_w\})$$

$$= \left[\prod_{i=1}^{t} P(X_i^m | n_i^m) \right] P(\{u_w\})$$

The distribution of $(X_i^m | n_i^m)$ is truncated binomial because $X_i^m \geq n_i^m$.

Since the estimation of N^u and N^m is restricted by the relationship $N = N^m + N^u$, it is natural to write N^u as being proportional to N^m with proportionality constant γ, say, and write $N^u = \gamma N^m$. Therefore, $N = N^m(1+\gamma)$. Further development of the model is simplified by this relationship. It is implicitly assumed that both N and N^m remain constant over time when γ is treated as a constant. Photographs of the same whale taken many years apart suggest that marks are not acquired frequently, so it is assumed that the closed population model is adequate for the marked as well as the total population over the two-year time period being considered.

The distribution of $(X_i^m | n_i^m)$ is expressed by:

$$P(X_i^m | n_i^m) = \frac{\binom{X_i}{X_i^m} \left[\frac{1}{1+\gamma} \right]^{X_i^m} \left[\frac{\gamma}{1+\gamma} \right]^{X_i^u} I\{(X_i^u, X_i^m) \in B\}}{\sum_{j=n_i^m}^{X_i} \binom{X_i}{j} \left[\frac{1}{1+\gamma} \right]^{j} \left[\frac{\gamma}{1+\gamma} \right]^{X_{i-j}}} \tag{5}$$

where $B = \{(X_i^u, X_i^m) : n_i^m \leq X_i^m \leq X_i\}$ and $I\{(X_i^u, X_i^m) \epsilon B\}$ is an indicator function for pairs (X_i^u, X_i^m) that belong to the set B.

The likelihood function is given by

$$L = \{ \prod_{i=1}^{t} P(X_i^m | n_i^m) \} P(\{u_w\})$$

$$= \prod_{i=1}^{t} \left[\frac{\binom{X_i}{X_i^m} \left[\frac{1}{1+\gamma} \right]^{X_i^m} \left[\frac{\gamma}{1+\gamma} \right]^{X_i^u} I\{(X_i^u, X_i^m) \in B\}}{\sum_{j=n_i^m}^{X_i} \binom{X_i}{j} \left[\frac{1}{1+\gamma} \right]^{j} \left[\frac{\gamma}{1+\gamma} \right]^{X_{i-j}}} \right]$$

$$\times \frac{N^m!}{(N^m - r)! \prod_w u_w!} \prod_{i=1}^{t} p_i^{n_i^m} q_i^{N^m - n_i^m} \tag{6}$$

The maximum likelihood estimators for p_i and N^m are given by:

$$\hat{p}_i = \frac{n_i^m}{\hat{N}^m}; \quad i = 1,...,t \tag{7}$$

$\hat{N}^m$ is obtained by solving the equation

$$\prod_{i=1}^{t}(N^m - n_i^m) = (N^m)^{t-1}(N^m - r) \tag{8}$$

as in Darroch (1958). This equation is solved iteratively using the technique of Robson and Regier (1968). Starting values are based on recommendations by Chapman (1952) or, when there are few recaptures, the estimator of Schnabel (1938); see da Silva (1999) for details. When $n_i^m / X_i \simeq 0$,

$$\hat{\gamma} = \frac{\sum_{i=1}^{t} X_i^u}{\sum_{i=1}^{t} X_i^m} \tag{9}$$

closely approximates the maximum likelihood estimate of γ. In the case of the bowhead whales, this condition is satisfied since large numbers of photographs are taken and the number of marked whales in the sample is small compared to the number of photographs. The adequacy of the approximation was checked by comparing estimates of γ obtained using equation (9) to those obtained by maximising the likelihood function using the NAG library Fortran routine E04KDF (16 AUGUST 1993). Estimates differed only beyond the third decimal place.

The population size estimator obtained from the above equations is the same as that of Williams *et al.* (1993).

Parametric bootstrap standard error

Following Buckland (1980) and others, the parametric bootstrap is used to estimate standard error. Bootstrap methods depend on the notion of a bootstrap sample (Efron and Tibshirani, 1993). If the distribution from which the bootstrap samples are drawn provides a good approximation to the distribution from which the original data were drawn, then the standard deviation of the estimates of the parameter of interest (in this case, population size N) computed from the bootstrap samples will provide a good estimate of the standard error of the parameter estimate (in this case, $\hat{N}$) computed from the original data. When we obtain bootstrap samples by re-sampling the original data, giving each of the original n data points equal weight, we are using the empirical distribution function $\hat{F}_n$ to approximate the true distribution F. Estimates of standard error obtained in this way are called non-parametric bootstrap estimates because $\hat{F}_n$ is the non-parametric estimate of F. The parametric bootstrap uses a different estimate of F. In the parametric bootstrap setting we draw B samples of size n from the distribution $\hat{F}_{par}$, an estimate of F derived from a parametric model for the data. Where parameters are needed to specify the distribution, estimates of these parameters computed from the original data are used.

The choice between non-parametric and parametric bootstrap in capture-recapture is addressed by Buckland and Garthwaite (1991). They note that even though the non-parametric bootstrap is more widely used and more familiar than the parametric bootstrap, the latter allows a choice of which underlying distribution model to assume for the data. Mark-recapture provides an example where the nonparametric bootstrap makes specific parametric assumptions that are not immediately apparent. That may lead a user to bootstrap on the wrong sampling unit, or to conclude erroneously that the results are more robust than those from a parametric approach. In fact, Bickel and Freedman (1981), in examining the theoretical basis for the bootstrap, developed a number of examples in which the nonparametric bootstrap fails to provide a consistent estimate of standard error while the parametric bootstrap succeeds. This occurs when $\hat{F}_n$ provides a poor approximation to F and the probability model used in the parametric bootstrap is correct.

The model presented in expressions (3) and (5) was used in the parametric bootstrap approach used here, which involves the steps given below.

(1) Obtain the ‘original data’ by running the data simulation program once or by using the real bowhead data.
(2) Estimate the parameters $\hat{N}^m$, $\hat{p}_1, \ldots, \hat{p}_4, \hat{\gamma}$ and $\hat{N}$, from the data obtained in step 1.
(3) Using the estimated capture probabilities ($\hat{p}_1, \ldots, \hat{p}_4$ and population size of the marked whales ($\hat{N}^m$), simulate the number of individuals with a given capture history w, u_w^*, by using Darroch’s multinomial model (equation 3). This yields the sample sizes ($n_1^{m*}, \ldots, n_4^{m*}$) and r^* to be used in calculating the estimate N^{m*} and p_i^*’s from the bootstrap sample.
(4) Simulate truncated binomial distributions. The total number of photographs X_i obtained in the ‘data’ at each occasion i was kept fixed at its value in the original data and was divided among marked and unmarked whales as follows:
    (i) A truncated binomial distribution with parameters

$$\left(X_i, \frac{1}{1+\hat{\gamma}}\right)$$

    was simulated to obtain the number of good photographs of marked whales;
    (ii) The number of good photographs of unmarked whales was obtained by subtraction.
    This provides the data needed to calculate the estimate γ^*.
(5) Calculate the parameter estimates, including N^*, from the bootstrap sample.
(6) Repeat steps 3-5 B times.

In the above steps, * denotes data or an estimate from the bootstrap sample.

The standard deviation of N^* over the B bootstrap samples, $s.e.^*$, estimates the standard error of $\hat{N}$. The difference between the mean of the B values N^* and $\hat{N}$, $bias^*$, estimates the bias of $\hat{N}$. Confidence intervals may be found using the percentile method (Efron, 1981; Buckland and Garthwaite, 1991) as follows. Order the N^* from smallest to largest, and denote the ordered list by $\hat{N}_{(j)}$. Approximate 100(1-2α)% confidence limits are then given by $\hat{N}_{(k)}$ and $\hat{N}_{(k')}$, where $k = (B + 1)\alpha$ and $k' = (B + 1)(1 - \alpha)$, both rounded to the nearest integer value.

The determination of the number B of bootstrap replications depends on the application. Efron (1981) suggests that bootstrap estimates of standard error usually have relatively little bias, and that seldom are more than B =

200 replications needed for estimating a standard error. Many more replications are needed to obtain a good estimate of bias or for construction of confidence intervals. The percentile method depends on the tail of the distribution where fewer samples occur. Buckland and Garthwaite (1991) advocate that for a 95% confidence interval B = 1,000 should be satisfactory whilst B = 200 will be inadequate.

The choice of B made here was based on the analysis of changes in coefficient of variation (CV) when 1,000, 2,000 and 3,000 bootstrap replications were drawn. Efron (1981) argued that the increased variability due to stopping after B bootstrap replications, rather than going on to infinity, was reflected in an increased CV. Working with simulations of data under closed and open population assumption, we observed that the CV based on 1,000 bootstrap samples compared to 2,000 and 3,000 presented much more dispersion than when the CV was calculated using 2,000 and 3,000 bootstrap replications. The last cases presented roughly close CVs, but to guarantee accurate results a value B = 3,000 was chosen. This value of B should achieve sufficient precision to estimate bias and obtain reliable confidence intervals.

Comparison of estimation methods using simulated data

Although our estimator $\hat{N}$ is the same as that obtained by Williams *et al*. (1993), the method of estimating its standard error, bias and confidence interval differs. The two approaches are compared below using the simulated data described in Section 1. In addition to considering the symmetric confidence intervals of Williams *et al*. (1993), the delta method variance (equation 2) is also used to compute the confidence intervals suggested by Burnham *et al.* (1987).

According to Burnham *et al*. (1987), the symmetric 95% confidence interval for N

$$\hat{N} \pm 1.96\sqrt{V(\hat{N})}$$

can be improved by using transformations that better approximate normality. They recommend the log-transformation. Transformation makes little difference if the CV of the parameter estimator in question is small, say ≤ 0.1 (10% when expressed as a percentage). It makes a difference at moderate (near 20%) and large ($\geq 40\%$) CV. For an approximate (1 - α)100% log-based confidence interval for a parameter, say N, Burnham *et al.* (1987) recommend the calculation of lower and upper bounds, $\hat{N}_L$ and $\hat{N}_U$, as

$$(\hat{N}_L, \hat{N}_U) = (\hat{N}/C, \hat{N}C) \tag{10}$$

where

$$C = \exp\left(z_{\alpha/2}\sqrt{\log\left(1 + \left[\mathrm{CV}(\hat{N})\right]^2\right)}\right) \tag{11}$$

and $\mathrm{CV}(\hat{N})$ is the estimated standard error of $\hat{N}$ from equation (2) divided by $\hat{N}$.

Some further definitions are needed. In the following expressions, N is the population size assumed in generating the simulated data, means are computed over the s simulated samples and $bias^*$ is computed for each simulated sample as

$$bias^* = \frac{1}{B}\sum_{b=1}^{B} N^* - \hat{N} \tag{12}$$

where the N^* are the estimates of N from the B bootstrap samples and $\hat{N}$ is the estimate of N from the simulated sample. With this notation,

(1) true bias of $\hat{N}$ = mean($\hat{N}$) – N;
(2) bias corrected $\hat{N}$ = uncorrected $\hat{N}$ – $bias^*$;
(3) bias of corrected $\hat{N}$ = mean(bias corrected $\hat{N}$) – N

(4) RMSE $= \sqrt{\frac{1}{s}\sum_{i=1}^{s}(\hat{N} - N)^2}$.

The summary statistics presented below are based on s = 500 simulated samples. For each of them, B = 3,000 parametric bootstrap replications were performed.

Results

In this section the multinomial model is analysed for five different cases in order to gain some insight into how the multinomial model works for varying values of total population size, capture probabilities and population size of unmarked individuals. For all of the cases studied, the number of marked individuals in the population was the same: 1,886 marked individuals. That has the advantage of keeping the number of cases in the study small, without leaving out the most interesting ones. In addition, the impact of departures from the closed population assumption on estimated population size for bowhead whales is investigated.

Firstly, the multinomial model is compared using simulations for closed and open populations. The rest of the population parameters needed in those simulations were kept fixed. Secondly capture probabilities and total population size were varied. A summary of the cases is presented in Table 1. For example, 'case 0' differs from 'case 1' because in the former it was assumed that the population was closed whereas in the latter it was open. The capture probabilities

Table 1

Description of the main parameters used in the simulation of five different cases for capture-recapture estimation of population size - 1,886 marked individuals for each case.

	Population Size		Number of individual photographs			
Case	1985	1986	Spring 1985	Summer 1985	Spring 1986	Summer 1986
0	6734	6734	641	1069	401	441
1	6649	6820	641	1069	401	441
2	6734	6734	2x641	2x1069	2x401	2x441
3	2x6734	2x6734	2x641	2x1069	2x401	2x441
4	2x6734	2x6734	641	1069	401	441

are almost the same for both cases since the same sizes were used in those simulations. 'Case 2' is an ideal case where the number of unmarked individuals in the population is not as bad as for cases 3 and 4, the population is closed and capture probabilities are twice as high as those for cases 0, 1 and 3. Case 4 is expected to give the worst results: large bias and standard deviation for the estimated value of N.

Table 2 contains uncorrected and bias corrected summary statistics for 500 estimated values of N under each case. The first column of Table 2 gives the average of $\hat{N}$ for 500 datasets. The next two columns give true bias and RMSE as defined above. The fourth column gives the average bootstrap standard error over the 500 samples. The last three columns give the statistics on true bias and RMSE for bias corrected $\hat{N}$.

These results show that even when the percentage of unmarked individuals in the population is as high as 72% (cases 0-2) to 86% (cases 3-4), as is the case for the bowhead whale population, the proposed model works well as long as it is possible to draw large samples from the population. As expected, case 2 yields the smallest bias and variance. Cases 0 and 1 show when it is not possible to draw large samples, the estimator still works well, even when the population is not closed. Of course the time period considered here is just two years. Further investigation is required if longer time periods are to be considered. Cases 3 and 4 show that the availability of large samples leads to better estimates. Although 86% of the whales were unmarked in both cases, case 3, with a sample size twice as large as case 4, resulted in more accurate and precise estimates of N. Case 0 contrasted with case 3 reveals that the variance of $\hat{N}$ increases with N and with the number of unmarked individuals in the population; the CV is small in both cases, being only slightly larger in case 3.

Bias is negligible in all cases but case 4. Otis *et al.* (1978) explain that the bias of the estimated value of a population size using Darroch's model is not significant when capture probabilities (p_i's) are, on average, close to 0.1 or larger. However, if the p_i's are smaller than 0.1, significant bias results. Otis *et al.* (1978) showed by simulation that positive bias is observed when capture probabilities are low. Seber (1982, p.72), showed that there is positive bias associated with the correction for unmarked whales when n and $\hat{p}^*$ are small relative to N. Both these problems occur in case 4. Bias correction has a negligible effect in cases 0-3. For case 4, the correction substantially reduces bias and RMSE.

The parametric bootstrap estimates of the variance of $\hat{N}$ are generally larger than the delta method estimates (equation 2). That may be because $\hat{p}^*$ and $\hat{N}^m$ were considered as being uncorrelated in the latter. Also of interest in evaluating the performance of the bootstrap is the study of coverage performance of confidence intervals for N. Percentile parametric bootstrap confidence intervals are compared below with intervals calculated from the delta method variance estimates (equation 2), both the symmetric intervals used by Williams *et al.* (1993) and the log-normal intervals proposed by Burnham *et al.* (1987).

Table 3 shows the percentage of the times that the symmetric, log-normal and percentile confidence intervals (CI) missed the true value of N on the left or right side. For example, *miss left* means that the left endpoint was larger than N, i.e. the population size was overestimated. The desired coverage is 95%, so we expect *miss left* and *miss right* to be roughly 2.5%. Table 3 also gives mean CI widths. A total of 500 samples (500 CI realisations) with 3,000 bootstrap replications for each sample were considered.

Table 3

Percentage of 500 samples in which the confidence interval (CI) missed the true value N. Average CI widths are also given. Symmetric and log-normal intervals use the delta method variance estimate. Percentile parametric bootstrap intervals are based on 3,000 bootstrap replications.

Case	Type of CI	% Miss Left	% Miss Right	Mean CI width
0	Symmetric	0.2	5.8	3,038
	Log-normal	1.2	3.8	2,988
	Percentile bootstrap	0.6	3.0	3,169
1	Symmetric	1.2	5.6	3,106
	Log-normal	2.4	3.8	3,102
	Percentile bootstrap	1.6	3.4	3,256
2	Symmetric	2.0	4.0	1,377
	Log-normal	2.6	3.2	1,351
	Percentile bootstrap	2.6	3.0	1,392
3	Symmetric	1.4	6.4	6,121
	Log-normal	1.8	4.6	5,863
	Percentile bootstrap	2.6	3.4	6,361
4	Symmetric	0.0	6.4	14,490
	Log-normal	2.2	3.4	13,740
	Percentile bootstrap	5.6	1.4	18,490

Overall Table 3 shows that the percentile intervals achieve better balance than the delta method based symmetric intervals on the left and right sides. The symmetric CI overcover on the left and undercover on the right in all cases, and their overall coverage is 94% or less in all cases. The parametric bootstrap comes closest, averaged over cases 0-4, to overall 95% coverage. However, the log-normal intervals deviate less from 95% coverage on average, and they have the shortest mean confidence interval widths in all cases. The percentile bootstrap intervals compare particularly badly with the log-normal intervals (poorer coverage and much larger mean CI width) in case 4, the sparse data case.

Case 4 in Table 2 suggests that bias correction can be important. Improved percentile bootstrap CI could no doubt be calculated, for example by using the bias corrected and

Table 2

Summary statistics for the estimated values of $\hat{N}$ based on 500 samples.

Case	$\hat{N}$ not bias corrected: Mean	True bias	RMSE	$\overline{s.e.}$*	Bias corrected: Mean	Bias	RMSE
0	6,770	36	767	859	6,759	25	769
1	6,812	78	860	917	6,800	66	866
2	6,724	-10	360	355	6,730	-4	358
3	13,530	62	1,699	1,628	13,367	-101	1,652
4	14,410	942	4,689	5,072	13,311	-157	3,545

accelerated percentile method discussed by Efron and Tibshirani (1993). However, the delta method based log-normal CI do quite well overall, and they require much less computing effort than the parametric bootstrap percentile intervals. When the approximation (equation 9) can be shown to be adequate, so that our estimate of N and that of Williams *et al.* (1993) is the same, the delta method variance estimate (equation 2) might be judged to provide adequate CI, as long as log-normal rather than symmetric intervals are used.

3. HETEROGENEITY: HIGHLY VS. MODERATELY MARKED WHALES

Introduction

The capture-recapture model presented in the previous section, while taking account of unmarked whales, assumed that all marked whales had the same capture probability on a given sampling occasion. The goal of the quality scoring proposed by most researchers interested in using photographic data for capture-recapture (for example, Hammond, 1986; Hammond *et al.*, 1990; Friday *et al.*, 2000; Rugh *et al.*, 1998) has been to identify a quality level that is good enough to permit whales in photographs of that quality to be identified regardless of whether they are highly or moderately marked. It was hoped that by restricting capture-recapture analyses to photographs of this quality, heterogeneity in capture probabilities, with highly marked whales more likely to be captured than moderately marked whales, could be avoided. However, this goal has proved elusive. Positive correlations are generally found between quality and distinctiveness scores (Friday *et al.*, 1997), implying that photographs of highly marked whales are more likely to be scored as of good quality.

Consideration of the scoring system developed and tested by Rugh *et al.* (1998) indicates that this positive correlation is inevitable for bowhead whales. Only in photographs of excellent quality can one be certain that marks are not obscured. Thus a highly marked whale may well be scored as moderately marked in a photograph of only good quality. However, as shown below, if capture-recapture analyses are restricted to photographs of excellent quality, the dataset will be too small to provide a useful population estimate.

White *et al.* (1982) pointed out that numerous published studies demonstrate heterogeneity in capture probabilities for a wide range of species. In studies where the true population size was known (for example, Carothers, 1973), the commonly used estimators were biased severely by this heterogeneity. Computer simulation studies have also shown that heterogeneity can cause substantial negative bias in the commonly used estimators of population size (Carothers, 1973; Otis *et al.*, 1978).

Pollock (1991) reviewed efforts to develop models and estimation procedures that can handle heterogeneity without producing biased estimates of population size. A set of models that allow capture probabilities to vary due to heterogeneity (h), trap response (b), time variation (t) (i.e. capture probability for time i differs from that for time j) and all possible two- and three-way combinations of these factors is now available. The eight models (M_o, M_h, M_b, M_{bh}, M_t, M_{th}, M_{tb}, M_{tbh}) were first considered as a set by Pollock (1974) and were more fully developed by Otis *et al.* (1978), White *et al.* (1982), and Pollock and Otto (1983). According to Pollock (1991), models M_{th}, M_{tb} and M_{tbh} do not usually permit estimation of population size due to non-identifiability issues although they are often necessary for real populations. In the case of bowhead whales, for example, time variation t is certain because of widely differing effort on the different sampling occasions.

Pollock *et al.* (1984) introduced a logistic regression technique to account for observable population heterogeneity in capture probabilities. In other words, the characteristics of the captured individuals were used to explain their probabilities of capture. They examined inferences based on the full likelihood, which necessitated the construction of categories of individuals according to the values of the covariates and estimation of the number of individuals in each category. This was necessary to overcome the problem that the covariates for uncaptured individuals were not known. Cormack (1989) also advocated the use of generalised linear models for capture-recapture but noted that, while his approach could diagnose heterogeneity, population size estimation was hampered by the lack of knowledge of the distribution of covariates in the unobserved part of the population.

Huggins (1989; 1991) and Alho (1990) independently suggested the use of a likelihood conditional on the captured individuals. The approach of Huggins is expanded below to develop a model that allows and accounts for heterogeneity in capture probabilities between moderately marked M and highly marked H whales. Although restricting photographs to those of highest quality might prevent heterogeneity, this would waste useful data and reduce precision. This is not necessary if models that allow for heterogeneity are used.

Heterogeneity in capture probability via logistic models

Initially a method to create the likelihood function for the marked individuals only is considered. This will then be combined with the unmarked whales to arrive at an estimate of total population size. The proposed model is based on Huggins (1989; 1991) who introduced a model that explains heterogeneous capture probabilities via observable characteristics of the individuals and time dependence via observable characteristics of the sampling occasions. Individual capture probabilities are expressed by linear logistic models with coefficients assumed to be the same for individuals in the same group as specified by the animals' covariates. That provides the homogeneity assumption needed to enable estimation of the parameters involved in the model. Population size is estimated by the method of moments as a function of the individual estimated values of the capture probabilities.

It is assumed here that the captures of a whale are independent from the previous occasions and that the individuals behave independently. That does not imply that all the whales have the same capture probabilities; this probability is allowed to be a function of the covariates included in the analysis. Under the assumptions, the full likelihood for the marked individuals is:

$$L^* = K\prod_{i=1}^{N^m}\prod_{j=1}^{t} p_{ij}^{\delta ij}(1-p_{ij})^{(1-\delta_{ij})} \tag{13}$$

where

N^m is the total number of marked individuals; and

p_{ij} is probability that animal i is captured at time j, where $i = 1, \ldots, N^m$, and $j = 1, \ldots, t$,

$$\delta_{ij} = \begin{cases} 1, \text{ if individual } i \text{ is captured at time } j, \\ 0, \text{ otherwise} \end{cases}$$

and K may depend on N^m but not on the parameters that define p_{ij}.

One complication with estimation of the parameters using the model above is that the covariates for the uncaptured marked individuals are not known. Huggins (1989) showed that inference can be based on the conditional likelihood,

$$L = \prod_{i=1}^{r}\prod_{j=1}^{t}\gamma_{ij}^{\delta_{ij}}(1-\gamma_{ij})^{(1-\delta_{ij})}$$

where r is the number of captured individuals over the experiment and γ_{ij} is defined by Huggins (1989) as follows:

$$\gamma_{ij} = \frac{p_{ij}}{1-(1-z_{ij})\prod_{l=j}^{t}(1-p_{il}^{*})} \tag{14}$$

where z_{ij} is the indicator of past capture history of individual i, i.e.

$$z_{ij} = \begin{cases} 1, \text{ if individual } i \text{ has been captured before } j, \\ 0, \text{ otherwise} \end{cases}$$

and p_{il}^{*} is p_{il} evaluated when $z_{il} = 0$. Notice that when $z_{ij} = 1$, $\gamma = p_{ij}$.

Huggins (1989; 1991) modelled the p_{ij} using logistic regression. The same approach is followed here by considering a linear logistic model for the capture probabilities,

$$p_{ij} = \frac{\exp(\beta_0+\beta_1 z_i+\beta_2 x_j+\beta_3 z_{ij})}{1+\exp(\beta_0+\beta_1 z_i+\beta_2 x_j+\beta_3 z_{ij})} \tag{15}$$

where z_i is an individual covariate and x_j is an occasion covariate.

Notice that when, for example, $z_{i2} = 1$ that means that up to time 2 either capture history 11 or 10 was observed, with 11 representing individual i was captured at times 1 and 2, and that $\gamma_{i2} = p_{i2}$. However, if $z_{i2} = 0$, then γ_{i2} is given by

$$\gamma_{ij} = \frac{p_{ij}}{1-\prod_{l=j}^{t}(1-p_{il}^{*})} \tag{16}$$

with $j = 2$; γ_{ij} of equation (16) is denoted by γ_{ij}^{*}.

Huggins (1991), in his appendix 2, shows that the likelihood function defined in equation (13) can be re-expressed by

$$\prod_{i=1}^{r}\prod_{j=1}^{t}\frac{p_{ij}^{\delta_{ij}}(1-p_{ij})^{(1-\delta_{ij})}}{1-\prod_{l=1}^{t}(1-p_{il}^{*})} \tag{17}$$

In the following sections, equation (17) is used in the likelihood.

In the logistic model of equation (15), the presence of z_{ij} allows capture probabilities to be modelled to vary according to an individual's capture history, i.e. allow for behavioural effect. In the case of the bowhead whales, the sampling procedure is not believed to produce any behavioural effect since systematic diving behaviour when the plane flies over the animals when the photographs are taken has not been observed. Therefore, in the model used here, $\beta_3 = 0$ and $p_{il}^{*} = p_{il}$. In the model proposed below for the bowhead whale, capture probabilities will be considered to vary only according to *occasion* and according to a group specific covariate that describes the amount of marking: $z_i = 1$ for highly marked whales and $z_i = 0$ for moderately marked whales. The effort in hours expended to take pictures on occasion j is defined as x_j. Therefore, the capture probabilities as defined by the linear logistic model are,

$$p_{ij} = \frac{\exp(\beta_0+\beta_1 z_i+\beta_2 x_j)}{1+\exp(\beta_0+\beta_1 z_i+\beta_2 x_j)} \tag{18}$$

The likelihood function

Estimating the total population size N requires the use of the available information on the marked individuals and some simplifying assumptions that also make sense in a biological context. Before describing the likelihood function which incorporates the unmarked whales, further notation must be introduced:

- p_{ij} is the capture probability of individual i at time j;
- θ_j represents the encountering probability at time j;
- λ_j^H is the average number of good photographs of highly marked whales;
- λ_j^M is the average number of good photographs of moderately marked whales;
- v is the probability of a marked whale being highly marked in the population;
- N^u is the population size of unmarked individuals;
- Z_i^H is an indicator, 1 if the marked whale is highly marked, and 0 if moderately marked;
- δ_{ij} is the indicator of capture of individual i at time j;
- X_{ij}^m is the number of good photographs of marked individual i at time j;
- X_j^u is the number of good photographs of unmarked whales at time j;
- r^H is the number of different highly marked individuals captured over the experiment;
- r^M is the number of different moderately marked individuals captured over the experiment; and
- r is the total number of different marked individuals captured over the experiment, $r = r^H + r^M$.

The model developed here follows Hammond (1986). He noted that the process of 'capture' and marking can be divided into three component parts:

(1) the whale must be sighted (encountered) by conducting a sample survey of some kind, usually from a boat or an aeroplane;
(2) once a whale has been seen it must present itself in such a way that a photograph of its natural markings can be taken; and
(3) once the best photograph of a particular whale has been selected, a decision must be made concerning how it should be treated.

Hammond (1986) argued that for all whales to have equal probability of capture they must all have the same probability of being sighted (encountered) and of presenting their markings (although strictly this need not be true in the unlikely event that the product of these probabilities were the same for all animals). Below it is assumed that all individuals have the same chance of being encountered (sighted) regardless of their amount of markings. The probability of an individual presenting its markings, that can be interpreted as the probability of a whale having at least one photograph that is good enough to be used in capture-recapture analysis, may differ according to the amount of markings the individual possesses. There is some evidence that highly marked whales have a higher average number of photographs good enough to be used in capture-recapture studies than moderately marked ones. Such distinction gives rise to the following expressions for capture probabilities, with θ_j representing the encountering probability and $(1 - e^{-\lambda_j^H})$ the probability of a highly marked whale having at least one good photograph,

$$p_{ij} = \theta_j(1 - e^{-\lambda_j^H}) \quad i = 1, \ldots, r^H;\ j = 1, \ldots, t \tag{19}$$

and similarly for moderately marked whales

$$p_{ij} = \theta_j(1 - e^{-\lambda_j^M}) \quad i = r^H + 1, \ldots, r;\ j = 1, \ldots, t \tag{20}$$

Notice that actually only two different capture probabilities must be estimated at each sampling occasion: one for the highly marked and one for the moderately marked ones, since all the animals are assumed to be equally affected by the occasion variable.

The encountering probabilities at each time are estimated by summing up the capture probabilities above at each time. By using simple algebraic calculation, the encounter probabilities θ_j are described as

$$\theta_j = \frac{r^H p_{1j} + r^M p_{rj}}{r^H(1 - e^{-\lambda_j^H}) + r^M(1 - e^{-\lambda_j^M})} \tag{21}$$

The encounter probabilities above are needed in the formulation of the modelling for the number of good photographs of unmarked whales as expressed by a random sum. Another assumption in our model is that the average number of good photographs of unmarked individuals is the same as for the moderately marked whales. That assumption, even though not ideal, is appropriate if the degree of marking affects the selection of photographs. It is not an unreasonable assumption, and it translates our belief that the average number of good photographs of unmarked whales is more likely to be closer to the moderately marked than to the highly marked ones.

The likelihood function is given by

$$L = P(\{\delta_{ij}, X_{ij}^m, Z_i^H\}, \{X_j^u\})$$

$$= P(\{X_{ij}^m | \delta_{ij}, Z_i^H\}) P(\{\delta_{ij}\} | \{Z_i^H\}) P(\{Z_i^H\}) P(\{X_j^u\})$$

$$= \prod_{i=1}^{r} \prod_{j=1}^{t} P(X_{ij}^m | \delta_{ij}, Z_i^H) P(\delta_{ij} | Z_i^H)$$

$$\times \prod_{j=1}^{t} P(\{X_j^u\}) \prod_{i=1}^{r} P(Z_i^H)$$

$$= \prod_{i=1}^{r_h} \prod_{j=1}^{t} P(X_{ij}^m | \delta_{ij}, Z_i^H = 1) P(\delta_{ij} | Z_i^H = 1)$$

$$\times \prod_{i=r^H+1}^{r} \prod_{j=1}^{t} P(X_{ij}^m | \delta_{ij}, Z_i^H = 0) P(\delta_{ij} | Z_i^H = 0)$$

$$\times \prod_{j=1}^{t} P(X_j^u) \prod_{i=1}^{r} P(Z_i^H)$$

$$= \prod_{i=1}^{r_h} \prod_{j=1}^{t} \frac{e^{-\lambda_j^H \delta_{ij}} (\lambda_j^H)^{x_{ij}^h \delta_{ij}}}{(x_{ij}^h \delta_{ij})!} \frac{1}{[(2 - \delta_{ij}) - e^{-\lambda_j^H \delta_{ij}}]}$$

$$\times \prod_{i=r^H+1}^{r} \prod_{j=1}^{t} \frac{e^{-\lambda_j^M \delta_{ij}} (\lambda_j^M)^{x_{ij}^m \delta_{ij}}}{(x_{ij}^m \delta_{ij})!} \frac{1}{[(2 - \delta_{ij}) - e^{-\lambda_j^M \delta_{ij}}]}$$

$$\times \prod_{i=1}^{r} \frac{p_{ij}^{\delta_{ij}} (1 - p_{ij})^{(1-\delta_{ij})}}{1 - \prod_{l=1}^{t} (1 - p_{il})}$$

$$\times \prod_{j=1}^{t} \left[\sum_{k=1}^{N^u} \frac{e^{-k\lambda_j^M} (k\lambda_j^M)^{x_j^u}}{x_j^u!} \binom{N^u}{k} \theta_j^k (1 - \theta_j)^{(N^u - k)} \right]$$

$$\times \prod_{i=1}^{r} v^{Z_i^H} (1 - v)^{(1 - Z_i^H)} \tag{22}$$

Note that $\{Z_i^H\}$ has been included in the likelihood to make it possible to allow highly marked whales to have a different average number of good photographs than moderately marked ones. That is essential in the characterisation of the model since capture probabilities are related to the number of good photographs, as can be seen from equation (19) and (20). The presence or not of the parameter v in the likelihood does not affect the estimation of the other parameters in the model, but it is necessary for some calculations involved in the unconditional parametric bootstrap procedure for this model since the total number of distinct individuals observed over the sampling experiment is random and so are the respective numbers of highly and moderately marked individuals.

The log of the likelihood function (equation 22) is maximised using an iterative procedure that consisted of maximising the function with respect to its continuous parameters when N^u was given a fixed initial value. Using the continuous parameter estimates, the function was then maximised with respect to N^u. This process was repeated until convergence. The stopping rule was based on the comparison of successive values of the log-likelihood function. The maximisation with respect to N^u was performed by finding the value of N^u which solved the difference equation

$$\log L(N^u) = \log L(N^u - 1) \tag{23}$$

This is the value of N^u that solves

$$\sum_{j=1}^{t} \log \left(\frac{\sum_{k=1}^{N^u} \frac{e^{-k\lambda_j^M} (k\lambda_j^M)^{x_j^u}}{x_j^u!} \binom{N^u}{k} \theta_j^k (1 - \theta_j)^{(N^u - k)}}{\sum_{k=1}^{N^u - 1} \frac{e^{-k\lambda_j^M} (k\lambda_j^M)^{x_j^u}}{x_j^u!} \binom{N^u - 1}{k} \theta_j^k (1 - \theta_j)^{(N^u - 1 - k)}} \right) = 0 \tag{24}$$

The estimated value of N^u obtained by fitting the multinomial model described in Section 2 was used as an initial value. Initial values for parameters β_0, β_1, and β_2, were estimated from a logistic regression that was fitted using *Splus* function glim. The dependent variable consisted of counts of the number of highly and moderately marked whales at each sampling occasion out of their respective estimated population sizes. Independent variables were Z_i^H and hours of sampling effort.

Estimation of *N*

Having estimated the population size of unmarked individuals, N^u, the next stage is to estimate the total population size, N, where $N = N^m + N^u$. The population size of the marked individuals, N^m, is composed of highly (H) and moderately (M) marked individuals, and it is described by the relationship $N^m = N^M + N^H$.

Following Huggins (1989), the method of moments is used to estimate N^H and N^M. Suppose that the full parameter vector denoted by θ is known. Let the probability that an individual is captured at least once during the course of the sampling experiment be denoted by

$$p_i(\theta) = 1 - \prod_{j=1}^{t}(1 - p_{ij}) \qquad (25)$$

Thus,

$$\hat{N}^m(\theta) = \sum_{i=1}^{r} p_i^{-1}(\theta) \qquad (26)$$

while N^H and N^M may be estimated separately by adding over the appropriate indices in the summation in equation (26) according to the probabilities expressed by equation (19) and (20).

Simulations

A set of 100 simulated datasets was generated. They comprise capture histories of the individuals and their respective number of good photographs, and a variable describing if a photographed individual was marked or unmarked and whether or not a naturally marked individual was highly or moderately marked. The average number of good photographs for the highly marked individuals was set to be slightly higher than for the moderately marked ones to mimic the true situation for bowhead whales. It is also assumed that unmarked individuals had the same average number of good photographs as moderately marked individuals.

The model described by equation (22) was fitted for the 100 generated datasets, and the multinomial model (equation 6) was also fitted for comparison (Table 4). The population size for the simulations was 6,734. Table 4 shows that both models seem, on average, to estimate the population size reasonably. The results do not strongly suggest that the model allowing for heterogeneity in capture probabilities is better than the simple multinomial model, although the standard deviation for the 100 $\hat{N}$ for the former model is smaller. The similarity in performance may be related to the fact that large differences in capture probabilities of highly and moderately marked individuals were not allowed for.

Table 4

Multinomial versus heterogeneity model.

Model	Mean	Bias	*s.d.*	*c.v.*
Multinomial	6,643	-91	774	0.12
Heterogeneity	6,841	107	701	0.10

In practice, the analyst has one dataset and, after estimating the population size, wishes to obtain a standard error for the estimate and a confidence interval. Huggins (1989) suggested a conditional parametric bootstrap, and da Silva (1999) also developed an unconditional parametric bootstrap. However, since each bootstrap sample requires a time-consuming iterative analysis, it is not feasible to compute many bootstrap replicates $\hat{N}$. The best approach for obtaining standard errors and confidence intervals for the heterogeneity model is the subject of ongoing work.

4. NONRANDOM SAMPLES: AGE SEGREGATION

Introduction

Capture-recapture theory is primarily based on the assumption that samples are drawn randomly. If this assumption is to hold, all animals in the population must be present in the survey area during each sampling occasion. For bowhead whales, that is unlikely given limitations in the time and area covered by the photographic surveys and age segregation in the population, both during the spring migration and on the summering grounds.

Hammond (1986) noted that if a group of animals is consistently less available to be sampled, heterogeneity in capture probabilities is present because this group will have a lower probability of being photographed than the rest of the population. Because age segregation is likely to define such a group in the bowhead case, here we develop methods of accounting for non-random sampling based on categorising the photographed whales as either immature ($\leq$ 13.0m long) or mature ($>$ 13.0m long). This is possible because a major goal of most of the photographic surveys was to determine the distribution of lengths, and ultimately ages, in the population. To achieve this, most of the photographed whales were measured using photogrammetric techniques and their lengths included in the data base.

The photographic surveys have provided good information about the proportions of mature and immature whales in the population, summarised by Angliss *et al.* (1995). While it is certainly possible that a sample with the expected proportion of mature and immature whales could nevertheless be non-random, samples with a greatly disproportionate number of mature or immature whales are certainly non-random. For example, the sample taken in summer 1985 had far too many immature animals to be a random sample of the whole population. The one taken in spring 1986 had too many mature animals, while the sample taken in summer 1986 had again too many immature animals.

Reasons why some samples were not random are known, and are related to severe weather conditions or logistical problems that prevented conducting surveys throughout the season. Withrow and Angliss (1992) noted that the spring 1986 study began two weeks late and missed the beginning of the migration. Angliss *et al.* (1995) demonstrated temporal age segregation during the spring migration. The earliest whales to migrate tend to be small. The later migrants are mostly adults (mature whales). Therefore, since the spring 1986 survey missed the first portion of the migration, which is composed primarily of immatures, this segment is underrepresented in the spring 1986 sample.

In summer 1985, a major photographic effort was conducted with the goal of estimating bowhead gross annual reproductive rate. Virtually all of the known summer range of the bowhead whales was searched, but, for unknown reasons, very few adults were found. Davis *et al.* (1986) speculated that

The unusually heavy ice conditions in the Beaufort Sea in 1985 apparently caused major shifts from the normal patterns of summer whale distribution. Results from aerial photography suggest that the actively breeding segment (adults with calves) of the population was essentially absent from the study area in 1985. In late August, only an estimated 229 of the 2251 bowhead whales accounted for were adults.

Koski *et al.* (1988) summarised the evidence for age segregation on the summering grounds.

Table 5, compiled from the above sources, summarises the percent mature/immature whales by sampling occasion and reveals wide differences in percent mature. An attempt to estimate population size with the methods developed so far would be suspect. Methods for estimating population size in the presence of non-random samples are highly desirable because this occurs frequently in capture-recapture studies involving cetaceans. To accommodate non-random sampling, the model in Section 3 is adapted by defining a new covariate that accounts for departures from random sampling. The additional covariate will help correct the magnitude of the capture probabilities to reflect the effect of the non-random samples. The idea is to treat non-randomness as a form of heterogeneity in capture probabilities as Hammond (1986) suggested.

Table 5

Percent mature whales by sampling occasion.

Sample	% Mature
Spring 1985	42.5
Summer 1985	13.6
Spring 1986	61.7
Summer 1986	23.3

Notation

p_{ij} is the capture probability of individual i at time j.
v_j represents the encountering probability at time j.
λ_j is the average number of good photographs of an encountered whale.
φ_{1j} is the conditional probability of encountering an individual at time j given that it is mature.
φ_{2j} is the conditional probability of encountering an individual at time j given that it is immature.
θ is the probability of an individual being mature.
N^u is the population size of unmarked individuals.
δ_{ij} is the indicator of capture of individual i at time j.
X_{ij}^m is the number of good photographs of marked individual i at time j.
X_j^u is the number of good photographs of unmarked whales at time j.
I_j is a vector describing whether a good photograph is from a mature (1) or immature (0) whale at time j.
ef_j^{mat} is the sampling effort expended to catch mature whales at time j.
ef_j^{imm} is the sampling effort expended to catch immature whales at time j.
mat_i is an indicator variable that assumes value 1 if marked whale i is mature and 0 elsewhere.

Let us now define some events that will be needed in the description of some probabilities that are used in the formulation of the model.

C_j is the event a whale is captured at time j.
E_j is the event a whale is encountered at time j.
M is the event a whale is mature.
I is the event a whale is immature.
1_j^+ is the event an encountered whale has at least one good photograph at time j.

The model idealised to allow for non-randomness is based on the estimation of the population size of the marked individuals when non-random samples were taken, but at least one random sample is available. A covariate describing departures from that random sample was defined and it accounts for differences in effort per maturity class. Once population size of the marked whales is estimated, the unmarked part of the population is accounted for via the random sum model for the number of good photographs of the unmarked individuals. The encountering probabilities needed in the random sum model are a function of capture probabilities and probability of a whale being mature. This assumes that encounter probabilities are related to maturity but not amount of markings since marked and unmarked mature individuals tend to migrate together. The same occurs with marked and unmarked immature individuals. Therefore φ_{1j} and $\varphi_{2j}, j = 1,\ldots, t$ are assumed to be the same for marked and unmarked individuals.

The conditional probability of capturing a whale at time j, given that it is mature, is expressed by the product of the probability of a whale having at least one good photograph taken given that it was encountered, times the conditional probability of encountering that whale at time j given that it is mature,

$$P(C_j \mid M) = P(E_j \cap 1_j^+ \mid M) = P(1_j^+ \mid E_j)P(E_j \mid M) = (1 - e^{-\lambda_j}) \times \varphi_{1j}$$

so the conditional probability of encountering a whale at time j given that it is mature is

$$\varphi_{1j} = P(C_j \mid M) / (1 - e^{-\lambda_j}) \tag{27}$$

The probability of encountering a whale at time j is defined as a function of the probabilities above,

$$\begin{aligned} P(E_j) &= P(E_j \mid M)P(M) + P(E_j \mid I)P(I) \\ &= \varphi_{1j}\theta + \varphi_{2j}(1-\theta) \\ &= v_j \end{aligned} \tag{28}$$

Only the data from the random sampling occasions are used in the estimation of the probability of a whale being mature. The best data available for the estimation of a whale being mature in the population are from the spring 1985 survey. The probability of a whale being mature must not change in the small time window being considered here (two years). All good photographs of whales from the spring 1985 survey are used in the estimation of the probability of a whale being mature (θ). These whales are categorised as being mature or immature based on their length. In the simulated data, it is assumed that lengths are available for all whales. Although an immature whale could reach maturity between the 1985 and 1986 samples, this possibility is ignored because the slow growth of bowhead whales and the small sample size make it unlikely that such a whale would be sampled.

The 'outcome' of a photograph being from a mature whale is being modelled as a Bernoulli trial, although such modelling has limitations since some of its assumptions are violated, because multiple photographs of some whales are not independent. However, the violation is mild because few photographs are taken of each whale (1.5 photographs/whale).

The likelihood function

$$L = P(\{I_1\},\{X_j^u\},\{X_{ij}^m\},\{\delta_{ij}\})$$

$$= P(\{I_1\} \big| \{X_{i1}^m\},\{X_1^u\}) \times P(\{X_{ij}^m\} \big| \{\delta_{ij}\}) P(\{\delta_{ij}\}) P(\{X_j^u\})$$

$$= \prod_{g=1}^{x_1^u + x_1^m} \theta^{I_{1g}} (1-\theta)^{(1-I_{1g})}$$

$$\times \prod_{j=1}^{t} \prod_{i=1}^{r} \frac{e^{-\lambda_j \delta_{ij}} (\lambda_j)^{x_{ij}^m \delta_{ij}}}{(x_{ij}^m \delta_{ij})!} \frac{1}{[(2-\delta_{ij}) - e^{-\lambda_j \delta_{ij}}]}$$

$$\times \prod_{j=1}^{t} \prod_{i=1}^{r} \frac{p_{ij}^{\delta_{ij}} (1-p_{ij})^{(1-\delta_{ij})}}{1 - \prod_{l=1}^{t} (1-p_{il})}$$

$$\times \prod_{j=1}^{t} \left[\sum_{k=1}^{N^u} \frac{e^{-k\lambda_j} (k\lambda_j)^{x_j^u}}{x_j^u!} \binom{N^u}{k} v_j^k (1-v_j)^{(N^u-k)} \right] \quad (29)$$

where v_j is the conditional probability of encountering a whale at time j given that it is unmarked, and it is given by

$$v_j = \varphi_{1j}\theta + \varphi_{2j}(1-\theta) \quad (30)$$

Capture probabilities are described by:

$$p_{ij} = \frac{\exp(\beta_0 + \beta_1(mat_i \times ef_j^{mat} + (1-mat_i) \times ef_j^{imm}))}{1 + \exp(\beta_0 + \beta_1(mat_i \times ef_j^{mat} + (1-mat_i) \times ef_j^{imm}))} \quad (31)$$

As earlier, the log of the likelihood (equation 29) is maximised using an iterative procedure that consisted of maximising that function with respect to its continuous parameters when N^u was given a fixed initial value. Using the resulting estimates of the continuous parameters, equation 29 was then maximised with respect to N^u. Once the likelihood was maximised, N^m is estimated as in Section 3.

Simulations

A set of 100 datasets was generated. The composition of the drawn samples at each sampling occasion in terms of proportion mature was simulated to reflect the values displayed in Table 5. It is assumed that 58% of the whales in the population were immature. Among the unmarked whales, 70% were assumed to be immature. The simulated population size was 6,734.

The data consisted of capture histories of the individuals and their respective number of good photographs, a variable describing if a photographed individual was marked or unmarked, one describing whether the individual was mature or immature, and the effort data for mature and immature whales on each sampling occasion. The average number of good photographs was constant for all the whales.

The model described by equation (29) was fitted for the 100 generated datasets. A comparison of the results with the multinomial model (equation 6) is given in Table 6.

Table 6

Multinomial versus non-random sample model.

Model	Mean	Bias	*s.d.*	*c.v.*
Multinomial	8254	1520	1211	0.15
Non-random	6815	81	984	0.14

As expected, the non-random sample model performed better than the multinomial model for estimating the total population size N. While the bias of the estimated value of N obtained by the multinomial model is 1,520, that value is only 81 for the non-random sample model. There is also a considerable gain in precision. As in Section 3, a parametric bootstrap was developed to estimate standard errors (da Silva, 1999), but only a few bootstrap replicates for a few samples could be computed because of time constraints.

5. RESULTS FROM THE ACTUAL DATA

Datasets used

Rescoring of the 1985 and 1986 photographs using the scoring system of Rugh *et al.* (1998) was completed by two of us (LB and GM) and a data base containing all the data from these years prepared by WK. Preliminary analyses by JZ confirmed that the mid-back region (Rugh *et al.*, 1998) provided the most good photographs (quality at least 2) and the most recaptures, compared to the rostrum, lower back and fluke. Four sampling occasions (spring 1985, summer 1985, spring 1986, and summer 1986) were considered. The variables in the data base were used to create a dataset containing records with the following information:

(1) WHALE: whale's number. Each marked whale has a unique number, but the same unmarked whale could occur in the dataset more than once with different numbers.
(2) H: a categorical variable indicating whether the photographed whale in a good photograph was unmarked (-1), moderately marked (0), or highly marked (1).
(3) MAT: a categorical variable indicating whether the photographed whale in a good photograph was immature (0), or mature (1).
(4) Four columns indicating the capture histories of the bowhead whales, with 1 indicating that the whale was, and 0 that it was not, captured in the sample represented by the column.
(5) Four columns indicating the number of good photographs obtained for each of the captured individuals by sampling occasion.

There were 1,190 records in the dataset, of which only 175 belonged to marked individuals. The subset of 175 marked whales was used for capture-recapture estimation. Only 12 of the 175 were captured more than once over the four sampling occasions.

It is important to recognise that there are many more than 175 identified bowhead whales in the photographic collection, and many more re-identifications than 12, even when attention is restricted to the years considered here. The capture-recapture dataset does not contain them all because many did not provide good photographs of the mid-back, or the marks by which they are identified occur on different parts of the body.

The requirement for length data so that the whale could be categorised as mature or immature also reduced the dataset. To mitigate this problem, a larger dataset was created that could be used in all of the estimation procedures discussed here with the exception of that which allows for non-random samples. The larger dataset contained 1,677 records, 229 belonging to marked individuals, with 16 of 229 captured on more than one occasion.

Table 7 compares the real to simulated data with respect to the frequencies of recaptured individuals with a given capture history w. For the simulated data, the average number of individuals (rounded to the nearest integer) with a given capture history was taken, based on 100 simulated datasets where the samples were not random. The simulations were used under the non-random sampling model for the comparisons of this section because they are expected to more closely match the actual data than the other simulations.

Notation w_{12}, for example, means that an individual was captured on sampling occasions 1 and 2. Although the number of captured individuals in the simulated data is much larger than in the real data (see Table 8), the scarcity of recaptures between the spring and summer 1985 samples (w_{12}) causes some concern. The 1985 sampling occasions had the largest sampling effort. In the simulated data the average number of recaptures for that category was 11, the largest in the table. This issue requires further investigation.

Table 8 shows the number of marked individuals captured by sampling occasion and the percentage of the marked population (estimated by $\hat{N}^m$ from the model of Section 2) for the real datasets. Averages over the simulated datasets are also given, and for these, the average numbers of marked individuals captured are small, representing at most 8.2% of the total number in the marked population. For the real data the situation is worse. The largest estimated capture probability, even in the large dataset, is 6.7% in the spring 1985 sample. On the fourth sampling occasion, only 18 individuals were captured in the smaller and 26 in the larger dataset. This cannot be expected to yield very reliable estimated values for N.

There were also fewer highly marked whales than expected in the real datasets. Table 9 compares the numbers in the real datasets with the numbers in the simulated data. In the actual data, only 18% to 19% of the marked whales captured were highly marked, compared to 27% in the simulated data.

Table 10 compares the numbers of marked mature individuals captured at each sampling occasion in the actual data with the average numbers obtained in the simulated data. There is good agreement between the percent values for real and simulated data. Note these numbers differ from the percentage of mature whales in the population, estimated at around 50% of the 1+ population in our capture-recapture analyses and around 43% by Angliss *et al.* (1995) from a larger dataset. This is because many more immature than mature whales are unmarked.

Precise information about effort is not available and it was necessary to use an *ad hoc* procedure to obtain a crude estimate of hours of effort expended to catch individuals in a given maturity class. This was done by counting up hours in which any whale of that maturity class (whether marked or unmarked) was photographed. The effort data by maturity class used in the analyses are summarised in Table 11. This table also gives the overall effort expended to capture whales of either maturity class. This overall effort is used in estimation of capture probabilities under the heterogeneity model. It is less than the sum of the separate efforts because both mature and immature whales were captured during some hours.

Table 8

Real versus simulated data - number of marked individuals captured by sampling occasion.

Real data, 175 individuals	Sp85	Su85	Sp86	Su86
Number captured	59.0	52.0	58.0	18.0
% of marked population	5.8	5.1	5.7	1.8
Real data, 229 individuals	Sp85	Su85	Sp86	Su86
Number captured	87.0	56.0	76.0	26.0
% of marked population	6.7	4.3	5.8	2.0
Average of simulated data	Sp85	Su85	Sp86	Su86
Average number captured	147.0	153.0	115.0	77.0
% of marked population	7.9	8.2	6.2	4.2

Table 9

Real versus simulated data - number of highly marked captured individuals.

Data set	Sp85	Su85	Sp86	Su86
Real data, 175 individuals	13	9	9	3
Real data, 229 individuals	22	9	10	6
Average of simulated data	46	25	43	18

Table 10

Real versus simulated data – mature individuals in samples.

Real data, 175 individuals	Sp85	Su85	Sp86	Su86
# of mature whales	44	20	49	7
% of sampled marked whales	75	38	84	39
Average of simulated data	Sp85	Su85	Sp86	Su86
# of mature whales	109	59	97	42
% of sampled marked whales	74	39	87	55

Table 7

Real versus simulated data - number of recaptured individuals with capture history *w*.

Case	w_{12}	w_{13}	w_{14}	w_{23}	w_{24}	w_{34}	w_{123}	w_{124}	w_{134}	w_{234}	w_{1234}
Real data 175 ind.	0	5	0	1	6	0	0	0	0	0	0
Real data 229 ind.	1	7	0	1	7	0	0	0	0	0	0
Sim. data 446 ind. avg.	11	8	5	6	7	3	0	1	0	0	0

Table 11

Effort data (hours).

Maturity	Sp85	Su85	Sp86	Su86
Mature	55	56	52	32
Immature	57	108	27	42
Either	80	132	68	58

Results

Using the multinomial model described in Section 2, the total population size N is estimated using the datasets containing 175 and 229 marked individuals. The results are summarised in Table 12. The dataset containing 229 whales led to a higher estimated value of N. The standard error, estimated from 3,000 bootstrap replications, was also somewhat higher but the CV lower. The bootstrap bias estimate was around 400 in both cases.

Table 12

Real bowhead data – summary of the results.

Model	Marked captured	$\hat{N}$	*s.e.**	C.I.	*bias**
Multinomial	175	5,071	1,871	(3,195,10,002)	410
Heterogeneity	175	5,116	–	–	–
Non-random	175	4,719	1,696	(2,382,9,343) –	–
Multinomial	229	7,022	2,017	(4,701,12,561)	413
Heterogeneity	229	7,331	–	–	–

It is encouraging that both the population estimates and their standard errors are comparable to the estimates obtained from the combined visual and acoustic census efforts conducted near Point Barrow in 1985 and 1986. Raftery and Zeh (1998) applied the generalised removal method to the combined visual and acoustic data collected during those years and obtained estimates of the size of the Bering-Chukchi-Beaufort Seas stock of bowhead whales (including calves) of 6,039 (SE = 1,915) and 7,734 (SE = 1,450) for 1985 and 1986 respectively. Both these years were ones in which environmental conditions for conducting a census were not ideal. Estimates with smaller standard errors are obtained from the census when conditions are better, but years with ideal conditions are relatively rare.

Recall that the estimates in Table 12 exclude calves, unlike the estimates of Raftery and Zeh (1998). The estimates of Table 12 also compare well with the 1985 and 1986 estimates of 6,649 and 6,820 (excluding calves) from the Bayesian synthesis analysis of Givens (pers. comm.). The estimates of Givens incorporate the estimates of Raftery and Zeh (1998) and additional data on bowhead whale population dynamics.

Table 12 also shows population estimates from the heterogeneity model of Section 3. They are slightly higher than the multinomial model estimates, suggesting that there may be some negative bias in the multinomial model values because highly marked whales are more likely to be captured than moderately marked. However, convergence to the estimates given was slow. Consideration of the number of parameters in this model and the limitations of the actual data, as compared to simulated data (Table 9), leads to the conclusion that the heterogeneity model has too many parameters for the data to support. There were too few highly marked whales in the actual data.

Bootstrap standard errors for the heterogeneity model were not estimated since the slow convergence would have made computing time prohibitive.

Only the smaller dataset could be used in the model of Section 4 that accounts for non-random sampling because length data are required to assess maturity. Using the model for non-random sampling described in Section 4, we obtained the population estimate given in Table 12 for that model. As expected, it is smaller than the multinomial model estimate since it was designed to avoid the positive bias exhibited by the multinomial model estimate (Table 6) when applied to simulated bowhead data with non-random sampling. Some 100 bootstrap replications were carried out for the non-random sampling model to obtain the standard error given in Table 12, too few to provide an estimate of bias or permit use of a percentile confidence interval. The confidence interval given in Table 12 is the log-normal confidence interval (equation 10) using the parametric bootstrap standard error estimate. It covers the values of 1985 and 1986 population size obtained from the ice-based census and population dynamics modelling.

Discussion

The estimated values of N in Table 12 agree with the results from the simulations discussed in the previous chapters. When applied to the data, the heterogeneity model corrects for negative biases resulting from highly marked whales being more likely to be recaptured than moderately marked ones. The non-random sample model corrects for positive biases caused by the reduced number of recaptures that can occur when samples are non-random. Both $\hat{N}$ and its standard error are somewhat smaller than the values from the multinomial model. Those results are in agreement with the simulations. The estimates in Table 12 are not precise enough to provide a clue as to whether the two kinds of bias partially cancel each other in the multinomial model estimates.

The differences between the models are small compared to the differences that result from increasing the number of marked whales by 31%, and the number of photographs of both marked and unmarked whales correspondingly, by relaxing the requirement for length data. While da Silva (1999) has outlined a model that accounts for both heterogeneity and non-random sampling, it is clear that model will have too many parameters for the data to support. Even if the heterogeneity side of the model is refined to reduce the number of parameters, it is unlikely that the 1985 and 1986 bowhead data can support its use. We hope, however, that the approaches developed in this paper prove useful for other photographic studies.

Regardless of the possible sources of bias mentioned, the confidence intervals of Table 12 all cover the 1985 and 1986 estimates of 6,649 and 6,820 (excluding calves) from the Bayesian synthesis analysis of Givens (pers. comm.), which reflect the best information available on the size of the Bering-Chukchi-Beaufort Seas stock of bowhead whales in those years. Since the population size data and estimation methods on which the Givens estimates are based are completely different from the data and methods used here, our results provide independent confirmation for the population estimates currently accepted by the IWC Scientific Committee.

The real dataset available so far is too small to provide a precise estimate of population size for the bowhead whale. In addition, refinements in both data and methods are needed. Work that needs to be undertaken includes the following:

(1) more thorough review of the data to ensure that all matches have been located and other data errors have been corrected;
(2) more refined estimation of sampling effort;
(3) measurement or estimation of length data for as many whales as possible and development of an estimator that allows for missing length data;
(4) development of models that permit use of rostrum, lower back and/or fluke data in addition to mid-back data, so the sample size of marked whales can be increased – this may require developing models that allow for matching error, i.e. the failure to recognise that two photographs are of the same whale;
(5) implementation of a model that accounts for both heterogeneity and non-random sampling but is as parsimonious as possible, and testing of that model on simulated data;
(6) extension of the methods developed here to open population models so data collected over the last two decades can be included to improve precision;
(7) the extended models need to allow for changes in markings and maturity status over the years;
(8) completion of scoring and matching work and incorporation of all years of data, not just 1985 and 1986, into the data base so that the extended methods can be used;
(9) examination of the relative cost and difficulty of the census effort, compared to the collection and analysis of several large photographic samples, since we have shown that comparable population estimates can be obtained using the two methods.

ACKNOWLEDGEMENTS

This research was supported by the North Slope Borough (Alaska) and the National Oceanic and Atmospheric Administration (through the National Marine Mammal Laboratory, both directly and via the Alaska Eskimo Whaling Commission). We thank Craig George, Department of Wildlife Management, North Slope Borough; Tore Schweder; and an anonymous reviewer for their helpful reviews of the manuscript.

REFERENCES

Alho, J.M. 1990. Logistic regression in capture-recapture models. *Biometrics* 46:623-35.

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.

Bickel, P. and Freedman, D.A. 1981. Some asympotic theory for the bootstrap. *Ann. Stat.* 9:1196-217.

Buckland, S.T. 1980. A modified analysis of the Jolly-Seber capture-recapture model. *Biometrics* 36:419-35.

Buckland, S.T. and Garthwaite, P.H. 1991. Quantifying precision of mark-recapture estimates using the bootstrap and related methods. *Biometrics* 47:255-68.

Burnham, K.P., Anderson, D.R., White, G.C., Brownie, C. and Pollock, K.H. 1987. *American Fisheries Society Monographs*. No. 5. *Design and Analysis Methods for Fish Survival Experiments Based on Release-Recapture*. American Fisheries Society, Bethesda, Maryland. x+437pp.

Carothers, A.D. 1973. Capture-recapture methods applied to a population with known parameters. *J. Anim. Ecol.* 42:125-46.

Chapman, D.G. 1952. Inverse, multiple and sequential sample censuses. *Biometrics* 8:286-306.

Cormack, R.M. 1989. Log-linear models for capture-recapture. *Biometrics* 45:395-413.

Cormack, R.M. 1992. Interval estimation for mark-recapture studies of closed populations. *Biometrics* 48:567-76.

Darroch, J.N. 1958. The multiple-recapture census. I: Estimation of a closed population. *Biometrika* 45:343-59.

da Silva, C.Q. 1999. Capture-recapture estimation of bowhead whale population size using photo-identification data. Doctoral Thesis, University of Washington, Seattle, WA. i-x + 190pp.

Davis, R.A., Koski, W.R., Miller, G.W., McLaren, P.L. and Evans, C.R. 1986. Reproduction in the bowhead whale, summer 1985. Paper SC/38/PS2 presented to the IWC Scientific Committee, May 1986 (unpublished). 127pp. [Available from the Office of this Journal].

Efron, B. 1981. Nonparametric standards errors and confidence intervals (with discussion). *Can. J. Stat.* 9:139-72.

Efron, B. and Tibshirani, R.J. 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New York. 436pp.

Friday, N., Smith, T.D., Stevick, P. and Allen, J. 2000. Measurement of photographic quality and whale distinctiveness for the photographic identification of humpback whales, *Megaptera novaeangliae*. *Mar. Mamm. Sci*. 16: 355-74.

Garthwaite, P.H. and Buckland, S.T. 1990. Analysis of a multiple-recapture census by computing conditional probabilites. *Biometrics* 46:231-8.

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 acid racemization. *Can. J. Zool.* 77:571-80.

Givens, G.H. 1993. A Bayesian framework and importance sampling methods for synthesising multiple sources of evidence and uncertainty linked by a complex mechanistic model. Doctoral Dissertation Thesis, University of Washington, Seattle, WA 98195. i-xii + 157pp.

Hammond, P.S. 1986. Estimating the size of naturally marked whale populations using capture-recapture techniques. *Rep. int. Whal. Commn* (special issue 8):253-82.

Hammond, P.S., Mizroch, S.A. and Donovan, G.P. (eds.). 1990. *Rep. int. Whal. Commn*. (special issue 12). *Individual Recognition of Cetaceans: Use of Photo-Identification and Other Techniques to Estimate Population Parameters*. International Whaling Commission, Cambridge, UK. [vi]+440pp.

Huggins, R.M. 1989. On the statistical analysis of capture experiments. *Biometrika* 76:133-40.

Huggins, R.M. 1991. Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics* 47:725-32.

International Whaling Commission. 2000. Report of the Scientific Committee. Annex E. Report of the Standing Working Group on the Development of an Aboriginal Subsistence Whaling Management Procedure. *J. Cetacean Res. Manage. (Suppl.)* 2:125-54.

Johnson, K. 1986. The bizarre diet of the walrus. A look at the history and natural history of the most unusual animal on two teeth. *Discovery* 19(1):2-7.

Koski, W.R., Miller, G.W. and Davis, R.A. 1988. The potential effects of tanker traffic on the bowhead whale in the Beaufort Sea. Report from LGL Limited, King City, Ontario, to Department of Indian Affairs and Northern Development, Hull, Quebec, K1A 0H4, Canada. 150pp.

Miller, G.W., Davis, R.A., Koski, W.R., Crone, M.J., Rugh, D.J., Withrow, D.E. and Fraker, M.A. 1992. Calving intervals of bowhead whales - an analysis of photographic data. *Rep. int. Whal. Commn* 42:501-6.

Otis, D.L., Burnham, K.P., White, G.C. and Anderson, D.R. 1978. Statistical inference from capture data on closed animal populations. *Wildl. Monogr.* 62:1-135.

Pollock, K.H. 1974. The assumption of equal catchability of animals in tag-recapture experiments. Doctoral Thesis, Cornell University, Ithaca, New York.

Pollock, K.H. 1991. Modelling capture, recapture and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present and future. *J. Am. Stat. Assoc.* 86:225-38.

Pollock, K.H. and Otto, M.C. 1983. Robust estimation of population size in closed animal populations from capture-recapture experiments. *Biometrics* 39:1035-49.

Pollock, K.H., Hines, J.E. and Nichols, J.D. 1984. The use of auxiliary variables in capture-recapture and removal experiments. *Biometrics* 40:329-40.

Raftery, A.E. and Zeh, J.E. 1998. Estimating bowhead whale population size and rate of increase from the 1993 census. *J. Am. Stat. Assoc.* 93:451-63.

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

Richardson, W.J., Koski, W.R., Miller, G.W. and Würsig, B. 1995. Bowhead whale results: movements, behavior and basic biology. pp. 138-246. *In:* W.J. Richardson, W.R. Koski, G.W. Miller and B. Würsig (eds.) *Acoustic effects of oil production activities on bowhead and white whales visible during spring migration near Pt. Barrow, Alaska-1991 and 1994 phases*. Unpublished report number MMS 95-0051 by LGL Ltd., King City, Ont., and Greeneridge Sciences Inc., Santa Barbara, CA, for Minerals Management Service. xxxii+539pp. [Available from U.S. Minerals Management Service, Procurement Operations, 381 Elden St./MS 2500, Herndon, VA 22070-4817].

Robson, D.S. and Rieger, H.A. 1968. Estimation of population and mortality rates. pp. 124-58. *In:* W.E. Ricker (ed.) *IBP Handbook No. 3. Methods for Assessment of Fish Production in Fresh Waters*. Blackwell Scientific Publications, Oxford. xv+365pp.

Rugh, D.J. 1990. Bowhead whales reidentified through aerial photography near Point Barrow, Alaska. *Rep. int. Whal. Commn* (special issue) 12:289-94.

Rugh, D.J., Braham, H.W. and Miller, G.W. 1992. Methods for photographic identification of bowhead whales, *Balaena mysticetus*. *Can. J. Zool.* 70:617-24.

Rugh, D.J., Zeh, J.E., Koski, W.R., Baraff, L.S., Miller, G.W. and Shelden, K.E.W. 1998. An improved system for scoring photo quality and whale identifiability in aerial photographs of bowhead whales. *Rep. int. Whal. Commn* 48:501-12.

Schnabel, Z.E. 1938. The estimation of the total fish population of a lake. *Amer. Math. Month.* 45:348-52.

Seber, G.A.F. 1982. *The Estimation of Animal Abundance and Related Parameters*. 2nd Edn. Charles Griffin and Company Ltd., London. i-xvii+654pp.

Whitcher, B.J., Zeh, J.E., Rugh, D.J., Koski, W.R. and Miller, G.W. 1996. Estimation of adult bowhead whale survival rates. Paper SC/48/AS12 presented to the IWC Scientific Committee, June 1996 (unpublished). 22pp. [Available from the Office of this Journal].

White, G.C., Anderson, D.R., Burnham, K.P. and Otis, D.L. 1982. *Capture-Recapture and Removal Methods for Sampling Closed Populations*. Los Alamos National Laboratory, Los Alamos, NM, USA. xvi+235pp. [Available from NTIS, 5285 Port Royal Road, Springfield, VA 22161].

Williams, J.A., Dawson, S.M. and Slooten, E. 1993. Abundance and distribution of bottlenose dolphins (*Tursiops truncatus*) in Doubtful Sound, New Zealand. *Can. J. Zool.* 71:2080-8.

Withrow, D. and Angliss, R. 1992. Length frequency of bowhead whales from spring aerial photogrammetric surveys in 1985, 1986, 1989 and 1990. *Rep. int. Whal. Commn* 42:463-7.