Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística

A Comparison Between Bayesian and Frequentist analyses for estimating bowhead whale population size using photo-identification

C. Q. da-Silva, J. Rodrigues, J. G. Leite and L. A. Milan

Relatório Técnico
RTP-03

Relatório Técnico
Série Pesquisa

A Comparison Between Bayesian and Frequentist analyses for estimating bowhead whale population size using photo-identification

Cibele Q. da-Silva

UFMG-ICEx, 30123-970, Belo Horizonte, MG, Brazil

Josemar Rodrigues, José G. Leite e Luis A. Milan

Des-UFSCar, 13565-905, São Carlos, SP, Brazil

Abstract

We develop a Bayesian statistical model for estimating bowhead whale population size from photo-identification data. The proposed conditional likelihood function is a product of Darroch's model, formulated as a function of the number of good photos, and a binomial distribution of captured whales given the total number of good photos at each occasion. The full Bayesian model is implemented via adaptive rejection sampling for log concave densities. We apply the model on data from 1985 and 1986 bowhead whale photographic studies and the results compare favorably with the ones obtained in the literature. Also, a comparison with the maximum likelihood procedure with bootstrap simulation is considered using different noninformative priors for the capture probabilities.

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 Scientific Committee (IWC SC). 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 bowheads have been interested in obtaining an independent population size estimate from such data.

In capture-recapture studies it is common practice to take samples of animals from a given population and mark them for future recognition. There are several possibilities to do that and a choice may depend heavily on the species that is being studied. In capture-recapture studies involving birds it is common to mark a captured individual using a numbered ring placed on its legs. In other studies animals may be either painted or tattooed, or even get a radio transmitter. In those contexts, marked individuals are animals that were captured, handled and got a unique artificial mark.

In the case of the bowhead whale it is not possible to attach an artificial mark to the captured individuals, but the natural marks acquired throughout their lives are useful to allow the analyst to distinguish individuals. Contrary to the notion of a marked individual in the capture-recapture studies described above, a *marked bowhead* means that it has acquired natural marks enough to make reidentification possible.

Since the majority of bowheads are unmarked and therefore uncatchable 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 $\hat{N} = \frac{\hat{N}^m}{\hat{p}^*}$, 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 Seber's approach 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. They used photos 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. To address this issue, da Silva (1999) and da Silva *et al.* (2000) developed alternative interval estimates of population size from photo-identification data when the population includes unmarked animals and compared their approach to the one developed by Williams *et al.* (1993) using simulated bowhead data. The work of da Silva *et al.* (2000) showed that: (i) their results had a good agreement with previous works (Raftery and Zeh (1998), Givens (1993) - personal communication), and the simulations indicated that their bootstrap intervals did not outperform log-normal intervals (Burnham *et al.*, 1987); (ii) the large sample size theory of maximum likelihood estimators is not applied since the likelihood function does not belong to a regular family of distributions; (iii) the truncated binomial model in the likelihood function implied a sophisticated algorithm with restrictions to obtain the maximum likelihood estimator. These points motivated approaching the present problem with a simpler likelihood function under a Bayesian perspective via adaptive rejection sampling allowing a very natural derivation of inferences for the parameters of interest. Also, a comparison with the maximum likelihood procedure with bootstrap simulation (da Silva *et al.*, 2000) is considered using three different ways to express vague prior information. We start the next section by introducing notation.

2 Notation

Quality of photos and extent of natural marks of a whale are important variables in our model formulation. Some variables which are functions of *quality* and *extent* also play an important role in our analyses, and they represent basically all the information we have available for estimating the population size of the bowheads. A capture essentially means that a good quality photo of a whale was taken. In this case, if a natural mark is found then the whale is considered marked. We now introduce some notation.

- N^u : the total number of unmarked whales in the population.
- N^m : the total number of marked whales in the population.
- $N = N^m + N^u$: the total number of whales.
- X_j^m : the number of good photos of marked whales at occasion j, $j = 1, \ldots, t$, where good photos are those from which the identification of the whales are possible.
- X_{ij}^m : the number of good photos of the ith marked whale at occasion j, $j = 1, \ldots, t$ and $i = 1, \ldots, N^m$.
- X_j^u : the number of good photos of unmarked whales at occasion j.
- I_j^m : the number of bad photos of marked whales, a latent variable.
- I_j^u : the number of bad photos of unmarked whales, a latent variable.
- $I_j = I_j^m + I_j^u$: the total number of bad photos taken at time j.
- $T_j^m = X_j^m + I_j^m$: the total number of photos taken from marked individuals at time j, a latent variable.
- $T_j^u = X_j^u + I_j^u$: the total number of photos taken from unmarked individuals at time j, a latent variable.

- The total number of good photos at occasion j: $X_j = X_j^m + X_j^u$.
- S_j : the total number of photos at time j.
- n_j : the total number of marked whales captured at time j.
- r : the number of different whales marked over the experiment.

We can organize the information related to the quality of photos taken from the whales and extent of the natural marks of an individual at time j into a two-way contingency table. We are restricting our analysis to the case of whales classified either as *marked* or as *unmarked.*

Table 1: *Two-way table relating quality of photos and extent of natural marks at time j*

Photo	*extent of natural marks*		
Quality	*Marked-M*	*Unmarked-U*	
Good-G	X_j^m	X_j^u	X_j
Bad-B	I_j^m	I_j^u	I_j
	T_j^m	T_j^u	S_j

In our approach the parameter δ, defined as $\delta = \frac{N^u}{N^m}$, is important because it allows us to estimate N through the estimation of δ and N^m. We prove next that δ has to do with the ratio of conditional probabilities involving quality of photos and extent of the marks of a whale. We first define some quantities of interest.

Let us assume that the probability of getting a good photo is the same for marked and unmarked individuals. That is a reasonable assumption since it is in agreement with the analyst's procedure protocol to attach rank quality to a photo. The analyst gives a grade to each photo based only on photo quality, not on a whale's extent of marks. Thus the probability of taking a good photo in a given occasion equals the probabilities above. Therefore given that a good photo is taken at time j, the probability of it being from a marked individual is

$$P(M \mid G_j) = \frac{N^m}{N},\ j = 1, \ldots, t$$

and

$$\left(X_j^m, X_j^u, I_j^m, I_j^u,\right) \sim \text{Mult}\left(S_j, P_{MG_j}, P_{UG_j}, P_{MB_j}, P_{UB_j}\right).$$

Thus, according to Lehmann (1959), we have the following results:

1. $X_j^m \mid X_j, N, N^m \sim \text{Binomial}(X_j, \frac{N^m}{N})$.

2. $I_j^m \mid I_j$ and $X_j^m \mid X_j$ are independent.

The above results suggest the development of a Bayesian analysis based only on the total number of good photos, X_j, at the occasion j, for $j = 1, \ldots, t$, assuming that

$$X_j^m \mid X_j, N, N^m, \sim \text{Binomial}(X_j, \frac{N^m}{N}).$$

The relevant term in the log odds ratio for Table 1 is

$$\Delta \;=\; \log\left(\frac{P(U \mid G_j)}{P(M \mid G_j)}\right) = \log\left(\frac{N^u}{N^m}\right) = \log(\delta).$$

We now introduce more notation in order to define the likelihood function for our model based only on good photos. Let Z_{ij} be the indicator informing whether marked whale i had good photos at time j,

$$Z_{ij} = \begin{cases} 1 & \text{if} \quad X_{ij}^m > 0, \\ 0 & \text{if} \quad X_{ij}^m = 0. \end{cases}$$

Let ω_i represent the capture history of a marked whale i. Thus ω_i is a subset of $\{1, \ldots, t\}$ defined by

$$\omega_i = \{j \in \{1, \ldots, t\} \quad \text{s.t.} \quad Z_{ij} = 1\}.$$

Let $I_{i\omega}$ be the indicator informing whether or not marked whale i has capture history ω,

$$I_{i\omega} = \begin{cases} 1 & \text{if } \omega_i = \omega \\ 0 & \text{otherwise,} \end{cases}$$

where ω is any subset of $\{1, \ldots, t\}$ and let u_ω be the number of marked whales with history ω.

Therefore,

$$p_\omega = P(I_{i\omega} = 1 \mid \{X_j^m\}) = \prod_{j \in \omega} P(Z_{ij} = 1 \mid \{X_j^m\}) \prod_{j \notin \omega} P(Z_{ij} = 0 \mid \{X_j^m\})$$

The expression above represents the conditional probability of a whale having capture history ω considering the sampling effort needed to catch them along the sampling seasons. Such effort is accounted into the model by the $\{X_j^m\}$ good photos gotten along the capture experiment. Therefore, the resulting probabilities p_ω are similar to those described by Darroch (1958).

Consider p_{ij} as the capture probability of whale i given that X_j^m good photos of marked individuals have been taken at time j,

$$p_{ij} = P[Z_{ij} = 1 \mid X_j^m] = p_j,$$

we have that

$$n_j = \sum_{j \in \omega} u_\omega \quad \text{and} \quad r = \sum_\omega u_\omega.$$

Finally, Darroch's model based on $\{X_j^m\}$ is used to estimate population size for the marked part of the population and it is given by

$$P[\{u_\omega\} \mid p, N_m, \{X_j^m\}] = \frac{N^m!}{(N^m - r)! \prod_\omega u_\omega!} \prod_{j=1}^{t} p_j^{n_j} (1 - p_j)^{N^m - n_j} \tag{1}$$

In the next section we present the conditional likelihood function for the model we are proposing and develop a Bayesian procedure for estimating N via Gilks/Wild sampling algorithm.

3 The conditional likelihood function based on good photos

The conditional likelihood of $\theta = (\Delta, p, N^m)$, given $\{X_j\}$, is

$$L(\Delta, p, N^m) = P\left(\{u_\omega\}, \{X_j^m\} \mid \{X_j\}, \Delta, p, N^m\right)$$

$$\propto \frac{N^m!}{(N^m - r)!} \prod_{j=1}^{t} p_j^{n_j} (1 - p_j)^{N^m - n_j} \left[\frac{1}{1 + e^\Delta}\right]^{\sum_{j=1}^{t} X_j^m} \left[\frac{e^\Delta}{1 + e^\Delta}\right]^{\sum_{j=1}^{t} X_j^u} \tag{2}$$

This likelihood function is simpler than that formulated by da Silva *et al.* (2000) and suggests drawing the observations of the conditional posterior of Δ independently of the parameters N^m and p.

As in Liao (1999) we consider the following noninformative prior distributions:

$$\Delta \mid (\mu, \sigma^2) \sim N(\mu, \sigma^2)$$

$$\mu \sim N(0, \sigma_\mu^2) \quad \text{with} \quad \sigma_\mu^2 = 10^6$$

$$\sigma^{-2} \sim \text{Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0}{2}\right) \quad \text{with} \quad \frac{\nu_0}{2} = 10^{-3}$$

$$p_j \sim \text{Beta}(a, b), \quad j = 1, \ldots, t \quad \text{and} \quad \pi(N^m) \propto \frac{1}{N^m},$$

for the following specifications of a and b to express vague prior information: (0,0), (1/2,1/2) and (1,1) (see Smith, (1991)).

The full conditional posterior distributions for the Gibbs sampling are presented in two independent stages as follows:

Stage 1:

$$\sigma^{-2} \mid \Delta, \mu, \nu_0 \sim \text{Gamma}\left(\frac{1}{2} + \frac{\nu_0}{2}, (\Delta - \mu)^2 + \nu_0\right) \tag{3}$$

$$\mu \mid \Delta, \sigma^2, \sigma_u^2 \sim \text{N}\left(\frac{\Delta \sigma_u^2}{\sigma^2 + \sigma_u^2}, \frac{\sigma^2 \sigma_u^2}{\sigma^2 + \sigma_u^2}\right) \tag{4}$$

$$\Delta \mid \mu, \sigma^2, N^m, \{X_j^m\}, \{X_j\} \propto \left[\frac{1}{1+e^{\Delta}}\right]^{\sum_{j=1}^{t} X_j^m} \left[\frac{e^{\Delta}}{1+e^{\Delta}}\right]^{\sum_{j=1}^{t} X_j^u} \exp\left[-\frac{1}{2}\left(\frac{\Delta-\mu}{\sigma}\right)^2\right] \quad (5)$$

Stage 2:

$$N^m - r \mid p, \{X_j^m\}, \{X_j\} \sim \text{Neg-Bin}\left(r, 1 - \prod_{j=1}^{t}(1-p_j)\right) \quad (6)$$

$$p_j \mid N^m, \{X_j^m\}, \{X_j\} \sim \text{Beta}\,(n_j + a, N^m - n_j + b) \quad (7)$$

Each iteration of the Gibbs sampler cycles through the vector of parameters θ, which is divided into some subvector components (like p). Each subvector is drawn conditioned on the value of all the others. Since N is expressed as a function of Δ and N^m, its estimated conditional posterior distribution is obtained through the drawn values of these two parameters. Except for Δ, the other conditional posterior distributions described above are standard ones, representing no difficulty in the simulation process. However, drawing Δ requires more sophisticated simulation tools. We used the adaptive rejection sampling method (ARS) by Gilks and Wild (1992) for drawing Δ. The ARS is only valid for log-concave densities which is true for (5). Such class of densities is composed of functions that are bounded above by their tangents, and bounded below by their cords. Tangents and cords are used to perform rejection sampling with squeezing. The ARS public domain Fortran routine developed by Gilks and Wild (1992) was used in the Δ drawing process.

In the next section, we presented a study that aims to compare the performances of the ML and Bayesian methods developed in the previous sections using simulated bowhead whale datasets. Sensitivity analysis to noninformative prior distributions are presented.

4 Comparison of the performances of the ML and Bayesian methods

Our intention here is to examine some summary statistics, interval width and coverage as in Tables 2 and 3 of da Silva *et al.* (2000) in order to help the readers to decide whether the Bayesian approach is an improvement over the log-normal (Burnham *al.*, 1987), symmetric (Williams *at al.*, 1993) and bootstrap intervals (da Silva *et al.*, 2000). da Silva *at al.* (2000), conducted an experiment in which 500 four occasion capture-recapture bowhead whale samples were simulated under each of 5 cases. The occasions were spring and summer of 1985 and 1986. For all the cases were considered a fixed marked population of 1,186 individuals whereas the unmarked population varied from moderate to high. Capture probabilities varied from small to high, and one of the cases allowed for an open population. More specifically, case 0 stands for a closed population whose capture probabilities are small and not too many unmarked individuals are present. Case 1 is analogous to case 0 except that it allows for an open population. Case 2 is also analogous to case 0 except that capture probabilities are high. This case represents the most optimistic one of the five. Cases 3 and 4 are characterized by a population composed primarily of unmarked individuals with high and low capture probabilities respectively. The cases are summarized in Table 1 in da Silva *et al.* (2000) where a detailed description of the simulated data sets can also be found. In order to study the sensitivity of the Bayesian approach to choices of noninformative prior distributions for the capture probabilities, three noninformative beta priors were considered: beta(0,0), beta(1/2,1/2) and beta(1,1).

Table 2 presents summary statistics for the estimated values of N resulting from each of 500 simulated samples. Column 3 of Table 2 gives the mean of the estimates of N. For the likelihood-based estimates, the mean value represents the average over the 500 estimated values of N. For the Bayesian based ones that column gives the average of the posterior medians over the 500 estimated values of N. Since we know the value of the *true* N in each case, for descriptive purposes only, column 4 of Table 2 presents the corresponding ***bias***. Column 5 of Table 2 evaluates the dispersion of the estimated values in each case. For the MLE, column 5 gives the standard error of the 500 estimated values of N while for the Bayesian method it represents the standard deviation of the 500 posterior medians. From Table 2 we can observe that inferences obtained using different reference priors can differ considerably and | Bayes bias |>| MLE bias |. However, the results obtained using prior (0,0) are reasonably in close agree with the

Table 2: *Summary statistics for the estimated values of* $\hat{N}$ *based on 500 simulated samples*

case	method	mean	bias	s.e.
0	MLE	6,770	36	767
	Bayes (0.0,0.0)	6,814	80	762
	Bayes (0.5,0.5)	6,671	-63	710
	Bayes (1.0,1.0)	6,530	-204	690
1	MLE	6,812	78	858
	Bayes (0.0,0.0)	6,866	132	904
	Bayes (0.5,0.5)	6,715	-20	898
	Bayes (1.0,1.0)	6,570	-165	758
2	MLE	6,724	-10	360
	Bayes (0.0,0.0)	6,747	13	359
	Bayes (0.5,0.5)	6,728	-13	360
	Bayes (1.0,1.0)	6,698	-38	350
3	MLE	13,350	62	1,699
	Bayes (0.0,0.0)	13,557	90	1,724
	Bayes (0.5,0.5)	13,257	-211	1,607
	Bayes (1.0,1.0)	12,986	-482	1,533
4	MLE	14,410	942	4,597
	Bayes (0.0,0.0)	14,734	1,266	4,244
	Bayes (0.5,0.5)	13,268	-200	3,654
	Bayes (1.0,1.0)	12,040	-1,428	2,883

ones obtained by MLE. In case 4 the likelihood is concentrated around small capture probabilities and the beta prior (0,0) reinforces this information. As observed by Smith (1991), the beta prior (0,0) gives strong weight to extreme values of capture probabilities. On the other hand the beta prior (1,1) causes the largest negative biases. Using the latter prior, we expect to generate a large proportion of capture probability values that are larger than we would observe in practice, producing underestimates of N. Beta prior (1/2,1/2) seems to be the best choice among the noninformative priors in study since it induces to the smaller values of | bias | allied with not too high variances. Dispersion of the MLE's are are very comparable with the ones obtained when using beta prior (0,0).

Overall the huge observed differences in the estimated biases and variances in cases 3 and 4 provides evidence that, like the ML, the Bayesian method is also highly affected by the presence of too many unmarked individuals combined with low capture probabilities. The results observed for cases 0 and 1 using both ML and Bayesian methods are in reasonable agreement, and they reveal that relaxation of the closed population assumption for the extent of openness allowed in the simulations does not have too much impact in the estimates of N. Both methods have best performance in the ideal case 2.

It is also useful to study coverage probability of confidence and credible intervals. Columns 3 and 4 of Table 3 give the percentage of samples that missed the true value N either on the left or on the right hand side. Column 5 of Table 3 gives mean interval width.

Log-normal intervals (Burnham *et al.*, 1977) use the delta method variance estimate. The parametric bootstrap intervals (da Silva *et al.*, 2000) are based on 3000 bootstrap replications. The last remarks we have made about Table 2 seem to be also true when the aim is to compare confidence and credible intervals over the cases.

Table 3 suggests that the Bayesian intervals of N are asymmetric with a firmer lower than upper bound. This effect is also observed in other works like Raftery (1998a,b) and Raftery *et al.* (1990) using empirical Bayes procedures with visual and acoustic data. It is important to mention that the lower bound is of interest for bowhead whales, since it represents the most pessimistic view

Table 3: *Percentage of 500 samples in which the 95 %-confidence or credible interval (CI) missed the true value N. Average CI widths are also given.*

Case	*Type of CI*	*% miss*		*Mean*
		left	*right*	*CI width*
0	Log-normal	1.2	3.8	2,988
	Percentile bootstrap	0.6	3.0	3,169
	Bayesian (0.0,0.0)	1.2	2.8	3,025
	Bayesian (0.5,0.5)	0.6	4.6	2,911
	Bayesian (1.0,1.0)	0.4	6.0	2,811
1	Log-normal	2.4	3.8	3,102
	Percentile bootstrap	1.6	3.4	3,256
	Bayesian (0.0,0.0)	2.2	2.8	3,090
	Bayesian (0.5,0.5)	1.4	4.2	2,984
	Bayesian (1.0,1.0)	1.4	6.6	2,873
2	Log-normal	2.6	3.2	1,351
	Percentile bootstrap	2.6	3.0	1,392
	Bayesian (0.0,0.0)	2.8	5.2	1,307
	Bayesian (0.5,0.5)	2.2	5.8	1,293
	Bayesian (1.0,1.0)	1.6	6.8	1,284
3	Log-normal	1.8	4.6	5,863
	Percentile bootstrap	2.6	3.4	6,361
	Bayesian (0.0,0.0)	2.2	3.2	6,159
	Bayesian (0.5,0.5)	1.8	6.6	5,906
	Bayesian (1.0,1.0)	1.0	7.8	5,702
4	Log-normal	2.2	3.4	13,740
	Percentile bootstrap	5.6	1.4	18,490
	Bayesian (0.0,0.0)	7.0	3.6	13,243
	Bayesian (0.5,0.5)	2.4	8.0	13,268
	Bayesian (1.0,1.0)	0.4	15.6	9,569

in the light of data (Raftery *et al.*, 1990).

We can observe from Table 3 that credible intervals are in general narrower than percentile bootstrap confidence intervals. This may be due to the impact of a few very large bootstrap-generated values. Some of the credible intervals have coverage probabilities far from the 95% nominal level, ranging from 84.0% to 96.0%, while in the frequentist cases they range from 92.2 to 96.4%. In terms of coverage probabilities, the worst performance was achieved by the credible intervals in case 4: when using beta prior (0,0), around 7% of the left endpoints were larger than N, i.e., the population size was overestimated, when using beta prior (1/2,1/2), around 8% of the right endpoints were smaller than N, i.e. the population size was underestimated. In the case of beta prior (1,1) the coverage probability was 84% with left and right endpoints of 0.4 and 15.6%, respectively, meaning that too many estimated values underestimated N. The small mean CI width associated to this case means that the distribution of the estimated values is highly concentrated around a value much smaller than N.

Compared to the other options, log-normal confidence interval (Burnham, 1987) seems to be best, it produces not only narrow CIs but also coverage probabilities close to the 95% nominal level. In the next section we present an application using actual bowhead whale photo-identification data and compare the results obtained via G/W sampling with the one obtained with the simplest ML model developed by da Silva *et al.* (2000).

5 An application to bowhead whale photo-identi fication data

We begin this section by describing the data in da Silva (2000). 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 data set containing records with the following information:

- Each marked whale has a unique number, but the same unmarked whale could occur in the data set more than once with different numbers.
- Four columns indicate the capture histories of the bowheads, with 1 indi-

cating that the whale was, and 0 that it was not, captured in the sample represented by the column.

- Four columns indicating the number of good photos obtained for each of the captured individuals by sampling occasion.

There were 1,677 records in the data set, 229 belonging to marked individuals, with 16 of the 229 captured on more than one occasion. The data set was analysed using the model proposed in section 3. The results are described in the next section.

5.1 Numerical results using MLE and Bayesian approachs

In this section our aim is to compare the results obtained using the methods discussed in the previous sections with the ones presented by da Silva *et al.*(2000) who estimated N using real bowhead data and their version of ML based-model via bootstrap. Sensitivity analysis is performed using priors beta(0,0), beta(1/2,1/2) and beta(1,1) for the capture probabilities. As in the last section, the Adaptive Rejection Sampling (ARS) was used to draw samples from density (5). For the Bayesian method, Table 4 gives posterior median ($\hat{N}$), and 95% credible intervals for N. The confidence interval is a log-normal one.

Table 4: *Comparing Bayesian and MLE based estimation for N*

method	$\hat{N}$	Intervals (95%)
MLE	7,022	(4,701;12,561)
Bayes(G/W):		
$a = 0.0, b = 0.0$	7,109	(4,746; 11,138)
$a = 0.5, b = 0.5$	6,389	(4,466; 9,704)
$a = 1.0, b = 1.0$	5,935	(3,978; 8,311)

As noted in the simulated data, the results are sensitive to the choice of the noninformative beta prior for the capture probabilities and, based on the simulated data, we speculate that the estimated value using beta prior (0,0) is positively biased while the one got using beta(1,1) is negatively biased. In this fashion, the most reliable Bayesian estimated value of N should be the one got using beta prior (1/2,1/2) since it likely provides a conservative lower bound on population size if the multinomial model holds.

The inferences obtained using the ML and Bayesian methods when prior (0,0) is considered are in remarkable agreement. As we observed in the simulated cases the estimated values obtained by the ML may present a positive bias. It is true though that such bias may be negligible in some favorable situations such as in case 2. As described by da Silva *et al.*, (2000, section 5), this real data set may present some problems since for some of the sampling occasions it was not possible to get a random sample. In the same article the authors show that nonrandom sampling causes estimated values of N to be positively biased and present another model to correct for that. Using bootstrap methods they estimated that such bias should be around 413 (see table 12 from the cited reference). Despite the positive bias in the MLE of N, both estimated values obtained using the ML and using Bayesian with beta prior(1/2,1/2) compare favorably with the ones obtained by Raftery and Zeh (1998) (6,039 (s.e.=1,915) and 7,734 (s.e.=1,450) for 1985 and 1986 respectively), and with the 1985 and 1986 estimates of 6,649 and 6,820 (excluding calves) from the Bayesian synthesis analysis of Givens (personal communication).

It seems that Bayesian estimates got using beta prior (1/2,1/2) are not as affected by the nonrandom sampling as are ML-based ones. In order to address this issue we performed additional studies based on nonrandom sample simulated data whose description can be found in da Silva *et al.* (2000). The population size considered in those simulations was $N = 6,734$ and capture probabilities were low. The simulations were meant to describe a very dramatic nonrandom sampling scenario. Table 5 summarizes the results. When a simple ML model was fit to the simulated data, da Silva *et al.* (2000) got $\overline{\hat{N}}$=8,254, $bias$=1,520 and $s.e.$=1,211. From Table 5 we can observe that the ML and Bayesian (1/2,1/2) estimated N are in close agreement, with the Bayesian (1/2,1/2) being only slighty superior to the ML, but both do equally badly on estimating N due to the large positive biases. We then speculate that the data is not as badly affected by nonrandom sampling as we have thought before, otherwise estimated values of N from the ML and Bayes (1/2,1/2) should be closer as suggested by the simulations.

Table 5: Comparing Bayesian estimates obtained from nonrandom sample simulated data

Prior	$\bar{\hat{N}}$	*bias*	*s.e.*
$a = 0.0, b = 0.0$	8,497	1,763	1,279
$a = 0.5, b = 0.5$	8,235	1,501	1,188
$a = 1.0, b = 1.0$	7,946	1,212	1,090

The convergence of the MCMC procedure was verified by Gelman and Rubin (1992)'s convergence diagnostics available in the software CODA. In order to do the diagnostic, two sequences with 21000 elements were generated using the procedures described above.

6 Discussion

In this paper we compared two approachs, ML and Bayesian methods, to estimate the size of a population when photo-identification data are collected using a capture-recapture sampling. Simulated bowhead whale datasets under several cases have been used to help us evaluate the methods's performance. Such simulated datasets were the same used by da Silva *et al.* (2000). Sensitivity analysis has been performed using noninformative beta priors: beta(0,0), beta(1/2,1/2) and beta(1,1) for the capture probabilities. The Bayesian method is sensitive to the choice of such priors: the use of beta prior (0,0) causes positive bias while beta prior (1,1) causes negative bias. Vague beta prior (1/2,1/2) seems to be the best choice for this kind of data. The ML estimates agree remarkably well with Bayesian estimates obtained using beta prior (0,0). However, both estimates may be positively biased in the case of small capture probabilities. The estimates of N in Table 4 agree with results from simulations discussed in the previous sections. If samples are nonrandom both methods give positively biased estimates of N. For the actual data it seems that the nonrandom sampling does not have a strong effect on the results. Both ML and Bayesian estimates (with the ML-based one corrected for positive bias) should be equally good. Log-normal confidence intervals (see Burnham, 1987) are the best choice for the ML inferences.

When nothing is known about the distribution of the capture probabilities, the use of vague beta priors like the ones considered in this paper is a reasonable option. However, we realize that for the bowhead data it would be nice to have less vague prior information on the capture probabilities. In order to attack this problem, following Smith (1991), we intend to use either an empirical Bayes or Bayes empirical Bayes approaches.

References

Burnham, K.P., and Anderson, D.R., Wright, G.C., Brownie, C. and Pollock, K.H. (1987). Design and analysis methods for fish survival experiments based on release-recapture. *American Fisheries Society*, Monograph 5. Wenatchee, Washinghton.

Cormack, R.M. (1992). Interval estimation for mark-recapture studies of closed populations. *Biometrics* 48, pp. 567-576.

da Silva, C.Q. (1999). Capture-recapture estimation of bowhead whale population size using photo-identification data. Phd. Dissertation, Universtiy of Washinghton, Dept. of Statistics.

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

Darroch, J.H. (1958). The multiple-recapture census. I: Estimation of a closed population. *Biometrika*, 45, pp. 343-359.

Gelman, A., and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences (with discussion), *Statistical Science,* 7, 457-511.

Gilks, W.R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41, pp. 337-348.

Givens, G. (1993). A Bayesian framework and importance sampling methods for synthesizing multiple sources of evidence and uncertainty linked by complex mechanistic model. Doctorate dissertation. University of Washinghton, Dept. of Statistics.

Lehmann, E.L. (1959). Testing Statistical Hypotheses, Wiley.

Liao, J.G. (1999). A hierarchical Bayesian model for combining multiple $2x2$ tables using conditionals likelihoods, *Biometrics*, 55, pp. 268-272.

Raftery , A.E., Givens, G.H, and Zeh, J.H. (1995). Inference from deterministic population dynamics model for bowheads whales. *Journal of the American Statistical Association*, 90, pp. 402-416.

Raftery. A.E. (1998a). Binomial N estimation: A hierarchical Bayes approach. *Biometrika*, 75, pp. 223-228.

Raftery. A.E. (1998b). Analysis of a simple debugging model. *J. Roy. Statist. Soc., ser. C (Appl. Statist.)*, 37, pp. 12-22.

Raftery, A.E., Zeh, J.E., and Yang, Q. (1990). Bayes empirical Bayes interval estimation of bowhead whale, *Balaena mysticetus*, population size base upon the 1986 combined visual and acoustic census off Point Barrow, Alaska. *Rep. Int. Whal. Comm.*, 40, pp. 393-409.

Raftery, A.E., and Zeh, J.E. (1998). Estimating bowhead whale population size and rate of increase from the 1993 census. *Journal of the American Statistical Association*, 93, pp. 451-463.

Rugh, D.J. (1990). Bowhead whales reindentified through aerial photography near Point Barrow. Rep. International Whaling Commission (special issue 12), pp. 289-294.

Seber, G.A.F. (1982). Estimation of animal abundance, *2nd edition*. London: Griffin.

Smith, P.J. (1991). Bayesian analysis for the multiple capture-recapture model. *Biometrika*, 78, pp 399-407.

Williams, J.A., Dawson, S.M. and Slooten, E. (1993). Distribution of bottlenosed dolphins (*Tursiops truncatus*) in Doubtful Sound, New Zealand. *Can. J. Zool.*, 71, 2080-2088.