PRELIMINARY ABUNDANCE ESTIMATES OF HUMPBACK WHALES BREEDING OFF ABROLHOS BANK, BAHIA, BRAZIL

C.B.P.Bethlem[1], P.G. Kinas[2], M.H. Engel[3]

1-FURG, Graduate Course in Biological Oceanography (pgobcbpb@super.furg.br), P.O.Box 474, Rio Grande, 96201 900, RS,Brazil

2-FURG, Depatment of Mathematics (dmtkinas@super.furg.br), P.O.Box 474, Rio Grande, 96201 900, RS,Brazil

3-Instituto Baleia Jubarte, Praia do Kitongo,s/n, Caravelas, Bahia, 45990 000,Brazil.

Abstract

Flukes of humpback whales (*Megaptera novaeangliae*) are been photographed at Abrolhos Marine National Park, during their reproductive season (july to november) since 1989. If we consider a succesfull first identification as "marking" and resightings as "recaptures", then these data can be used in mark-recapture models to estimate abundance (N) and related parameters. We provide in-season estimates of N for years 1992 to 1994 assuming closed population models. The total number of sightings in the three consecutive years were 22, 34 and 61, of which null, three and two were resightings. Because of sparse data and few resightings, classical methods were avoided and a Bayesian viewpoint was taken instead. Firstly we applied a model that allows for stochastic variation in fluking behaviour. Using information provided by field personel we defined the required prior distributions. The 50% highest posterior density estimates, HPD(50%), were [581,771] in 1992, [733,930] in 1993, and [1660,1900] in 1994. Secondly we

applied a model that assumes constant fluking behaviour and taking each sighting as a sampling occasion of unit size. This resulted in an HPD(50%) of [109,230] in 1993, and [632,1333] in 1994. The first model gives higher and more precise estimates when compared to the second model. But, since available data are not very informative, prior specification is critical and therefore the reported estimates should be taken carefully. Independent estimates of abundance from aerial census or fixed point observations are strongly reccomended and could reduce current model uncertainties. Some of these alternative estimates are planned for the upcoming season of 1996.

Introduction

Humpback whales (*Megaptera novaeangliae*), migrate yearly from their polar feeding concentrations to wintering grounds at tropical coasts, using the shallow and warm waters for breeding.

One of these reproductive aggregations can be found at the Abrolhos Bank, off southernmost Bahia. The area holds peculiar reef formations and many endemic coral species, as well as an exuberant fauna. It was established as a Marine National Park by the federal government in 1983. The Park area includes the Abrolhos archipelago with five islands and the reefs of *Parcel dos Abrolhos* and *Timbebas* (IBAMA, 1991).

Visitation to the Park has grown from 850 persons/year in 1989 to 4000 persons/year in 1994 (Souza, 1995) and it is still increasing. The opening of a tourist season during winter for whale watching has started to rise some worries of whether or not it would be affecting the whales on their reproductive concentrations.

Photo-identification studies of the ventral patterns of the whales' flukes is being carried out since 1989 to evaluate residence time, but consistent efforts have actually started in 1992, with the need for a better understanding of the whales' population dynamics.

A study on abundance estimates and related parameters, using standardized photo-identification and mark-recapture techniques (Katona & Beard, 1990) started in 1995, when sampling efforts have been intensified in order to increase resighting rates and to obtain more precise estimates (Hammond, 1986).

The results presented here refer to the years of 1992 to 1994, and they represent preliminary estimates of abundance as the authors regard it as a pilot study.

Methods

Abrolhos Bank stands over an enlargement of the continental shelf, off southernmost state of Bahia, Brazil, at the limits of the coordinates 17°20' to 18°20'S and 038° to 039° W (Fig. 1).

Figure 1: Map of the Area

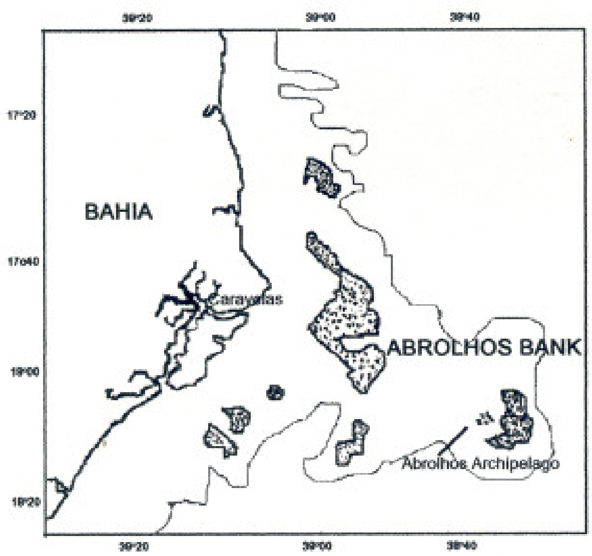

Photo-identification cruises departed from the coastal city of Caravelas, heading to the Abrolhos archipelago. The cruises were continuous throughout the season. The schedule consisted in staying three days at sea, overnight at the archipelago and a two days stop at land for reloading.

Alternate routes were followed, covering the Bank area, guided by sighting cues from the whales (nasal sprays, fluke or dorsal exposition).

The whales were approached and photographed with cameras equipped with 300mm lenses and motordrives. Black and white films ASA200 and colored ASA100 films were used. Data on weather conditions, sampling time, date, group composition and the number of the photographs were recorded. Their geographic position was plotted with a GPS (Global Positioning System).

Photographic analysis

Field records and photographs were carefully analyzed in order to identify individual humpback whales.

Fluke photographs received a degree from one to five, regarding the proportion of fluke exposition and the quality of the photograph; e.g. grade five meaning a hundred percent of exposition and a clear distinction of the patterns (Baker *et al.*,1985). Photographs graded over three received a number of reference as they were included at the *Humpback Whales Institute* catalogue. Those graded one and two were excluded as the individuals could not be identified.

The whales' flukes were categorised as described by Mizroch *et al.*(1990), in 10 stylised generic pattern groups, regarding the amount and disposition of black and white on the leading and trailing edges of the flukes as well as the characteristics of the medial line (broken, solid or absent).

The matching was also done with computer software specific for editing digitized photographs, were brightness, contrast and zoom effects were used to evidence the patterns. All the photographs and data related to it were inserted to the computer.

Abundance Estimation

Data from years 1989 to 1991 were used to observe residence time but were not used in mark-recapture models. The results described here refer to years 1992 to 1994 where sampling efforts became continuous throughout the reproductive season (July through November) and data were sufficient to produce in-season estimates of abundance.

A total number of 112 individuals were identified from years 1992 to 1994. The total number of resightings were five.

Here are in-season estimates of the population size N, for years 1992, 1993, and 1994. Let m denote the total number of sightings in a given year of which w ($w \leq n$) refer to distinct animals. The total searching time in hours is t. The observed values are given in Table 1.

Table 1: Total number of sightings (*m*) and of distinct individuals of humpbacks (*w*), with total sampling effort (*t*) in hours, for years 1992 to 1994.

Year	m	w	t (hours)
1992	22	22	285.30
1993	34	31	327.45
1994	61	59	289.35

There are no distinct sampling occasions in the sense of multiple recapture closed population models (Otis *et al.*, 1978). Therefore, each sighting is taken as a sampling occasion of unit size.

To estimate *N*, we could use the procedure of Seber (1982). This requires the assumption of constant capture probability *p*; that is, model M_0 of Otis *et al.* (1978). Because the data are sparse and resightings are few - null, three and two for years 1992, 1993 and 1994 - the authors use a Bayesian approach instead.

Two different models are considered. Model 1 is a Bayesian alternative to model M_h of Otis *et al.* (1978), as it allows for stochastic variation in fluking behaviour. Model 2 is a Bayesian version of model M_0 and it assumes constant fluking behaviour.

Model 1

This model was proposed by Yoshida et al. (1995). They assume resightings as Poisson processes. The model requires a prior Gamma distribution for the average time between resightings (λ) [lambda] and a discrete uniform prior for *N*. There is a relation between λ and the probability of seeing an animal at least once, *p*. This relation is $p = 1 - e^{-\lambda t}$.

For *p* we assume a Beta ($\alpha_p = 1.8$; $\beta_p = 16.2$) distribution and for *N* we define a non-informative (improper) prior on the positive integers; that is $\pi(N) = 1$, for $N \in \{1, 2, 3, ...\}$. A sample of 1000 values for *p* taken from the prior distribution is displayed in Figure 2 together with the true underlying density; this distribution has mean 0.10, mode 0.05 and a coeficient of variation of 69%. These values were used to determine empirical parameters for the Gamma prior for the less intuitive parameter λ.

The posterior distributions for *N* are displayed by the bold lines in Figure 3. Some posterior estimates are given in Table 2. Highest posterior density (HPD) estimates are a Bayesian version of confidence intervals. For instance, in 1993 the 50% most likely values for *N*, denoted HPD(50%), are given by the set of integers ranging from 733 to 930. Other HPD's are interpreted similarly.

Table 2: In - season posterior estimates of population size (*N*) for years 1992 to 1994.

Year	Mode	Mean	S.D.	HPD(50%)	HPD(90%)
1992	672	702	138	[581, 771]	[466, 933]
1993	828	854	141	[733, 930]	[641, 1093]
1994	1777	1725	196	[1660, 1900]	[1496, 2000]
1994*	1777	1725	160	[1142,1355]	[1003,1525]

* refers to changing of a prior and it is discussed further (*see* Robustness)

Model 2

This model was proposed by Leite and Pereira (1990). It only requires a proper prior distribution for the population size N. We define this distribution to be discrete uniform on the integers from 1 to some largest N^*. That is $\pi(N) = (N^*)^{-1}$ for $N \in \{1, 2, 3, ..., N^*\}$. In our analysis we felt that $N^* = 2000$, was large enough to assume $\pi(N>2000) = 0$.

The posterior distributions for N are given by the broken lines in Figure 3 and some posterior estimates are summarized in the table below. Because of no resightings in 1992, the estimate is unreliable and was therefore excluded.

Table 3: In- season posterior estimates of population size (*N*) for years 1993 and 1994.

Year	Mode	Mean	S.D.	HPD(50%)	HPD(90%)
1993	176	407	327	[109, 320]	[69, 836]
1994	895	1154	444	[632, 1333]	[494, 1913]

Discussion

Comparing the estimates obtained by both models:

(*i*) If we use the posterior mode as the most likely estimate for *N*, model 1 consistently gives higher estimates than model 2. The same trend, although less pronounced, is also observed for posterior means. Otis *et al.* (1978), using simulation studies, found that model M_0 always underestimates *N* when M_h is the true model. Although more analysis is needed, our results might suggest a Bayesian version of this phenomenon. As no model is an adequate description of an experiment, the authors recommend that independent estimates be obtained in order to test the degree of heterogeneity inherent at the samples.

(*ii*) Both models suggest substantial changes in the population size between years. Using the overlapping regions of the HPD(90%) for models 1 and 2 the data suggest that there has been an increase in population size from years 1992 to 1994. Such trend could be verified by continuing sampling the population over the time, to confirm whether population size is actually increasing or whether it would be a biased consequence of intensifying sampling efforts.

(*iii*) The posterior distribution for model 1 gives more precise estimates of *N* when compared to the results of model 2. The first model gives higher and more precise estimates when compared to the second model. It could reflect the importance of the heterogeneity of capture probabilities in capture - recapture experiments.

Robustness

Since the available data are not very informative, the influence of the prior distribution on the resulting estimates is expected to be strong and was no surprise. Its implications are discussed next.

If in model 1 the prior Beta distribution for capture probability *p* is chosen to have mean 0.15 and mode 0.10 (i.e.: α_p = 2.4; β_p = 13.6; coefficient of variation 58%) then the posterior distribution will change. The authors considered this change in the prior (p) in year 1994* (*see Table 3*) once sampling efforts were increased (one more vessel was used in photo-identification cruises). We would have mode 1245, mean 1266, standard deviation 160, HPD(50%): [1142, 1355] and HPD(90%): [1003,1525] as given in Table 2.

The sensitivity of the estimates displays the scientific uncertainty associated to current knowledge of N. It further indicates that, at the present stage, good judgment about values of p is critical for the resulting estimates of N and has been validated by those involved in the field work.

In model 2, there are weaker requirements for prior specification; only N^* needs to be given. However there is a price to pay: the (unreasonable) assumption of equal capturability (constant p) for all animals. Further, changing the maximum N^* from 2000 to 3000 will have a strong effect on the upper tail of the distribution; hence, posterior mean and variance respond to this change. But the mode remains unaffected by such changes. This suggests to use the mode and the HPD(50%) to make inter annual comparisons of N.

From the data at hand we can not perform any tests of the models. While recognizing this weakness, we know of no alternative to avoid this limitation. Simulation studies are necessary to understand the performance and biases of the models presented here. The authors plan to provide some results in the near future.

To reduce the uncertainties in the reported estimates of *N*, caused in part by the sensitivity to prior distributions, we suggest data acquisition for independent estimates of abundance. For instance, aerial census or fixed point observations could be used to obtain data to use with distance models (Buckland *et al.* 1995). They are a helpful device to check the estimates obtained in this report. Some of these procedures are being planned for the winter season of 1996.

Based upon this pilot study the authors had improved field methods for year 1995, expecting to obtain larger sample sizes and consequently reduce the role of heterogeneity in the precision of the estimates.

References

Baker,C.S., Herman, L.M., Perry, A., Lawton, W.S., Straley, J.M., Straley, J.H.1985. Population characteristics and migration of summer and late season humpback whales (*Megaptera novaeangliae*) in southeastern Alaska. *Mar. Mamm. Sci.* 1(4): 304-23.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L. 1993. *Distance Sampling: Estimating Abundance of Biological Populations.* Chapman and Hall, 446pp.

Hammond,P.S.1986.Estimating the size of naturally marked whale populations using capture-recapture techniques. In: *Rep. Int. Whal. Commn.* (Special Issue 8):253-282

Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis - IBAMA/ Fundação Pró-Natureza - Funatura. 1991. *Plano de Manejo do Parque Nacional Marinho dos Abrolhos.* Aracruz Celulose, Brasília. 96 pp.

Katona, S. K. & Beard, J. A. 1990. Population size, migrations and feeding aggregations of the humpback whale (*Megaptera novaeangliae*) in the western North Atlantic Ocean. In: *Rep. Int. Whal. Commn.* (Special Issue 12):295-306.

Leite, J.G., Pereira, C.A.B. 1990. Bayes Estimation of the Size of a Finite Population: Capture/ Recapture Sequential Data. *International Statistical Review* .58: 201-213.

Mizroch, S.A., Beard, J.A.,Lynde, M.1990. Computer assisted photo-identification of humpback whales. In: *Rep. Int. Whal. Commn.* (Special Issue 12):63-70.

Otis, D.L.; Burnham, K.P.; White G.C.; Anderson, D.R.1978. *Statistical Inference from Capture Data on Closed Animal Populations. Wildlife Monographs.* The Wildlife Society, Inc. 133 pp.

Seber, G.A.F. (1982). *The estimation of animal abundance and related parameters.* Second ed., MacMillan Publishing Co., New York. 654 pp.

Souza, C.A.A.S..1995. *Ecoturismo na Bahia: Estudo Analítico.* Edição Sebrae, Salvador.44pp.

Yoshida, O.S.; Leite, J.G.; Bolfarine, H. 1995. *Bayes estimation of the number of component processes of a superimposed process.* São Paulo, IME - USP 19p. (RT-MAE 9518).

Figure 2

A sample of 1000 values for *p* taken from the prior distribution is displayed together with the true underlying density; this distribution has mean 0.10, mode 0.05 and a coeficient of variation of 69%.

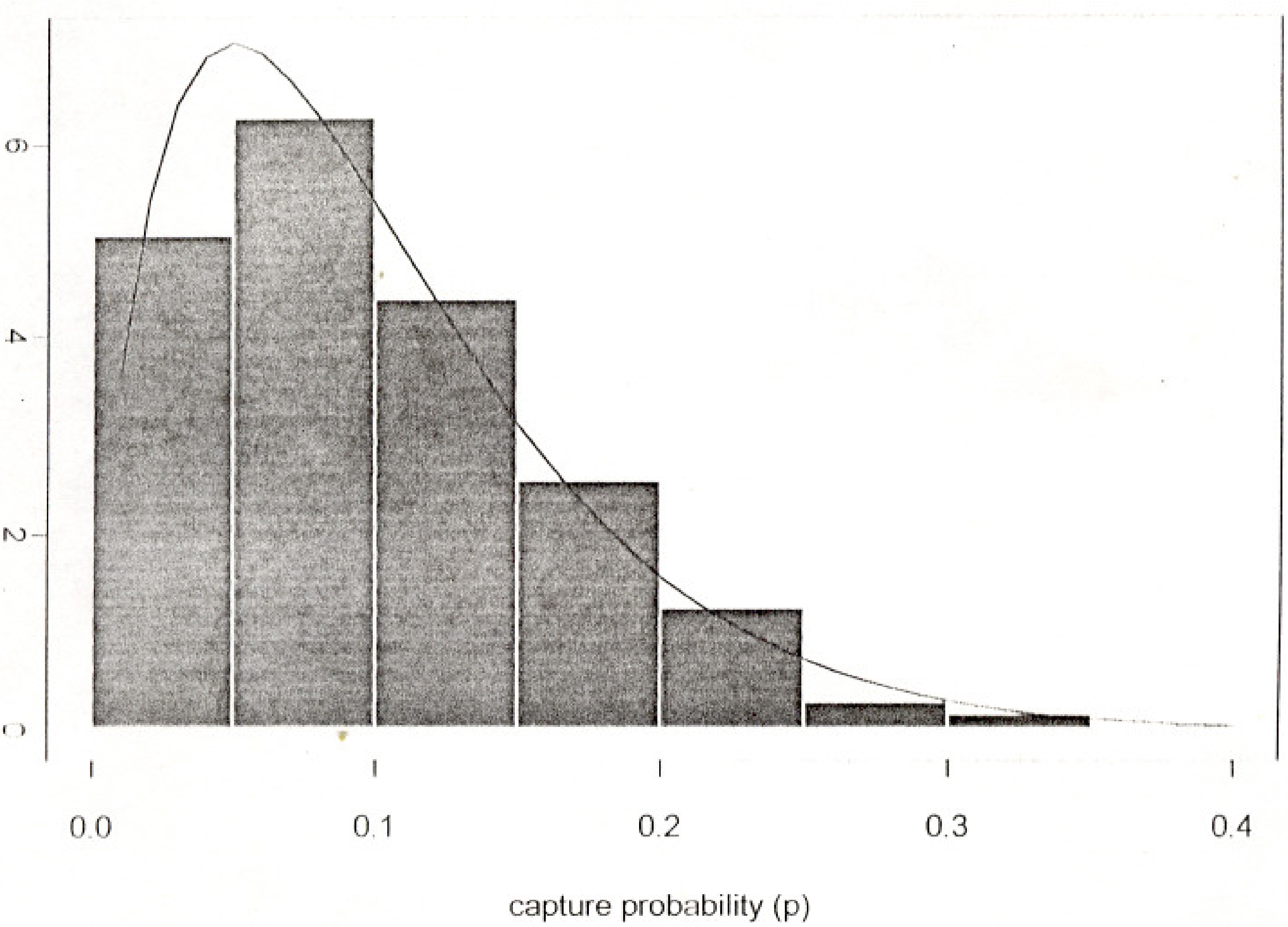

Figure 3

The posterior distributions for *N* are displayed by the bold lines for Model 1 and thin lines for Model 2.