(Draft version 3.0 - 16/03/2006)

INTEGRATING PHOTO-IDENTIFICATION AND AERIAL SURVEYS TO ESTIMATE G(0) FOR HUMPBACK WHALES (*Megaptera novaeangliae*) IN THE BRAZILIAN BREEDING GROUND.

Kinas, P.G. [2] ; Andriolo, A. [3]; Engel, M.H. [1]

[1]*Projeto Baleia Jubarte/Instituto Baleia Jubarte, Rua Barão do Rio Branco, 26 - Caravelas, BA - Brasil.*
[2]*Lab. de Estatística-DMat, Fundação Universidade Federal do Rio Grande, , Rio Grande, RS – Brasil.*
[3] *Depto. de Zoologia, Instituto de Ciências Biológicas da Universidade Federal de Juiz de Fora, MG, Brasil,*

Contact author:
Paul G. Kinas
Laboratório de Estatística – DMat -FURG
Av. Itália Km. 8
96201-900, Rio Grande, RS, Brasil.
Phone: (53) 3233 6496
Fax: (53) 3233 6604
E-mail: dmtkinas@furg.br

ABSTRACT

In ordinary distance sampling the detection probability on the trackline, $g(0)$, is assumed to be 1. However, when dealing with aerial surveys of aquatic mammals, this assumptions is unrealistic as animals on the trackline but not close to the surface remain undetected. Therefore, in order to obtain abundance estimates with distance sampling, a reliable estimate of $g(0)$ is needed. Various possibilities to obtain such estimates have been proposed, but none applies directly to the aerial survey data available for humpbacks (*Megaptera novaeangliae*) in the Brazilian breeding ground. We obtain an estimate based on the ratio between (*i*) a population size estimate from distance sampling assuming $g(0)$ = 1 and (*ii*) an independent population size estimate based on mark-recapture methods. Precision of this estimate is estimated with a bootstrap sample. The estimate $\hat{g}(0) = 0.432$ (*s.e.* = 0.108) is the average of the bootstrap sample and accounts for availability and perception bias combined.

Keywords: aerial survey, line transect sampling, mark-recapture, abundance, visibility bias, detection probability, *Megaptera novaeangliae*,

INTRODUCTION

In order to obtain abundance estimates using line transect sampling (Buckland *et al.* 1993) it is necessary to know the detection probability on the trackline, g(0). The usual assumption that all animals on the trackline are detected, *g*(0) = 1, is not appropriate for aerial surveys of humpback whales, since animals re not visible if they are not close enough to the surface. A true value of *g*(0) smaller than 1, when ignored, leads to visibility bias (Marsh & Sinclair, 1989) resulting in underestimated population abundance.

According to Marsh & Sinclair (1989) visibility bias for marine mammals can occur when animals are not close enough to the surface to be seen (availability bias) or when animals are visible but missed to a variety of other reasons like sun glare or observer fatigue (perception bias) . While perception bias can be minimized by a careful field protocol, availability bias is harder to control since it is determined by environmental conditions and animal behavior. No attempt will be made here to distinguish among these two causes of bias.

Pollock & Kendall (1987) review estimation procedures for visibility bias in aerial surveys. They propose methods that combine ground counts with aerial surveys in general but are not concerned with the specifics of the detection probability g(0) in distance sampling. Efforts to estimate g(0) directly include the use of breath rate data (e.g. Barlow et al. 1988,). Alternatively, a more elaborate procedure uses a double-platform in which two observers search the same covered regions independently and the data are used in a mixture of distance sampling and mark-recapture methods (Borchers et al. 2002)

Limitations imposed by the logistics of our aerial-surveys, preclude the use of estimation procedures of *g*(0) which are based on double-platforms. Breath-rate data have been used before (Andriolo et al. accepted) but were based only on the observed behavior of whales around the Abrolhos bank.

In order to overcome this difficulty, we propose an estimate of *g*(0) which integrates the data on photo-id of Freitas *et al.* (2004) with the aerial survey data collected between 2002 and 2005 and described in Andriolo et *al.* (2006, paper presented at this meeting). Our proposal is conceptually in line with Pollock & Kendall (1987) by using mark-recapture methods to replace a ground count with a population size estimate. The results are further used in Andriolo *et al.* (2006, this meeting) to produce *g*(0)-corrected abundance estimates in distance sampling.

METHODS

An estimate for *g*(0) and of its standard error are obtained from a bootstrap sample which is build in four stages as follows.

The first stage consists in a bootstrap sample of population size at the breeding ground in year 2000 from photo-id data. This bootstrap sample, denoted $N_{0,mr}^{(j)}$ for $j = 1, \ldots, J$, is obtained by sampling J positive integers with probability proportional to the likelihood function for mark-recapture model HG-r (model 3) in Freitas *et al*. (2004). That model estimates population size in year 2000 together with an annual population growth parameter *r* for years 1996 to 2000.

The second stage starts with population size estimates for years 2002 to 2005 using distance methods (Buckland *et al.*, 1993) and the assumption that $g(0) = 1$ in the aerial survey data of Andriolo *et al.* (2006, this meeting). A bootstrap sample of *J* population size estimates in year 2002, denoted $N_{2.dist}^{(j)}$ for j = 1,… *J*, is obtained from a Normal distribution with mean and standard deviation set at the corresponding estimates obtained from distance method. The population size estimates for years 2002 to 2005 are further used to calculate annual growth rates for years *t* from 2002 to 2004 as follows

$$h_t = \frac{N_{t+1} - N_t}{N_t}. \qquad \text{(eq. 1)}$$

The mean and standard error of the logistic transformations of h_t are taken as parameters of a Normal distribution to generate a random sample denoted $y^{(j)}$ for j = 1, …, *J*. The corresponding bootstrap sample of average annual population growth rate is obtained by back-transformation.

$$h^{(j)} = \frac{e^{y^{(j)}}}{1 + e^{y^{(j)}}} \qquad \text{(eq. 2)}$$

The third stage consists in a two-year forward projection of the bootstrap sample for year 2000 which was obtained previously (first stage). Given the *j*-th pair of estimates $N_{0.mr}^{(j)}$ and $h^{(j)}$, the projected population size estimate in year 2002 is obtained simply as

$$N_{2.mr}^{(j)} = N_{0.mr}^{(j)}(1 + h^{(j)})^2 \qquad \text{(eq. 3)}$$

The fourth and last stage is the bootstrap sample of *J* estimates for $g(0)$ defined as the ratio between population size estimates for year 2002 from distance sampling with the assumption that $g(0) = 1$ (stage two) and the corresponding estimate derived from mark-recapture (stage three). That is,

$$g^{(j)}(0) = \frac{N_{2.dist}^{(j)}}{N_{2.mr}^{(j)}}. \qquad \text{(eq. 4)}$$

This bootstrap sample is used to obtain a point estimate, standard error and confidence interval for $g(0)$.

RESULTS

The population size estimate for year 2000, obtained from distance sampling and the assumption $g(0) = 1$, resulted in $\hat{N}_{2.dist} = 2305$ humpbacks with standard error *s.e.* = 308.9. This model uses the combined four years of data to fit the same detection function of Andriolo *et al.* (2006, this meeting) since the distance data are the same. These estimates are used as parameters in a Normal distribution to generate the bootstrap sample $N_{2.dist}^{(j)}$ for $j = 1, \ldots, J$.

The estimated population annual growth rates (*h*) between years 2002-03, 2003-04 and 2004-05 are 0.102, 0.424 and 0.077, respectively. After the logistic transformation, these resulted in mean -1.66 and standard error 0.682 which were used as parameters in a Normal distribution to generate $y^{(j)}$´s. The back-transformed bootstrap sample for growth rate h shown in Figure 1, has a mean of 0.178 (mode = 0.125, median = 0.158) and 95% confidence interval (0.051; 0.425).

Finally, the bootstrap sample for $g(0)$ displayed in Figure 2 has mean 0.432 (mode = 0.398, median = 0.423), standard error 0.108 and 95% confidence interval (0.243; 0.666).

DISCUSSION

The estimate for $g(0)$ which we obtained accounts for availability and perception bias combined. This might help to explain why it turned out to be smaller than a previous estimate of 0.67 (± 0.15) which was based on breath rate data (Andriolo *et al. accepted*) since this would only account for availability bias.

Baird et al. (2000) studying the diving behavior for humpback whales off the west side of Maui, Hawaii, found that animals remained roughly 40% of the time within the upper most 10m of depth. Furthermore, in the region where our aerial survey was conducted, animals up to this depth are available to observation on the trackline while each point on the trackline remains within visual range for about 34 seconds. Connecting both information and assuming that humpback whales diving behavior at the Brazilian breeding ground is similar to what was observed by Baird et al. (2000) in the north Pacific, our estimate is quite reasonable.

In Freitas *et al.* (2004) various models were applied. The choice for HG-r (model 3) to obtain the bootstrap sample for mark-recapture population size estimate, was based on two considerations. First, this model incorporates an annual growth rate that was obtained from an entirely different data set and could be checked against the estimate obtained from distance sampling. The estimate of 0.26 (from foto-id) is comfortably within the range established by the confidence interval obtained for *h* (from aerial survey). Second, it gives a population size estimate for year 2000 which is closest in time to year 2002 so that only two years of forward projection is necessary.

According to Clapham et al. (2001), and based on the range of life history parameters observed for several humpback whale populations, the maximum growth rate for the species is 0.126. This value was used by Zerbini (SC/56/SH17) as the upper limit for the prior distribution for the maximum net recruitment rate r_{max}. For a population not subjected to harvest, the relation between r_{max} and the growth rate h_t as defined in (eq. 1) is

$$h_t = r_{\max}\left[1-\left(\frac{N_t}{K}\right)^z\right] \qquad \text{(eq. 5)}$$

where: K is the pre-exploitation size, and z a shape parameter which determines the population size where productivity is maximum. The term in brackets approaches one (from below) as $(N_t K^{-1}) \rightarrow 0$; therefore $h_t \leq r_{\max}$. For instance, if z = 2.39 (used by Zerbini) and taking $N_t K^{-1}$ equal to 0.1 and 0.3, the terms in brackets are 0.9959 and 0.9437, respectively. We assume that current population size is still much smaller than the pre-exploitation size K and make no further distinction between h_t and $r_{\max}$ to simplify our discussion.

The observed population growth rate of 0.424 between years 2003 and 2004 used in our analysis, is biologically implausible according to Clapham et al. (2001). To examine the impact of this value on the final estimate of *g*(0), we replaced it by 0.126 and rerun the analysis. The resulting bootstrap sample for *h* has mean 0.101 (*se* = 0.014) and represents a considerable change with respect to the results displayed in Figure 1. However the impact on *g*(0) is much smaller, resulting in mean 0.49 (mode = 0.45, median = 0.48) standard error 0.095 and 95% confidence interval (0.317; 0.692). This result would imply a reduction of approximately 12% in the *g*(0)-corrected population estimate of Andriolo *et al.* (2006, this meeting).

However, while considerations of the last paragraph suggest slightly larger value for *g*(0), the possibility that the denominator in (eq. 4) is a negatively biased estimate of population size, points into the opposite direction. This possibility is supported by two arguments. First, whenever there is heterogeneity in detection probability in mark-recapture, which is not accounted for by the model, simulation studies have shown that estimates usually are negatively biased (Borchers *et al.* 2002). This heterogeneity is certainly present in our photo-id data as movement pattern around Abrolhos bank and tail exposition is very different among animals. Second, the photo-id data refer to the population which visits the area around Abrolhos bank at least once during the wintering season. Although this might represent a large percentage of the humpback population visiting the Brazilian breeding ground, current knowledge on movement pattern of these animals is insufficient to know how large. Animal movements was documented for the Abrolhos Bank by Zerbini (2006 in press), and considerations about the implications in photo-ID were discussed. However the sample was small and further studies to document animal movements should be encouraged.

The combination of bootstrap samples for all components used in the estimation of *g*(0), allows for the appropriate modeling of uncertainty propagation. Hence, the bootstrap sample for *g*(0) displayed in Figure 2 gives a full account of the range of possibilities. Since the arguments in favor of a slightly large *g*(0) are balanced with arguments leading to the opposite direction, we conclude that further efforts should be directed towards studies to help reduce these uncertainties.

ACKNOWLEDGEMENTS

This study was supported by Aracruz Celulose S.A, Veracel Celulose and Petróleo Brasileiro S.A (PETROBRAS).

BIBLIOGRAPHY

Andriolo, A., Martins, C.C.A., Engel, M.H., Pizzorno, J.L., Más-Rosa, S., Freitas, A.C., Morete, M.E. and Kinas, P.G. (accepted). The first aerial surveys of humpback whale (*Megaptera novaeangliae*) to estimate abundance in the breeding ground, Brazil. Journal of Cetacean Research and Management.

Baird, R.W.; Ligon, A.D.; Hooker, S.K. 2000. Sub-surface and nighttime behavior of humpback whale off Mauis, Hawaii: a preliminary report. Report prepared under Order No. AB133F-03-SE-0986 to the Hawai‘i Wildlife Fund, Volcano, HI, Southwest Fisheries Science Center, National Marine Fisheries Service, La Jolla, USA, 15pp.

Barlow, J.; Oliver, C.W.; Jackson, T.D.; Taylor, B.L.1988. Harbor porpoise, Phocoena phocoena, abundance estimation of California, Oregon, and Washington: aerial survey. *Fish. Bull*. 86: 433-444.

Borchers, D.L.; Buckland, S.T.; Zucchini, W. 2002. *Estimating Animal Anundance: closed populations*. Springer. 313p.

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

Clapham, P.; Robbins, J.; Brown, M.; Wade, P.; Findlay, K. 2001. Appendix 5 – A note on plausible rates of population growth in humpback whales. *J. Cetacean Res. Manage.* 3 *(Suppl)*: 196-197.

Freitas, A.C; Kinas, P.G.; Martins, C.C.A.; Engel, M. H. 2004. Abundance of humpback whales on the Abrolhos Bank wintering ground, Brazil. *J. Cetacean Res. Manage*. 6: 225-230.

Marsh, H. & Sinclair, D.F. 1989. Corecting for visibility bias in strip transect aerial surveys of aquatic fauna. *J. Wildl. Manage*. 53: 1017-1024.

Pollock, K.H. & Kendall, W.L. 1987. Visibility bias in aerial surveys: a review of estimation procedures. *J. Wildl. Manage*. 51: 502-510.

Zerbini, A N, Andriolo, A, Heide-Jorgensen, M A, Pizzorno, J L, Maia, Y G, VanBlaricom, G R, DeMaster, D G, Simoes-Lopes, P, Moreira, S, and Bethlem, C. *In press* Movements of satellite monitored humpback whales (*Megaptera novaeangliae*) in the Southwest Atlantic Ocean. *Marine Ecology Progress Serie.*

FIGURES

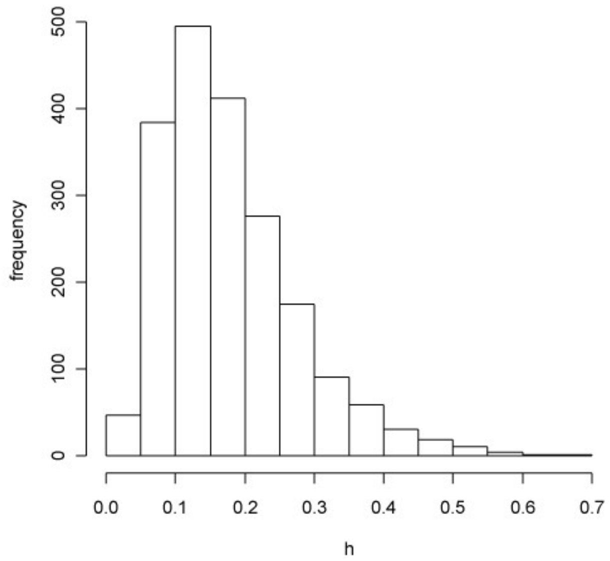

Figure 1: Boostrap sample of average annual population growth rate for humpbacks in the Brazilian breeding ground.

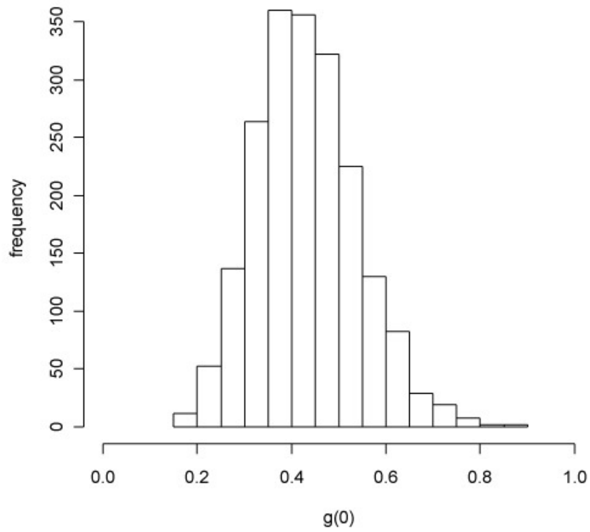

Figure 2: Bootstrap sample for g(0) estimates in aerial surveys for humpbacks in the Brazilian breeding ground.