CAPTURE-RECAPTURE ANALYSES OF HUMPBACK WHALE POPULATION SIZES AND INCREASE RATES: BREEDING SUB-STOCK B1

S.J. Johnston and D.S. Butterworth

Marine Resource Assessment and Management Group (MARAM)
Department of Mathematics and Applied Mathematics
University of Cape Town
Rondebosch, 7701
South Africa

Abstract

Maximum likelihood and Bayesian approaches are applied to estimate parameters of a simple experimental growth model for the humpback whales of breeding sub-stock B1 from photo-ID and genotypic capture-recapture data for all observation sites combined and for Iguela only. The maximum likelihood estimates are generally rather imprecise, and some estimates for the annual growth rate r are demographically unrealistic. Bayesian results, for which a prior restricts its r to a demographically plausible range of [0; 0.106], are more reliable and suggest a population of size 7600 [95% PI: 6600-8800] in 2003, with r at 8.5% [95% PI: 3.7%-10.4%]. These results, are however, intended primarily as illustrative of the methodologies, and to facilitate further discussion and model refinement. This will need, *inter alia*, to consider the appropriateness of aggregating data from different sites.

Introduction

This document reports results from fitting an exponential growth model to the capture-recapture data available for the B1 breeding sub-stock using both maximum likelihood and Bayesian estimation. Fewer scenarios are examined than for the similar analyses for the breeding stock C in Johnston and Butterworth (2008). The reason is that this paper is intended to be illustrative only, and to provide a basis from which to develop further runs and refinement of the approach during sub-committee discussions at the IWC Scientific Committee meeting.

Data

The capture-recapture data used here are those reported in Collins *et al*. (2008). These consist of both photo-ID and genotypic mark-recapture data from Gabon. The data span the period 2000-2006 and are reproduced in Appendix 1. The data are reported both for all data sites, and for Iguela only (for which the data cover the years 2001-2005 only).

Methods

Maximum Likelihood
The following simple exponential population growth model is considered in conjunction with the capture-recapture data (where "capture" can refer to either photographic or genetic identification):

$$N_{y+1} = N_y + rN_y \tag{1}$$

$$n_y = p_y N_y \tag{2}$$

$$\hat{m}_{y',y} = p_{y'} p_y N_y e^{-M(y'-y)} \tag{3}$$

where: n_y = number of animals captured in year y

$m_{y',y}$ = number of animals captured in year y that were recaptured in year y'

$\hat{m}_{y',y}$ = model predicted number of animals captured in year y that were recaptured in year y'

r = the population growth rate
M = natural mortality rate
p_y = probability animal is seen in year y
N_y = population size in year y.

For each potential recapture cell (ignoring some year recaptures), the likelihood contribution assuming a Poisson distribution is given by:

$$L = \frac{(\hat{m}_{y',y})^{m_{y',y}}}{m_{y',y}!} e^{-\hat{m}_{y',y}} \tag{4}$$

Hence the overall $-ln$L to be minimised is:

$$-lnL = \sum_{y=y_0}^{y_{f-1}} \sum_{y'=y+1}^{y_f} [-m_{y',y} \ln \hat{m}_{y',y} + \hat{m}_{y',y}] + const \tag{5}$$

where: y_0 = the first year of captures

y_f = the last year of captures

and the minimisation is over the estimable parameters r and N_{y_0}. Note that the approach makes allowance for the reduction over time in the numbers of animals potentially recaptured as a result of natural mortality.

Bayesian
Results are produced for a prior on $r \sim$ U[0; 0.106] to take into account Scientific Committee deliberations on demographically plausible bounds for this parameter. The prior for $N(2003)$ is [200; 80 000], i.e. essentially uninformative.

General
A "Reference Case" analysis is one which uses all the data available for a certain data type and sets $M = 0.03$ yr^{-1}. Hessian-based CVs are provided for the maximum likelihood parameter estimates, but to minimise covariance impacts, these are reported rather for N_{2003} than for year y_0 (N_{y_0}). For the Bayesian results, medians plus 95% PIs are reported for r and the N_y and p_y values.

Results for B1

Table 1 reports the MLE results for Breeding Sub-stock B1 (BS B1). Results are presented for both the photo-ID data and the genotypic data, as well as for the use of all the data or Iguela only data.

Figures 1a and b show the population trends as estimated by the Bayesian approach – medians and 95% PIs are shown.

Bayesian results assuming a prior for r of U[0; 0.106] are provided in both Table 2 and Figures 2a and 2b for both photo-ID and genetics information in combination, first for data from all sites, and then for Iguela only. The posterior distributions for r for both cases are shown in Figures 3a and 3b.

Note that Collins *et al.'s* (2008) preferred estimates to provide an upper and lower bound for the population size of B1 using MARK are the following:

MARK photo-ID data Iguela only (2001-2005): 6432 (CV=0.18), and
MARK genetic data Iguela only (2001-2004):   7196 (CV=0.15).

Discussion

The point estimates of r provided by maximum likelihood estimation are unrealistically large when using the photo-ID data. Estimates from genotypic data are more demographically realistic but imprecise.

Results from the Bayesian analyses using both data sets and for which r is appropriately bounded to ensure demographic realism, should be considered the more reliable. Precision is poor for data from Iguela only, with the data updating the prior for r only slightly. However, when the data from all sites are combined, they become more informative with a posterior median for r at 8.5% per annum and clear indications of population increase. Sub-committee discussions should though consider the appropriateness of combining the data from all sites in this way, as that may not be in line with the underlying assumption of the model applied that animals at the various sites mix fully between years.

Acknowledgements

References

Collins, T., Cerchio, S., Pomilla, C., Loo, J., Carvalho, I., Ngouessono, S. and Rosenbaum, H. 2008. Revised abundance estimates for breeding stock B1: Gabon. IWC document, SC/60/SH28.

Johnston. S.J. and Butterworth, D.S. 2008. Capture-recapture analyses of humpback population sizes and increase rates: Breeding stocks C1-C3. IWC document, SC/60/SH37.

Table 1: B1 MLE results - values in brackets are Hessian based CVs.

	Photo-ID data all sites	Genotypic data all sites	Photo-ID data Iguela only	Genotypic data Iguela only
*-ln*L	-22.24	-51.95	-16.43	-31.53
r	0.401 (0.32)	0.133 (0.65)	0.502 (0.48)	0.055 (2.84)
N(2000)	1697	5248		
N(2001)	2378	5944	1508	5407
N(2002)	3331	6732	2266	5706
N(2003)	4667 (0.16)	7625 (0.14)	3404 (0.19)	6022 (0.33)
N(2004)	6538	8636	5114	6355
N(2005)	9159	9781	7683	6707
N(2006)	12830	11078		

Table 2: B1 Bayesian results – medians and 95% PIs are reported. Note the prior for *r* is U[0; 0.106]. The "pho" refers to photo-ID data, and the "gen" refers to the genotypic data.

	Both Photo-ID data and genetic data from all sites	**Both Photo-ID data and genetic data from Iguela only**
r	0.085 [0.037; 0.104]	0.066 [0.011; 0.102]
N(2000)	5998 [5062; 7532]	
N(2001)	6507 [5549; 7896]	4778 [3883; 6028]
N(2002)	7045 [6079; 8325]	5080 [4208; 6239]
N(2003)	7618 [6628; 8835]	5398 [4541; 6522]
N(2004)	8235 [7207; 9457]	5730 [4844; 6892]
N(2005)	8892 [7903; 10214]	6096 [5106; 7395]
N(2006)	9603 [8350; 11089]	
p(2000)	pho: 0.004 [0.003; 0.005] gen: 0.014 [0.011; 0.016]	
p(2001)	pho: 0.017 [0.014; 0.020] gen: 0.024 [0.020; 0.028]	pho: 0.023 [0.018; 0.029] gen: 0.032 [0.026; 0.040]
p(2002)	pho: 0.033 [0.028; 0.038] gen: 0.026 [0.031; 0.042]	pho: 0.028 [0.023; 0.034] gen: 0.033 [0.027; 0.040]
p(2003)	pho: 0.021 [0.018; 0.024] gen: 0.035 [0.031; 0.040]	pho: 0.030 [0.025; 0.035] gen: 0.050 [0.041; 0.059]
p(2004)	pho: 0.017 [0.015; 0.019] gen: 0.023 [0.020; 0.026]	pho: 0.024 [0.020; 0.028] gen: 0.033 [0.027; 0.039]
p(2005)	pho: 0.024 [0.021; 0.028] gen: 0.033 [0.029; 0.038]	pho: 0.020 [0.016; 0.024] gen: 0.022 [0.019; 0.027]
p(2006)	pho: 0.021 [0.018; 0.024] gen: 0.022 [0.019; 0.025]	

Figure 1: BS B1 population maximum likelihood abundance estimates from the capture-recapture analysis compared with abundance estimates from Collins *et al.* (2008).
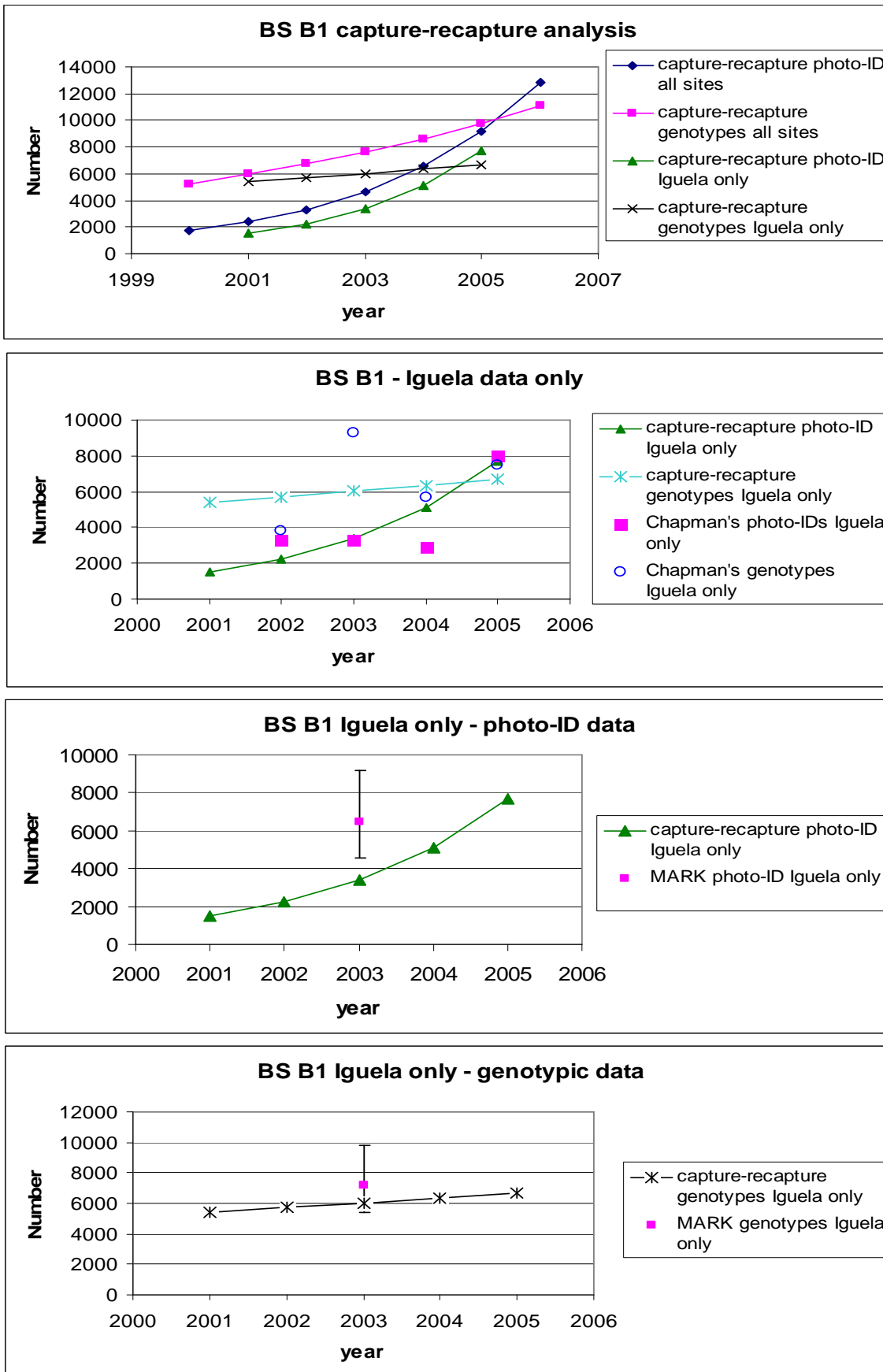
Figure 2a: Bayesian B1 assessment results using both photo-ID + genotypic data for the whole dataset; the prior for *r* is U[0; 0.106].
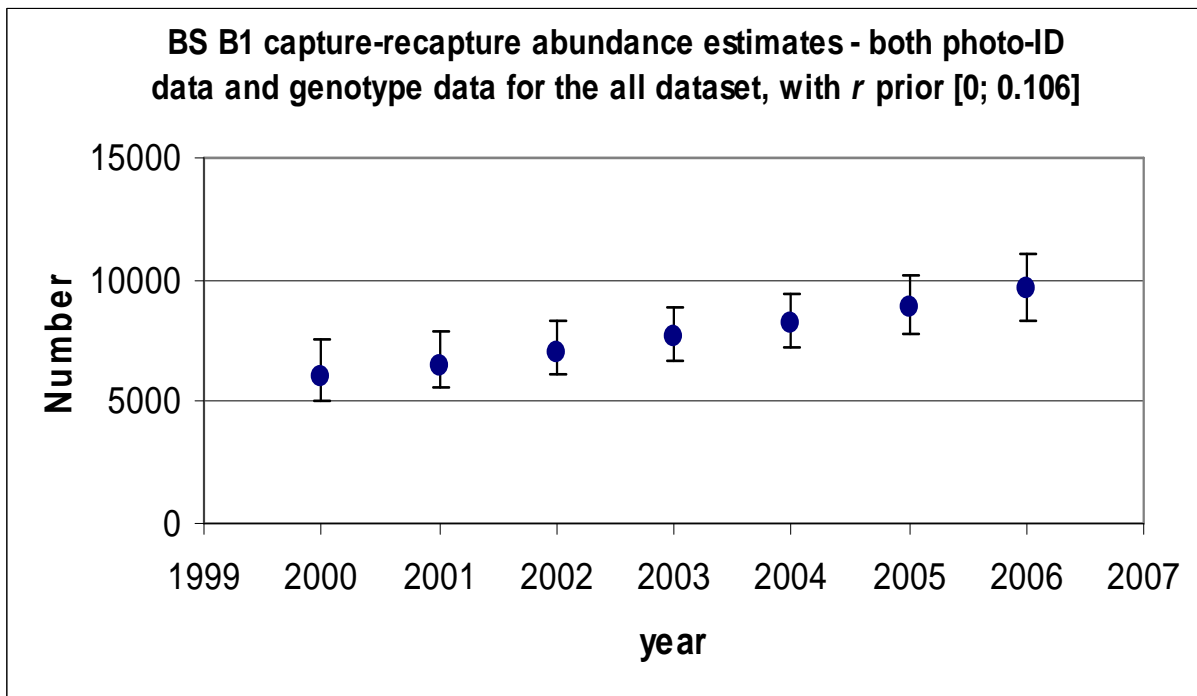

Figure 2b: Bayesian B1 assessment results using both photo-ID + genotypic data for only Iguela data; prior for *r* is U[0; 0.106].
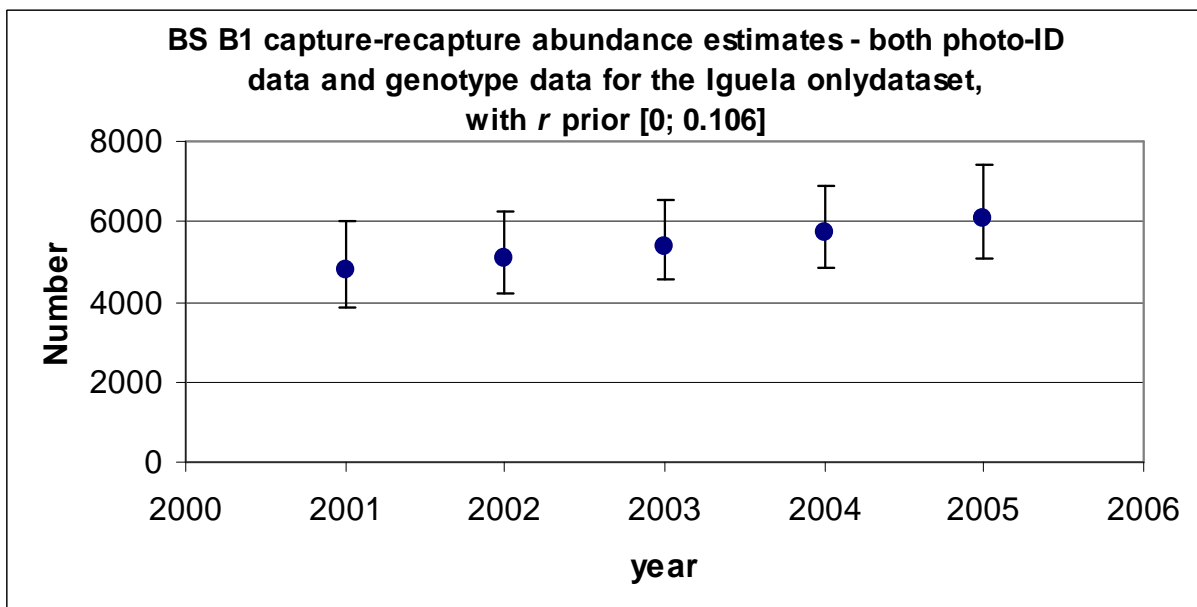
Figure 3a: Histogram of the *r* posterior distribution for the B1 Bayesian analysis that includes both photo-ID and genetic data for the whole dataset and has an *r* prior of *r* ~ U[0; 0.106]. The bars indicate the proportion of the distribution between the value shown and that immediately less; results shown as for 0.11 are between 0.10 and 0.106.

r posterior distribution

Figure 3b: Histogram of the *r* posterior distribution for the B1 Bayesian analysis that includes both photo-ID and genetic data for Iguela only and has an *r* prior of *r* ~ U[0; 0.106]. The bars indicate the proportion of the distribution between the value shown and that immediately less; results shown as for 0.11 are between 0.10 and 0.106.
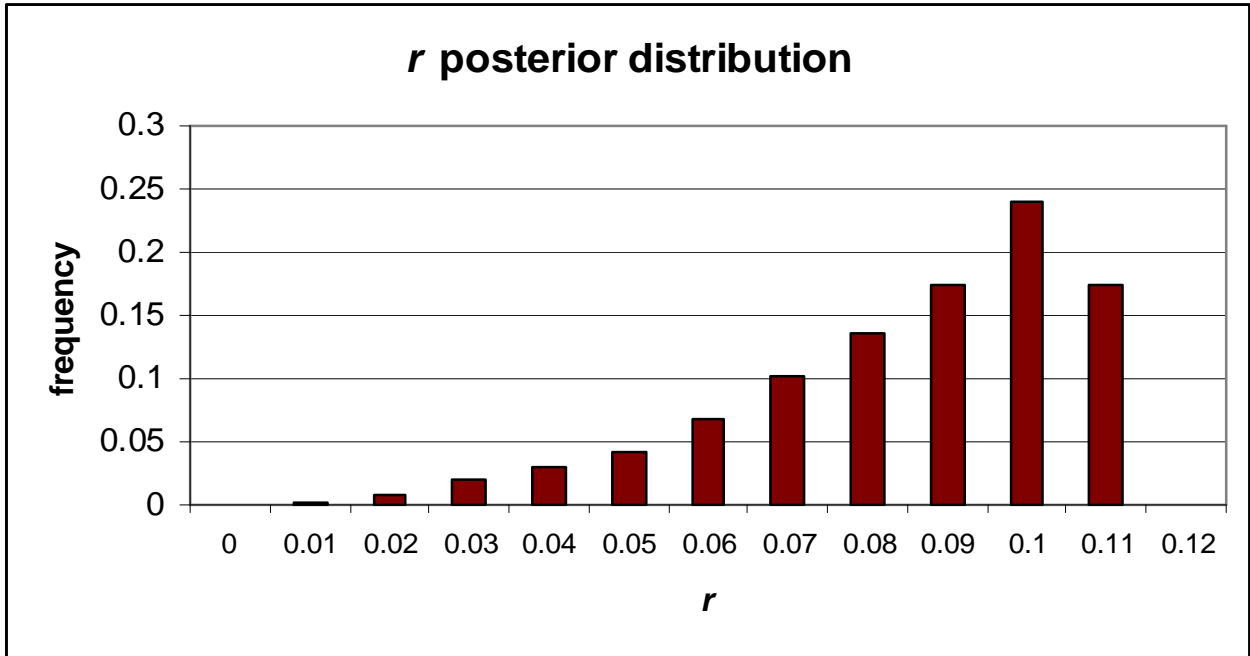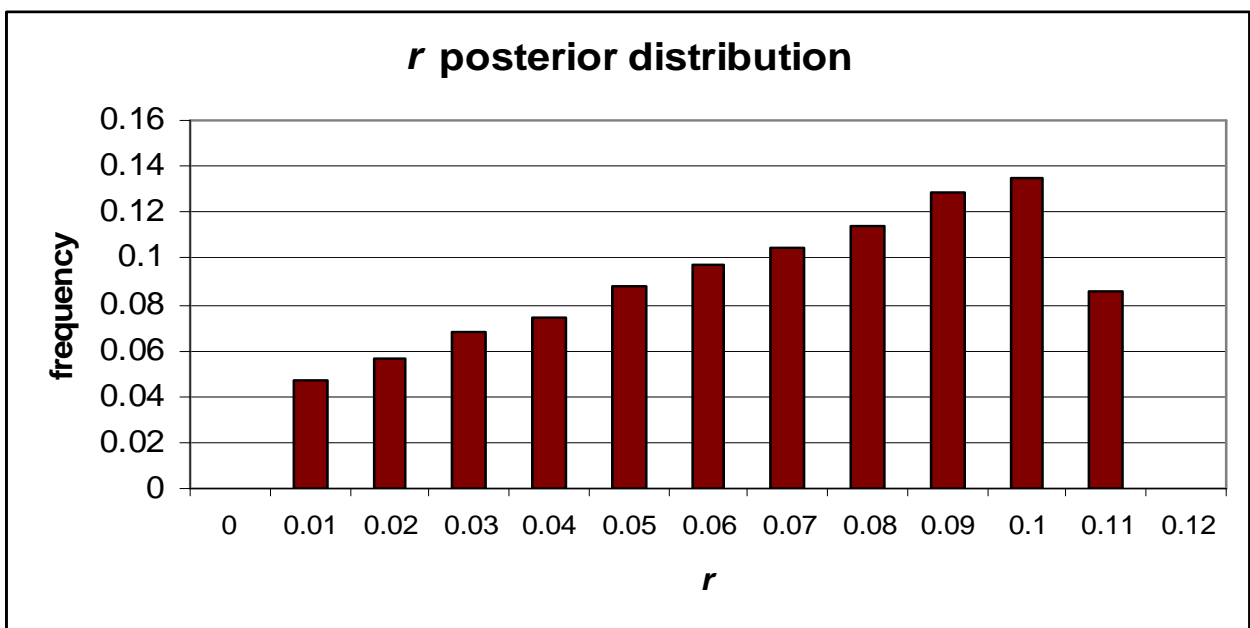

r posterior distribution

Appendix 1: Data from Collins *et al*. (2008) used in these analyses

Photographic capture-recapture data from B1 – from SC/60/SH28 (Collins *et al.* 2008) [n = number of different individuals sighted each year, m = total recaptures between pairs of years]. Note $m^{B1,B1}$ refers to individuals captured in B1 and recaptured in B1.

Table A1.1: Photo-ID Dataset – Total sample from all sites (2000-2006)

n							
	2000	2001	2002	2003	2004	2005	2006
	24	111	233	161	138	216	99

$m^{B1,B1}$							
	2000	2001	2002	2003	2004	2005	2006
2000	X	0	1	0	0	0	0
2001		X	5	6	5	3	2
2002			X	12	2	2	4
2003				X	7	2	1
2004					X	2	2
2005						X	6
2006							X

Table A1.2: Photo-ID Dataset – Iguela only (2001-2005)

n						
		2001	2002	2003	2004	2005
		111	143	161	138	121

$m^{B1,B1}$						
		2001	2002	2003	2004	2005
2001		X	4	6	5	1
2002			X	6	6	1
2003				X	7	1
2004					X	0
2005						X

Table A1.3: Genotypes – Total sample from all sites (2000-2006)

n							
	2000	2001	2002	2003	2004	2005	2006
	82	155	257	270	188	296	207

$m^{B1,B1}$							
	2000	2001	2002	2003	2004	2005	2006
2000	X	1	1	4	2	3	0
2001		X	6	8	6	3	2
2002			X	6	6	6	4
2003				X	8	7	1
2004					X	3	3
2005						X	11
2006							X

Table A1.4: Genotypes – Iguela only (2001-2005)

n							
		2001	2002	2003	2004	2005	
		155	170	270	188	137	

$m^{B1,B1}$							
		2001	2002	2003	2004	2005	
2001		X	6	8	6	0	
2002			X	4	2	6	
2003				X	8	4	
2004					X	6	
2005						X	