

Fitting both moult counts and tagging data to a population model for Robben Island penguins

William Robinson and Doug Butterworth

Marine Resource Assessment and Management Group, Department of Mathematics and Applied Mathematics, University of Cape Town

(Dated: 5 October 2009)

A population model for penguins at Robben Island is fit to both moult count and tag-resighting data. The fit is reasonable and provides an estimate of the over-dispersion of the resighting data. This is important for generating resightings in planned simulations for analysing the power of a future tagging programme to detect a change in mortality rate linked to the presence or absence of pelagic fish in the vicinity of Robben Island.

1. DATA

We have obtained data from Res Altwegg consisting of records of the resighting of 6694 banded adult penguins from Robben Island from 2001 to 2008. This data set was previously analysed with the Program MARK to estimate penguin survival from 2002 to 2006 (Altwegg, 2009). The data provided do not include information on when each penguin was banded, nor do we know how many penguins were banded in total. However, it is understood that the majority of birds were banded in 2000 during rehabilitation or transportation following the *Treasure* oil spill.

A previous paper describes the compilation of Robben Island moult count data series for both adult and juvenile birds (Plagányi and Robinson, 2008). Recent information led to the adjustment of historic Robben Island penguin moult counts (Robinson and Butterworth, 2009). These adjusted figures are used for this work. The data time series to which our model is fitted runs from 1989 to 2008.

2. PENGUIN POPULATION MODEL

The abundance of penguins resident at Robben Island is modelled using an age-structured production model. The number of female penguins $N_{y,a}$ in year y of age a is determined by the

following equations:

$$N_{y+1,a} = \begin{cases} H e^{\eta_y} \sum_{k=4}^A N_{y,k} & \text{if } a = 0 \\ N_{y,a-1} e^{-M_y} & \text{if } 1 \leq a < A \\ (N_{y,a-1} + N_{y,a}) e^{-M_y} & \text{if } a = A \end{cases} \quad (1)$$

where

H is the estimated average breeding success rate,

η_y is the residual breeding success rate in year y ,

A is the plus-group age, and

M_y is the estimated mortality rate in year y .

The initial number of birds of each age is

$$N_{y_0,a} = N_0 e^{-\lambda a} \quad (2)$$

where both N_0 and λ are estimated parameters.

The population model is fitted to annual moult count data using the negative log-likelihood function

$$-\mathcal{L}_M = \sum_y \left[\ln \sigma_M + \frac{1}{2\sigma_M^2} \left(\ln \bar{N}_y - \ln \hat{N}_y \right)^2 \right] \quad (3)$$

where

$\bar{N}_y = \sum_{a=2}^A N_{y,a}$ is the number of adult females (aged two and over) in year y ,

\hat{N}_y is the number of adult moulters observed in year y , and

σ_M is the standard deviation of the logarithms of the moult counts about the true population numbers.

We also fit the proportion of juvenile (age 1) moulting birds in the population to the observed counts:

$$-\mathcal{L}_J = \sum_y \left[\ln \sigma_J + \frac{1}{2\sigma_J^2} \left(\ln J_y - \ln \hat{J}_y \right)^2 \right] \quad (4)$$

where

TABLE I Constants specified in the penguin age-structured population model

| Constant | Symbol | Value |
|---------------------------------------------------|---------------|-------|
| Plus-group age | A | 10 |
| Standard deviation of logged adult moult counts | σ_M | 0.2 |
| Standard deviation of logged juvenile proportions | σ_J | 0.2 |
| Standard deviation of breeding success residuals | σ_η | 1.0 |

$J_y = \frac{N_{y,1}}{N_{y,1} + \bar{N}_y}$ is the proportion of juvenile birds in the model,

\hat{J}_y is the observed proportion of juvenile moulters in year y , and

σ_J is the standard deviation of the logarithms of the observed juvenile proportions about their true values.

A penalty term for the estimated breeding success residuals is added to the likelihood:

$$\mathcal{P}_\eta = \frac{1}{2\sigma_\eta} \sum_y \eta_y^2 \quad (5)$$

where σ_η is the standard deviation of the η_y residuals.

The values of the constants are given in Table I.

3. TAG-RESIGHTING MODEL

For this work we assume that all the resighted penguins were tagged in year $t = 2000$. The total number of penguins tagged is estimated in the model and denoted T .

For each year that we have resighting data, we estimate the probability p_y of resighting a tagged bird in year y . We denote the expected number of birds last seen in year y and resighted in year y' as $\hat{m}_{y,y'}$. The expected number of birds resighted for the first time since being tagged in each year y' is

$$\hat{m}_{t,y'} = \left\{ \left[(T e^{-M_t} - m_{t,t+1}) e^{-M_{t+1}} - m_{t,t+2} \right] e^{-M_{t+2}} - \dots - m_{t,y'-1} \right\} e^{-M_{y'-1}} p_{y'} \quad (6)$$

The expected number of birds seen in year y and then resighted in year y' is

$$\hat{m}_{y,y'} = \left\{ \left[\left(\sum_{k=t}^{y-1} m_{k,y} e^{-M_y} - m_{y,y+1} \right) e^{-M_{y+1}} - m_{y,y+2} \right] e^{-M_{y+2}} - \dots - m_{y,y'-1} \right\} e^{-M_{y'-1}} p_{y'} \quad (7)$$

Assuming an over-dispersed Poisson error distribution, the negative log-likelihood is:

$$\mathcal{L}_{\text{T-R}} = \frac{1}{D^2} \sum_{i=t}^{y_f-1} \sum_{j=i+1}^{y_f} \left(-m_{i,j} \ln \hat{m}_{i,j} + \hat{m}_{i,j} \right) \quad (8)$$

The likelihood is down-weighted by the over-dispersion coefficient

$$D^2 = \frac{1}{n} \sum_{i=t}^{y_f-1} \sum_{j=i+1}^{y_f} \frac{(m_{i,j} - \hat{m}_{i,j})^2}{\hat{m}_{i,j}} \quad (9)$$

where the number of entries in the resightings half-matrix is

$$n = \frac{1}{2} (y_f - t) (y_f - t + 1) \quad (10)$$

4. RESULTS

The likelihood function to be minimized is the sum of of the negative log-likelihoods and penalty terms for both the moult counts and the tag-resighting data. We denote this as Model A.

$$\mathcal{L} = \mathcal{L}_M + \mathcal{L}_J + \mathcal{P}_\eta + \mathcal{L}_{\text{T-R}} \quad (11)$$

As a sensitivity, we consider an alternative model in which terms of $\mathcal{L}_{\text{T-R}}$ in Equation (8) involving the $m_{2001,y'}$ data are excluded. This is denoted as Model B. The reason these particular data (the 2001 sightings) are excluded is that for Model A their residuals are not random, and are influential in forcing a worse fit to the moult count data.

Finally, we consider a variant of Model A in which the tag-resighting data are disregarded. Here, the $\mathcal{L}_{\text{T-R}}$ term is omitted from the likelihood completely (Model C).

For 20 years of penguin abundance data and 8 years of penguin resighting data, 50 parameters are estimated for Model A, as shown in Table II. (A lower bound is placed on the mortality rate so that $M_y \geq 0.02$.) The estimated values of selected parameters are given for both models.

In Fig. 1 the adult moult count time series is plotted along with the fits to Model A and Model B. As the likelihood vales \mathcal{L}_M in Table II suggest, Model B fits the moult count data better. (When the tag-resighting term is excluded from the likelihood, the moult count data are fitted exactly.)

Fig. 2 shows that Model B predicts a lower average breeding success rate, especially during the late 1990s.

Fig. 3 shows identical results for annual sighting probabilities for Model A and Model B, apart from the final year (2008).

Fig. 4 shows that Model B is better able to fit to the proportion of juvenile birds in the moult count data.

The resighting data are plotted in Fig. 5, along with the fits of Model A and Model B. The fits are good, and are indistinguishable for the two models except for $y = 2001$ which, of course, is the year for which the data were excluded from the likelihood in Model B.

Fig. 6 shows that the two models give identical estimates and Hessian-based confidence intervals (assuming log-normally distributed estimates) for annual mortality rates (converted here to survival rates $S_y = e^{-M_y}$), apart from the year 1996. The corresponding survival estimates when the tag-resighting data are omitted from the likelihood (Model C) are plotted in Fig. 7. The drop in estimate precision over the period 1998 to 2006 is evident. This corresponds to the period over which the tagging data are informative regarding mortality rates.

5. FUTURE WORK

The ultimate aim of this work is to contribute to an evaluation of the power of an experimental programme of closing foraging areas around islands where penguins breed to the pelagic fishery to determine whether such fishing has a meaningful impact on adult survival rates. A future tagging programme will provide the basis to estimate annual mortality rates and correlate these with whether or not pelagic fishing was excluded from the neighbourhood of the island that year.

The power of a tagging programme to detect such a differential effect, and the dependence of this power on the number of new birds to be tagged annually, is to be determined by simulation. The particular contribution of the analyses of this paper is in their provision of information on the statistical properties of the tag-resighting data, and in particular the value of the over-dispersion parameter D , which are to be used in generating future tag-resightings in the simulations planned.

Before that work is undertaken, efforts will be made to re-estimate D excluding the bias that arises from estimating a large number of parameters given few data. Changing the error distribution for the tag-resightings from over-dispersed Poisson to over-dispersed multinomial will also be considered as a sensitivity test. Future work will also attempt more accurate estimates of the precision of parameter values estimated than is provided by the Hessian.

TABLE II Model parameter estimates and likelihoods

| Parameter | Symbol | Model A | Model B | Model C |
|---------------------------------------|---------------------|---------|---------|---------|
| Annual mortality | M_{1989} | 0.165 | 0.167 | 0.167 |
| | M_{1990} | 0.020 | 0.020 | 0.020 |
| | M_{1991} | 0.067 | 0.064 | 0.063 |
| | M_{1992} | 0.020 | 0.020 | 0.020 |
| | M_{1993} | 0.020 | 0.020 | 0.020 |
| | M_{1994} | 0.161 | 0.160 | 0.159 |
| | M_{1995} | 0.346 | 0.347 | 0.347 |
| | M_{1996} | 0.145 | 0.227 | 0.240 |
| | M_{1997} | 0.020 | 0.020 | 0.020 |
| | M_{1998} | 0.020 | 0.020 | 0.057 |
| | M_{1999} | 0.020 | 0.020 | 0.020 |
| | M_{2000} | 0.020 | 0.020 | 0.020 |
| | M_{2001} | 0.403 | 0.020 | 0.020 |
| | M_{2002} | 0.271 | 0.272 | 0.235 |
| | M_{2003} | 0.277 | 0.277 | 0.020 |
| | M_{2004} | 0.477 | 0.477 | 0.464 |
| | M_{2005} | 0.372 | 0.374 | 0.649 |
| M_{2006} | 0.573 | 0.575 | 0.405 | |
| M_{2007} | 0.131 | 0.198 | 0.273 | |
| Initial population | N_0 | 531.6 | 530.8 | 530.5 |
| | λ | 0.191 | 0.190 | 0.190 |
| Breeding success | H | 0.568 | 0.517 | 0.501 |
| Breeding success residuals (19 years) | η_y | | | |
| Resighting probabilities (8 years) | p_y | | | |
| Number of birds tagged in 2000 | T | 22016 | 15162 | – |
| Over-dispersion coefficient | D^2 | 4.9 | 6.0 | – |
| Moult count likelihood | \mathcal{L}_M | -27.4 | -30.6 | -31.9 |
| Juvenile proportion likelihood | \mathcal{L}_J | -29.9 | -31.9 | -32.1 |
| Breeding success penalty | \mathcal{P}_η | 0.6 | 0.5 | 0.3 |
| Tag-resighting likelihood | \mathcal{L}_{T-R} | -10613 | -8612 | – |
| Total likelihood | \mathcal{L} | -10669 | -8674 | -63.7 |

Fig. 1 Model fits to moult count data: the Model A estimates are black and those for Model B are grey, both here and in subsequent plots.

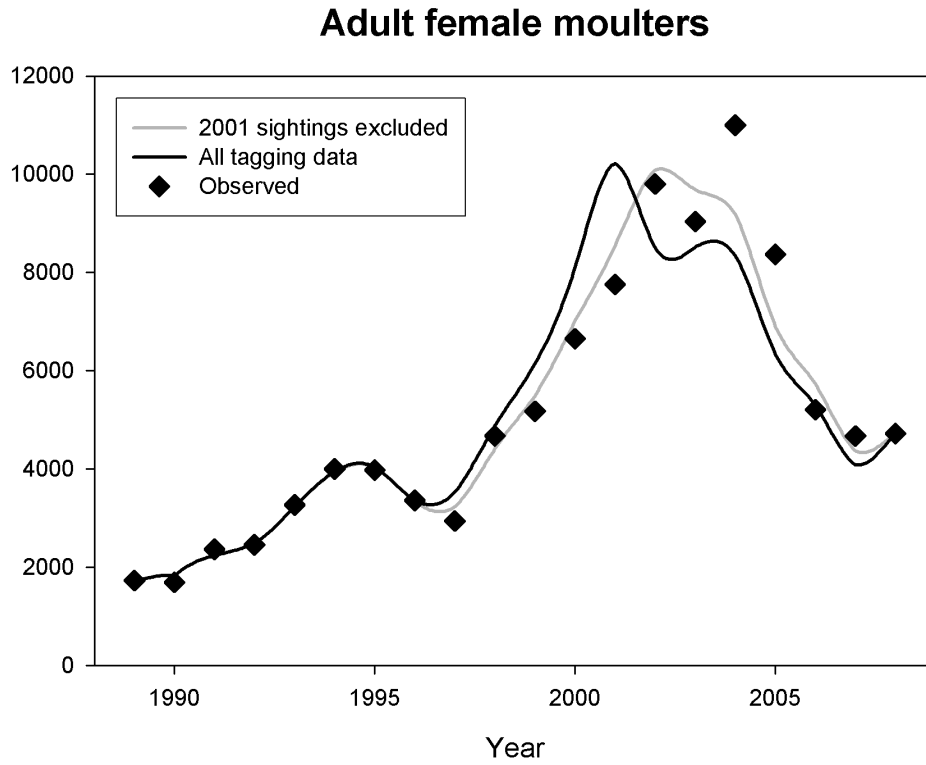


Fig. 2 Annual breeding success.

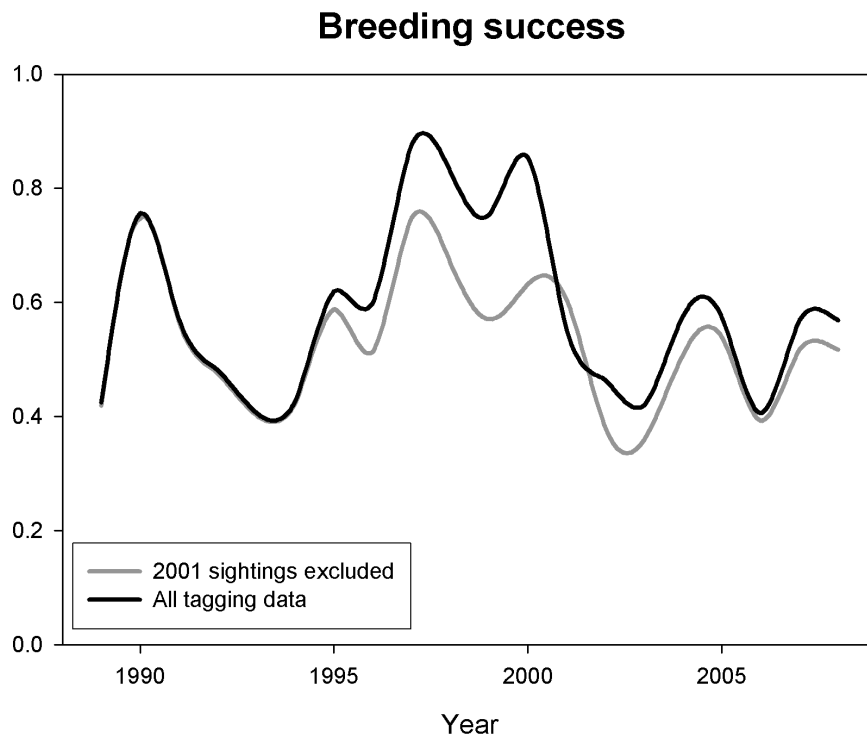


Fig. 3 Annual probability of sighting a tagged bird.

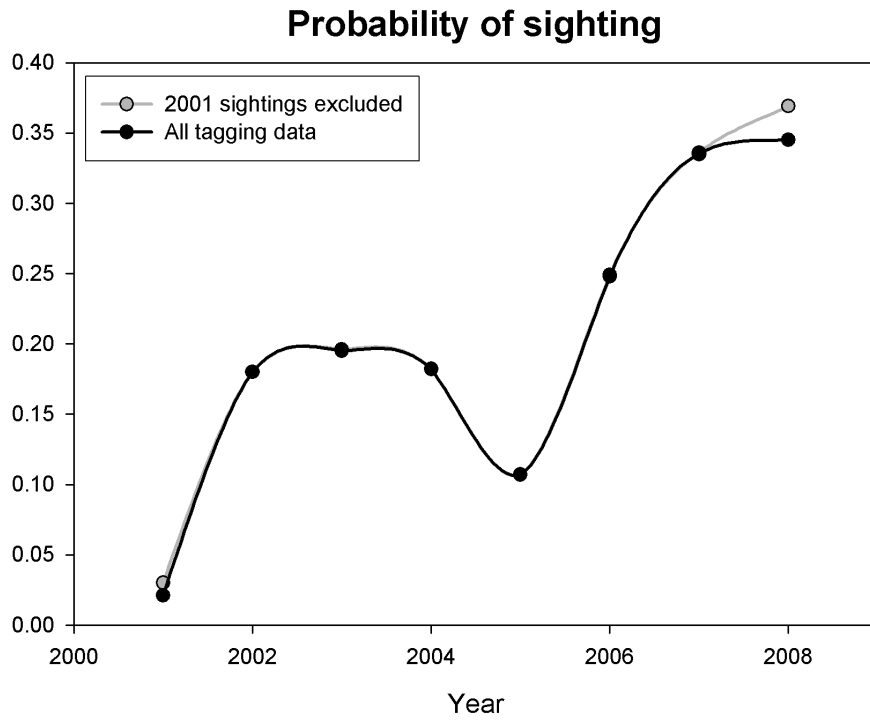


Fig. 4 Proportion of moulters which are juvenile.

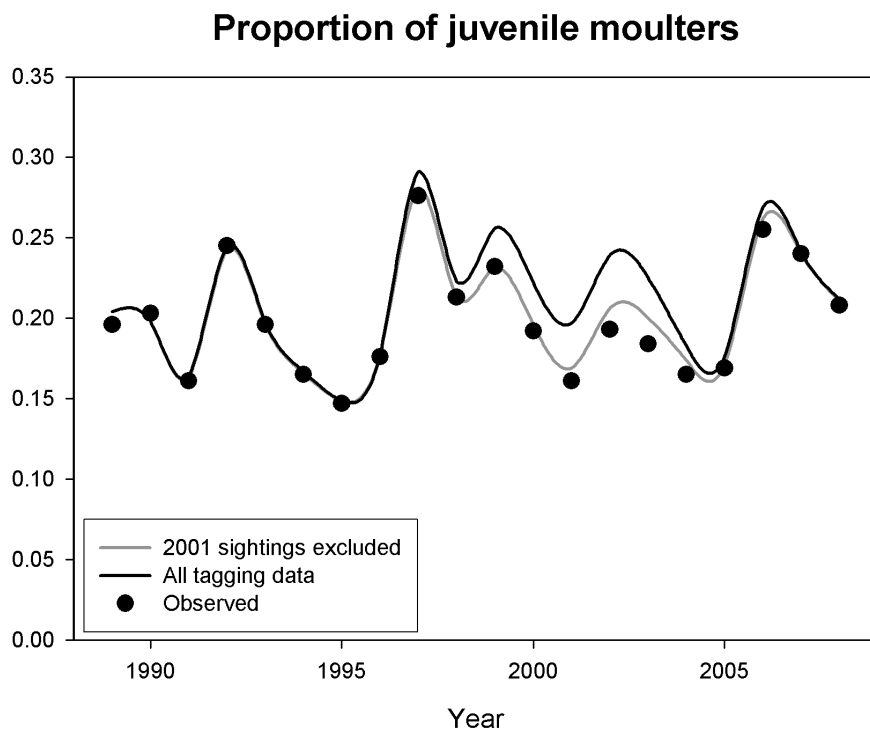


Fig. 5 Number of tagged birds sighted in year y (plot headings) and next resighted in year y' (horizontal axes). Fits are shown for Model A and Model B (often indistinguishable).

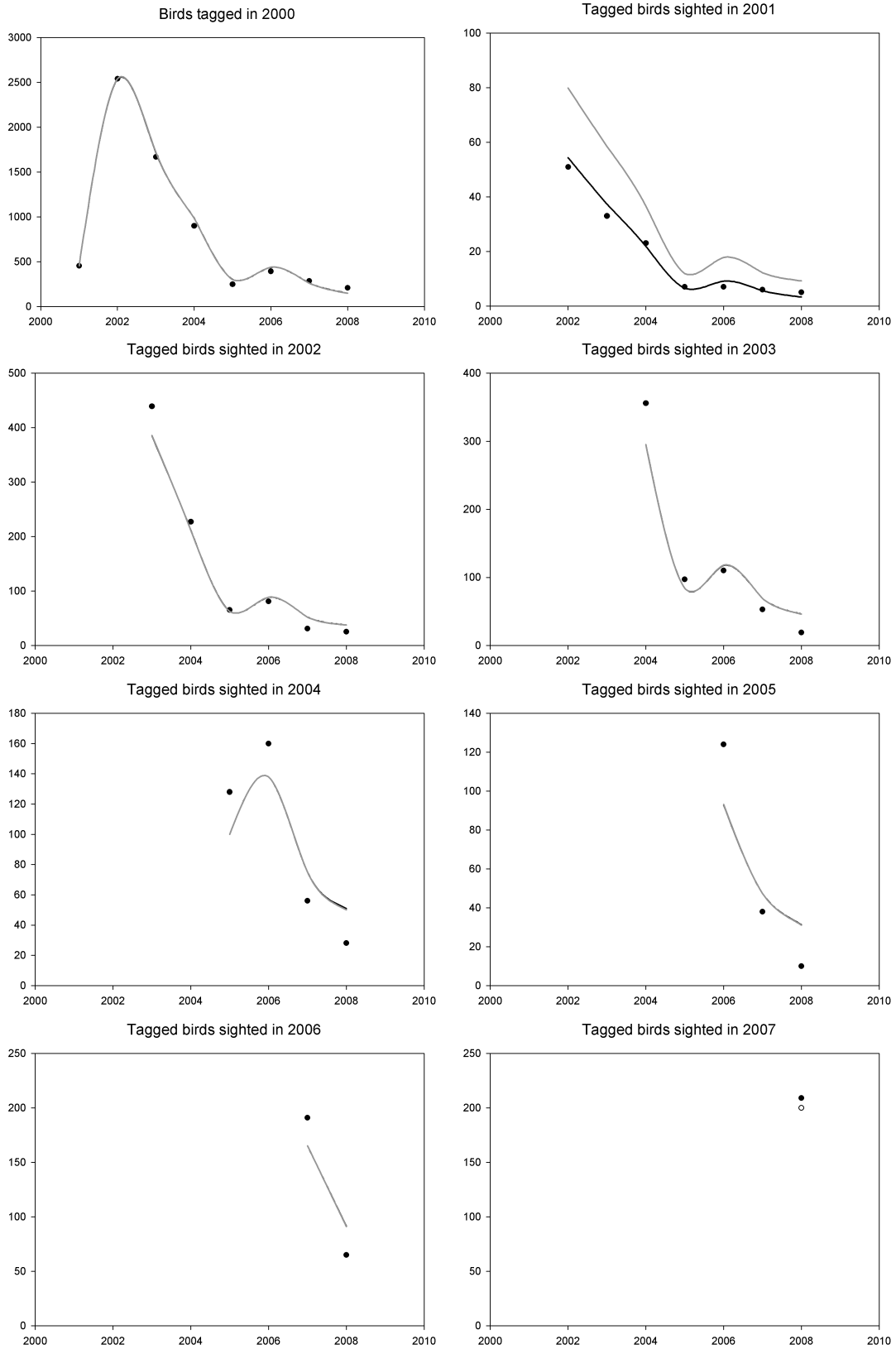


Fig. 6 Estimated annual survival rates with Hessian-based 95% confidence intervals for Model A (top) and Model B (bottom).

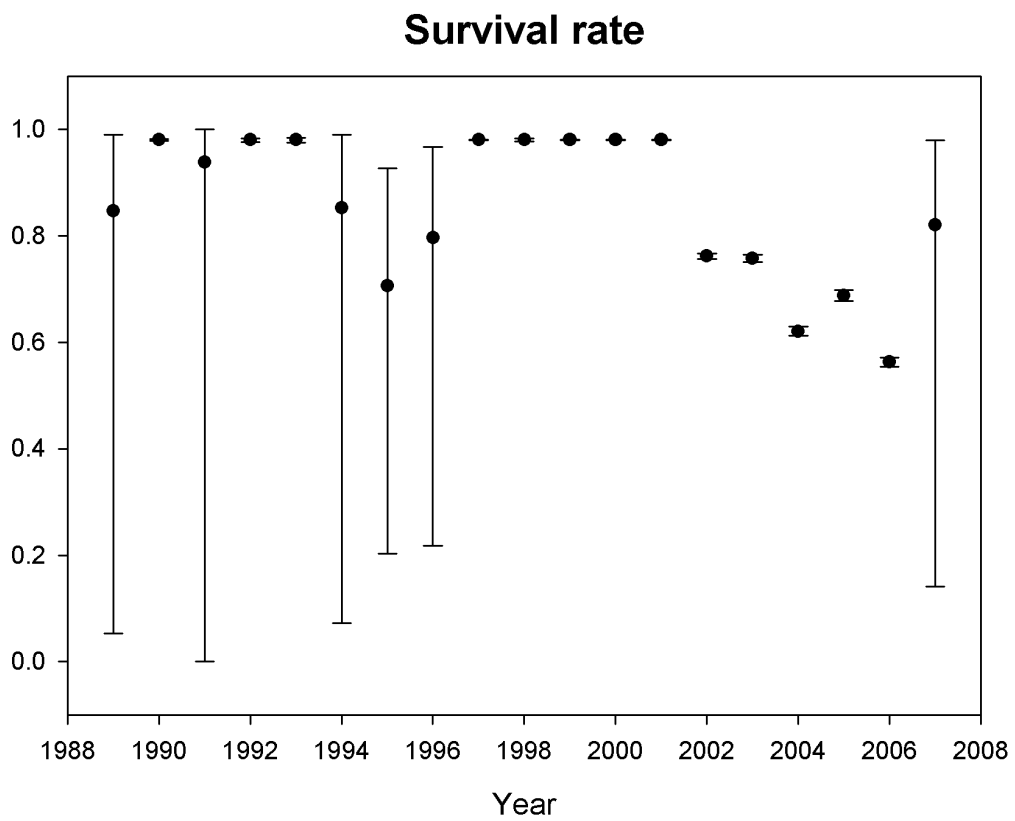
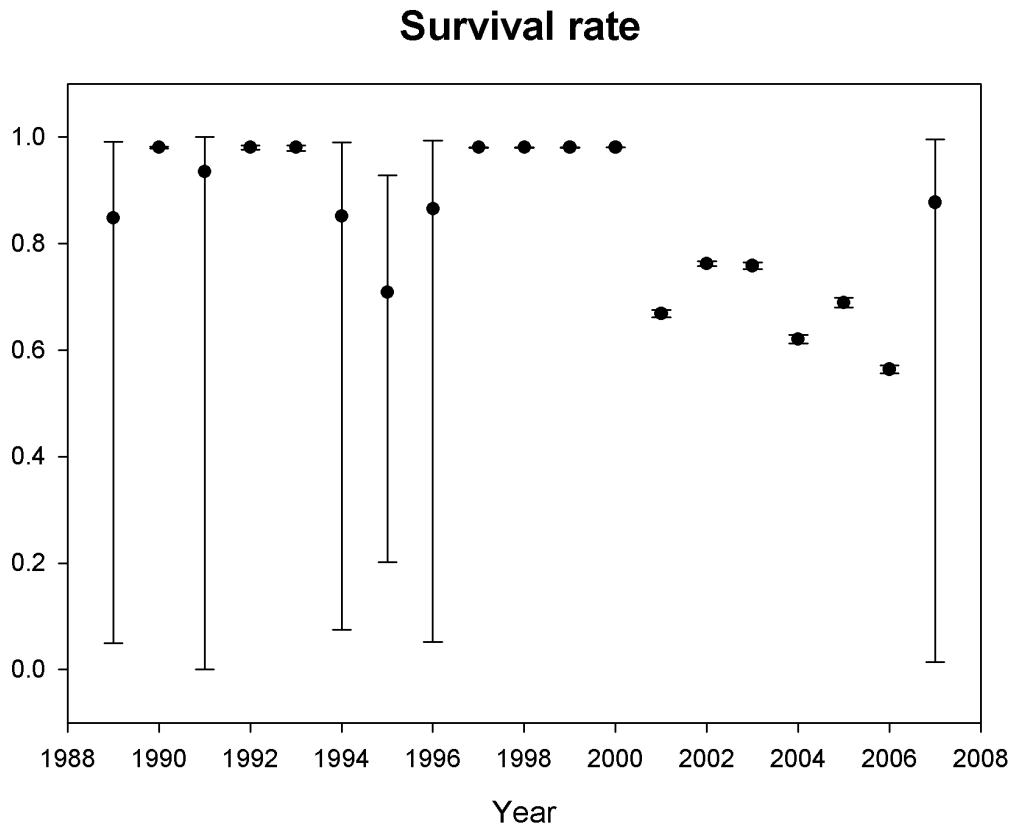
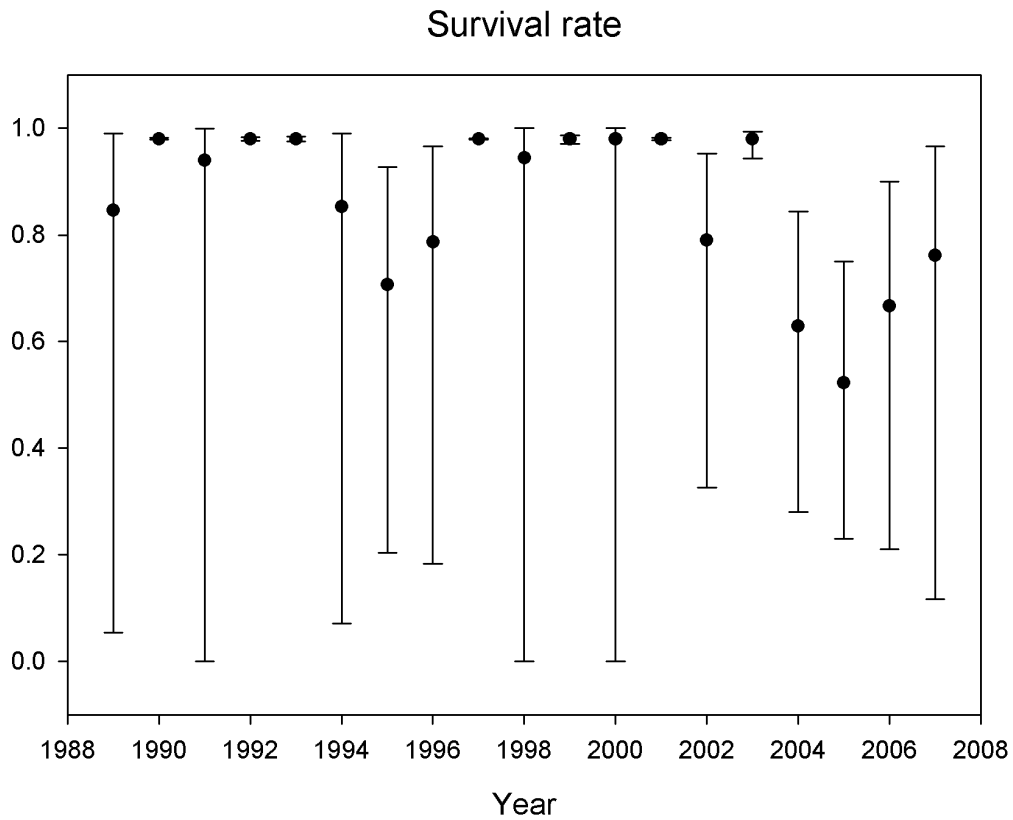


Fig. 7 Estimated annual survival rates with Hessian-based 95% confidence intervals for Model C.



References

- Altwegg, R. 2009. *Survival of African penguins on Robben and Dassen islands from 2002 to 2006*, Working Group document.
- Plagányi, É. and W. Robinson. 2008. *Data inputs for the African penguin *Spheniscus demersus* model to be coupled to the pelagic OMP*. Document MCM/2008/DEC/SWG-PEL/26.
- Robinson, W. and D. S. Butterworth. 2009. *Update to the African penguin population model taking account of new moult count information*. Document MCM/2009/MAY/SWG-PEL/14.