

Penguin population models for Robben Island

William Robinson and Doug Butterworth

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

November 2010

Abstract

As an initial illustrative exercise, a population model for Robben Island penguins, which includes dependency of reproductive success and survival rate on pelagic fish abundance, is fitted to moult count information for the colony. The results indicate a strong dependence of adult penguin survival rate on sardine abundance west of Cape Agulhas. However, the logistic transformation used to ensure respect of biological constraints on these demographic parameters leads to some problems in extending the approach to Bayesian estimation. Further work will explore use of the beta distribution and the incorporation of further data for the penguin population in the fitting process.

Introduction

Penguin–fisheries population modelling carried out in 2008 is summarised in MARAM IWS/DEC/PA/P1. MARAM IWS/DEC10/PA/P2 contains some recent comments on those analyses and responses thereto. This paper follows on from MARAM IWS/DEC10/PA/P2 in implementing some of the structural model changes put forward in those responses.

At this stage, the model has deliberately been kept simple to enable a focus first on key estimation properties. Thus the model is restricted to a closed population analysis of the Robben Island colony fitted to penguin moult count data. Other penguin data can be added in due course.

The paper sets out the basic methodology, and gives results for an initial reference case model fit. Results for variants of this case will follow.

Basic dynamics

The model considers the number of female penguins $N_{y,a}$ at the start (1 January) of year y of age a at Robben Island (see

Figure 1). The initial population size (at the start of year 1988) and structure is:

$$N_{y_0,a} = \begin{cases} N_0 e^{-a\lambda} & \text{for } 1 \leq a < A \\ \frac{N_0 e^{-a\lambda}}{1 - e^{-\lambda}} & \text{for } a = A \end{cases} \quad (1)$$

where A is the plus-group age. Both N_0 and λ are parameters whose values are estimated.

The following equations describe the population trajectory:

$$N_{y+1,a} = \begin{cases} \frac{1}{2} H_y \left(\sum_{a=a^*}^A N_{y,a} \right) S_y^{4/12} (1 - p_y^{\text{oil}}) & \text{if } a = 1 \\ N_{y,a-1}^* S_y & \text{if } 2 \leq a < A \\ (N_{y,a-1}^* + N_{y,a}^*) S_y & \text{if } a = A \end{cases} \quad (2)$$

with

$$N_{y,a}^* = N_{y,a} \left(1 - \frac{N_y^{\text{oil}}}{\left(\sum_{a=1}^A N_{y,a} \right) S_y^{m_y/12}} \right) \quad (3)$$

where

S_y is the adult (post 1 January of first year of life) annual survival rate in year y ,

H_y is the annual reproductive success (number of chicks per mature female reaching 1 January of the year following birth, where 50% of these chicks are assumed to be female),

a^* is the age at which the penguins first attempt to breed,

p_y^{oil} is the proportion of chicks thought to have died as a result of an oil spill in year y ,

N_y^{oil} is the number of juvenile and adult penguins thought to have died as a result of oiling in year y (see Table 6 in document MARAM IWS/DEC10/PA/P4), and

m_y is the month in which the oil spill occurred in year y .

Note that this (initial) version of the model is closed, i.e. without immigration or emigration.

Population model

Both the annual adult survival rate S_y and the annual reproductive success H_y are assumed to depend on some function of prey biomass (the deterministic effect), but to be influenced also by some noise (random effects). Two formulations have been implemented with the aims of respecting biologically plausible bounds and ensuring that the optimization routine is stable. In the first case an adjusted logistic transformation maps the estimated values onto the specified ranges, while in the second case the estimates are drawn from a beta distribution which is bounded by definition. The equations concerning the annual adult survival rate follow. The equations concerning reproductive success rate are completely analogous, and hence are not reproduced here.

Adult survival depends on the normalized annual biomass levels $B_{S,y}$, where the time series I_S is some function of the sardine and anchovy November spawner biomass and May recruit biomass survey results:

$$B_{S,y} = I_{S,y} / \max\{I_S\} \quad (4)$$

Method 1: Adjusted logistic

In logit space, expected survival $Z_{S,y}$ depends on the logarithm of the normalized fish abundance:

$$Z_{S,y} = f_{\text{logistic}}(B_{S,y}) = \mu_S \ln B_{S,y} + \eta_S \quad (5)$$

where the functional form assumed here is a reference case for which estimation robustness will be checked for alternative choices. The parameters μ_S and η_S are estimated. A logistic transformation relates $Z_{S,y}$ with added normally distributed random effects to the annual survival rate S_y , which is constrained to fall between S_{\min} and S_{\max} :

$$S_y = S_{\min} + \frac{S_{\max} - S_{\min}}{1 + e^{-Z_{S,y} - \varepsilon_{S,y}}}, \quad \varepsilon_{S,y} \text{ from } N(0, \sigma_{S,y}^2) \quad (6)$$

The variance of the annual survival rate $\text{var}(S_y) = \tilde{\sigma}_S^2$ which is taken to be fixed (i.e. independent of y or S). (Note that if instead the variance of $Z_{S,y}$ is fixed, this has the problem that the associated variance of S_y becomes very small when S_y is close to S_{\max} .) This is then related to $\sigma_{S,y}^2$, which is the variance of $Z_{S,y}$, as follows:

$$\begin{aligned} \tilde{\sigma}_S &\approx \left(\frac{dS_y}{dZ_{S,y}} \right) \sigma_{S,y} \\ &= \left[\frac{S_{\max} - S_{\min}}{(1 + e^{Z_{S,y}})^2} e^{-Z_{S,y}} \right] \sigma_{S,y} \end{aligned} \quad (7)$$

$$\begin{aligned} \sigma_{S,y} &= \frac{(1 + e^{-Z_{S,y}})^2 \tilde{\sigma}_S}{(S_{\max} - S_{\min}) e^{-Z_{S,y}}} \\ &= f(Z_{S,y}) \tilde{\sigma}_S \end{aligned} \quad (8)$$

where

$$f(Z_{S,y}) = \frac{(1 + e^{-Z_{S,y}})^2}{(S_{\max} - S_{\min}) e^{-Z_{S,y}}} \quad (9)$$

By rearranging equation (6) and ignoring the random effects, $f(Z_{S,y})$ may be expressed as a function of S_y :

$$\begin{aligned} f(Z_{S,y}) &= \frac{\left(1 + \frac{S_{\max} - S_y}{S_y - S_{\min}}\right)^2}{(S_{\max} - S_{\min}) \frac{S_{\max} - S_y}{S_y - S_{\min}}} \\ &= g(S_y) \end{aligned} \quad (10)$$

In order to prevent $\sigma_{S,y}$ from becoming too large (as this can destabilize the estimation process), equation (8) is replaced by:

$$\sigma_{S,y} = f^*(Z_{S,y}) \tilde{\sigma}_S \quad (11)$$

where

$$f^*(Z_{S,y}) = \begin{cases} f(Z_{S,y}) & \text{for } f(Z_{S,y}) \leq g(0.94S_{\max}) \\ g(0.95S_{\max}) - \frac{\epsilon^2}{2\epsilon - g(0.95S_{\max}) + f(Z_{S,y})} & \text{for } f(Z_{S,y}) > g(0.94S_{\max}) \end{cases} \quad (12)$$

where $\epsilon = g(0.95S_{\max}) - g(0.94S_{\max})$.

The penalty term added to the negative log likelihood accounting for the adult survival residuals $\mathcal{E}_{S,y}$ is:

$$P_S = \frac{1}{2} \sum_y \left(\frac{\mathcal{E}_{S,y}}{\sigma_{S,y}} \right)^2 + \left(\sum_y \mathcal{E}_{S,y} \right)^2 \quad (13)$$

The second term on the RHS was found necessary to stabilise the estimation – essentially it forces the mean of the residuals towards zero so that this mean cannot be “traded” against other parameters in the likelihood maximisation.

Following an identical process for annual reproductive success H_y , the corresponding penalty term is:

$$P_H = \frac{1}{2} \sum_y \left(\frac{\mathcal{E}_{H,y}}{\sigma_{H,y}} \right)^2 + \left(\sum_y \mathcal{E}_{H,y} \right)^2 \quad (14)$$

Method 2: Beta distribution

The parameter \bar{S}^* is estimated on the interval $[0,1]$. \bar{S}^* is the transformed average annual survival rate \bar{S} which falls in the interval $[S_{\min}, S_{\max}]$:

$$\bar{S} = (S_{\max} - S_{\min}) \bar{S}^* + S_{\min} \quad (15)$$

Similarly, the annual survival rate S_y^* is estimated for each year on the interval $[0,1]$ and then transformed to the range $[S_{\min}, S_{\max}]$ as follows:

$$S_y = (S_{\max} - S_{\min}) S_y^* + S_{\min} \quad (16)$$

where

$$S_y^* = f_{\text{beta}}(B_{S,y}) \quad (17)$$

Setting $\kappa_S = \bar{S}^* (1 - \bar{S}^*) / \sigma_S^2 - 1$, the beta distribution parameters α_S and β_S are:

$$\alpha_S = \bar{S}^* \kappa_S \quad \beta_S = \kappa_S (1 - \bar{S}^*) \quad (18)$$

The penalty term added to the negative log likelihood for each year, which assumes S_y^* to be beta-distributed, is:

$$P_S = -(\alpha_S - 1) \ln S_y^* - (\beta_S - 1) \ln (1 - S_y^*) \quad (19)$$

Similarly, for the random effects for reproductive success the penalty term for each year is:

$$P_H = -(\alpha_H - 1) \ln H_y^* - (\beta_H - 1) \ln (1 - H_y^*) \quad (20)$$

Likelihood function

The population model is fitted to annual moult count data for both adult and juvenile birds by taking account of the negative log-likelihood functions:

$$\begin{aligned} -L_M &= \sum_y \left[\ln \sigma_M + \frac{1}{2\sigma_M^2} (\ln N_y^{\text{obs}} - \ln N_y^{\text{model}})^2 \right] \\ -L_J &= \sum_y \left[\ln \sigma_J + \frac{1}{2\sigma_J^2} (\ln J_y^{\text{obs}} - \ln J_y^{\text{model}})^2 \right] \end{aligned} \quad (21)$$

where

σ_M and σ_J are respectively the standard errors of the logarithms of the adult moult counts and juvenile proportions of these counts about their true values (i.e. these reflect observation errors),

$N_y^{\text{model}} = q_M \sum_{a=2}^A N_{y,a} S_y^{11/12}$ is the number of female birds in adult plumage (aged 2 and over) counted in year y , and q_H is the proportion of these birds susceptible for observation (assumed here to be 0.9),

$J_y = \frac{p_J N_{y,1}}{\sum_{a=1}^A N_{y,a}}$ is the proportion of juvenile birds in the model at the time of the moult count (note that $S_y^{11/12}$ factors in numerator and denominator cancel), with p_J being the detectability of juvenile moulters relative to adults in the counting process,

N_y^{obs} is the number of female adult moulters observed in year y (taken to be one half of the total counted adult moulters), and

J_y^{obs} is the observed proportion of moulters in immature plumage counted in year y .

The overall (penalized) negative log-likelihood is thus:

$$-\ln L = -L_M - L_J + P_S + P_H \quad (22)$$

Computations are readily extended to a Bayesian framework by integrating over priors for the estimable parameters of the model. In such computations the penalty terms effect integration over the random effects.

Results

At this stage, results are presented only for an initial reference case for the adjusted logistic model where $I_{S,y}$ is the sardine spawner biomass west of Cape Aguhlas from the November acoustic survey, and $I_{H,y}$ is the recruit biomass west of Cape Infanta from the May acoustic survey. Further results and plans for future extensions will be presented in a following document.

The model values and constants for this reference case are given in Table 1, with the priors used for Bayesian computations listed in Table 2.

Figure 2 and Figure 3 show the penalized maximum likelihood estimates for the annual adult survival and the annual reproductive success.

Figure 4 and Figure 5 show the estimated annual survival and reproductive success rates plotted against the normalized pelagic indices. The model estimated relationships are shown.

Figure 6 and Figure 7 show the estimated random effects for annual survival and annual reproductive success.

Figure 8 shows the model fit to the observed female moult counts. Figure 9 shows the model fit to the juvenile proportion data. Figure 10 and Figure 11 show the residuals for these fits.

Figure 12 to Figure 15 show the Bayesian posterior medians and 90% probability intervals for the female moult counts, juvenile proportions, annual survival and annual reproductive success. The joint posterior modes are indicated by dashed lines.

Figure 16 shows the posterior distributions for the annual survival rate and reproductive success in the years 1995 and 2005.

Discussion

Figure 4 indicates a strong relationship between annual adult survival and sardine spawner abundance. However, no strong dependence has been found between annual reproductive success and any similar pelagic index including the anchovy recruit survey estimates for which results are reported here (Figure 5). A possible reason for this last failure is that different aspects of reproductive success may depend on different pelagic signals. For example, if sardine spawner abundance affects the proportion of birds which attempt breeding, while anchovy recruit abundance best accounts for variability in chick survival, more complex forms than considered thus far would need to be examined. Alternatively, it may be that pelagic fish abundance over the period considered has always been sufficiently high not to impact reproductive success at Robben Island.

Because of the logistic transformation, the uniform priors on the μ and η parameters translate to priors on S_y and H_y which are heavily U-shaped. This problem affects the posterior distributions in Figure 16, and explains why the joint posterior modes for survival rate in Figure 14 often differ from the posterior medians. This effect also influence the shapes of the marginal posterior distributions for survival rate shown in Figure 16. Continued use of the adjusted logistic approach requires the specification of priors for the μ and η parameters that correspond to less informative priors for H_y and particularly S_y .

Employing the beta distribution (method 2) instead would address the current logistic transformation's problem of informative priors for S_y in particular. However, early attempts to do this experienced difficulties in estimating the α and β parameters which might need to be constrained to secure stable estimation. These attempts were however for models without dependence on fish abundance, and inclusion of this factor may assist estimation stability. This is a high priority for continuing analysis.

Further data which are available and could readily be included in the model are annual nest counts and tag data. As reported in MARAM IWS/DEC10/PA/P3, the tag data have been used to increase the precision of the annual survival estimates during the years for which sighting histories are available.

References

MARAM IWS/DEC10/PA/P1. Robinson W, Plagányi ÉE, Butterworth DS, de Moor CL. 2010. Summary of penguin–pelagic fish interaction modelling during 2008.

MARAM IWS/DEC10/PA/P2. Butterworth DS, Robinson W, Plagányi ÉE, de Moor CL. 2010. Response to MCM/2010/PEL/48: Suggestions for improving the penguin modelling framework developed by William Robinson.

MARAM IWS/DEC10/PA/P3. Robinson W, Butterworth DS. 2010. Penguin survival estimates from tag data using a multinomial likelihood.

MARAM IWS/DEC10/PA/P4. Robinson W, Butterworth DS. 2010. Data inputs for the African penguin population model.

Tables

Table 1: Penguin population model constants and values used for the analyses of this paper.

Constant	Symbol	Value
Plus-group age	A	5
Age of first breeding attempt	a^*	4
Standard error of logged adult moult counts	σ_M	0.2
Standard error of logged juvenile proportions	σ_J	0.1
Standard deviation of survival random effect	$\tilde{\sigma}_S$	0.1
Standard deviation of reproductive success random effect	$\tilde{\sigma}_H$	0.15
Juvenile : adult relative moult undercount	p_J	1.0
Proportion of moulters which are observable	q_M	0.9
Minimum annual survival rate	S_{\min}	0.1
Maximum annual survival rate	S_{\max}	0.96
Minimum annual reproductive success rate	H_{\min}	0.1
Maximum annual reproductive success rate	H_{\max}	1.8

Table 2: Parameters which are estimated for the reference case model.

Parameter	Symbol	Prior
Initial population	$\ln N_0$	$U[1,10]$
	λ	$U[0,3]$
Relationship between survival and fish abundance	μ_S	$U[-10,10]$
	η_S	$U[-10,10]$
Relationship between reproductive success and fish abundance	μ_H	$U[-10,10]$
	η_H	$U[-10,10]$
Random effects in annual survival	$\mathcal{E}_{S,y}$	$N(0, \sigma_{S,y}^2)$
Random effects in annual reproductive success	$\mathcal{E}_{H,y}$	$N(0, \sigma_{H,y}^2)$

Figures

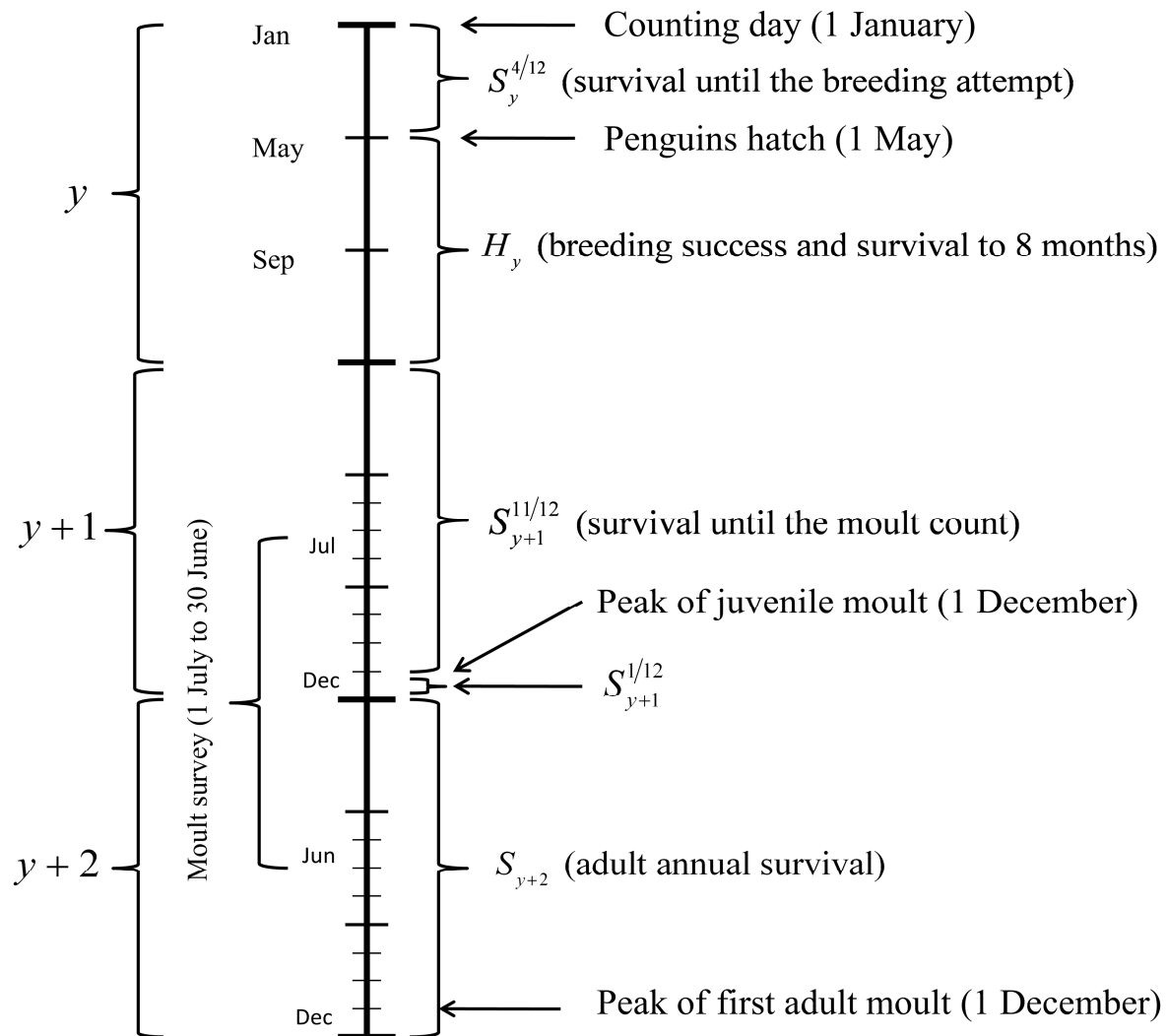


Figure 1: The timeline for an "average" penguin shows the model counting day (1 January), the peak of the breeding season (1 May) and the peak of the moult count (1 December). The observations of each moult season are made over the split year from 1 July until 30 June. Note that in the model the count for year y refers to the moult season where the December peak falls in year y , which is different to the convention used in Table 1 of document MARAM IWS/DEC10/PA/P4. Also shown are the proportion of adults which survive from the model counting day to the hatching day (4 months) and the proportion of adults which survive from the model counting day until the peak of the moult season (11 months).

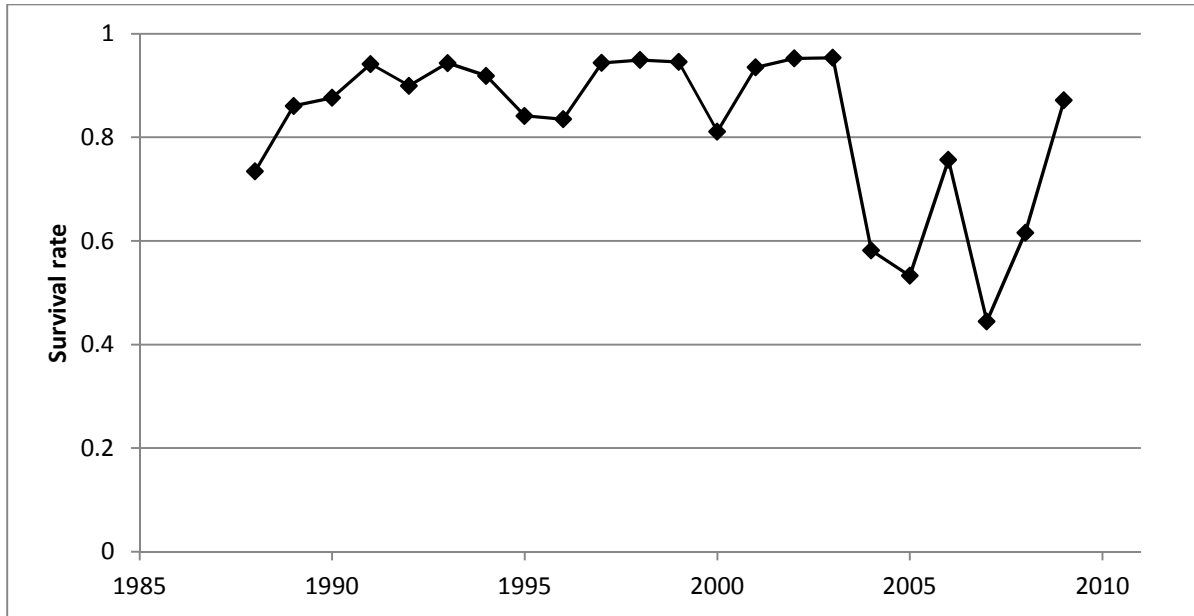


Figure 2: Penalized maximum likelihood estimates of the annual survival rate of penguins.

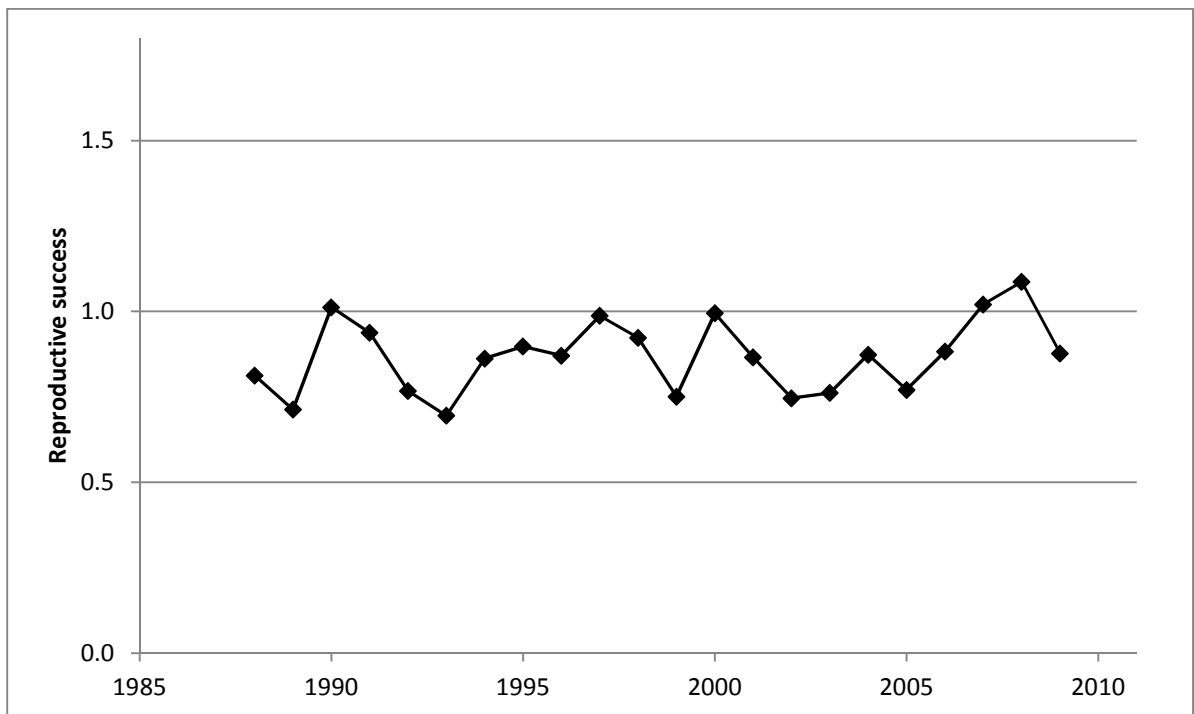


Figure 3: Penalized maximum likelihood estimates of annual reproductive success of penguins, which comprises the proportion of birds which attempt breeding, the average number of eggs laid (including double clutches), and the proportion of chicks which survive until 1 January of the following year.

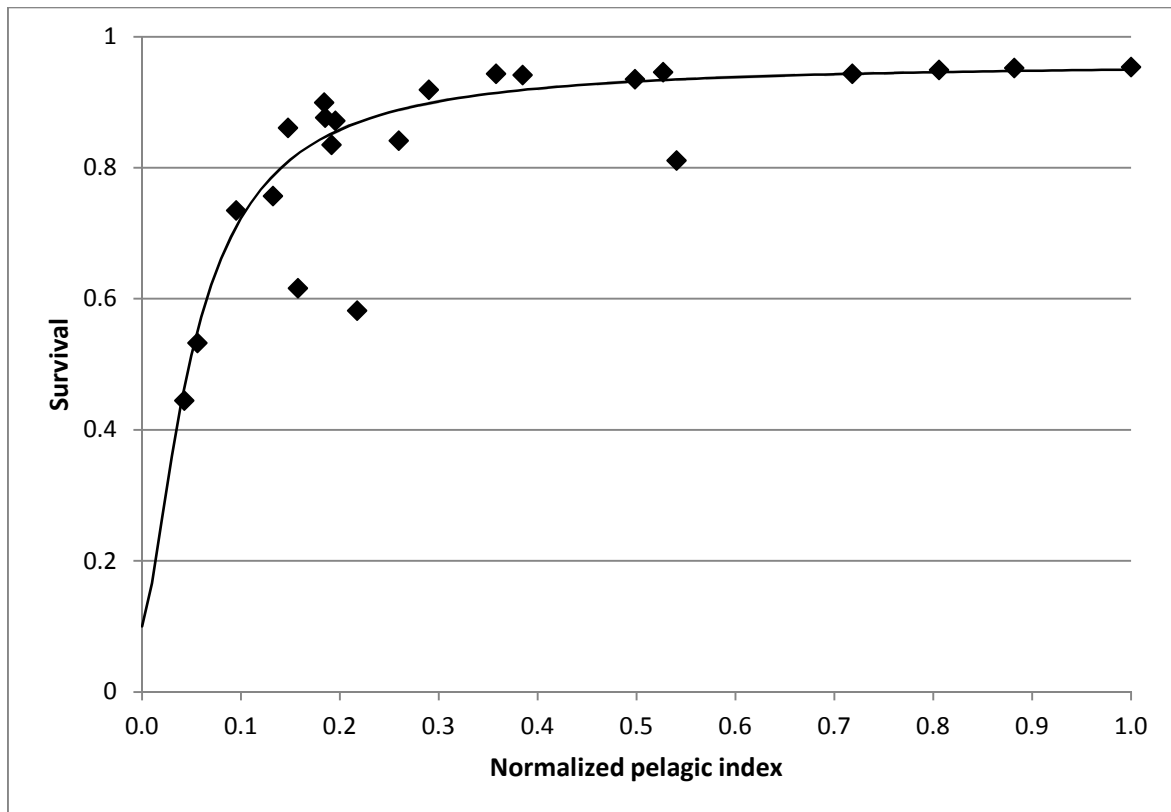


Figure 4: The thin line is the deterministic relationship between the normalized pelagic index (sardine spawner biomass west of Cape Agulhas) and model estimated penguin survival.

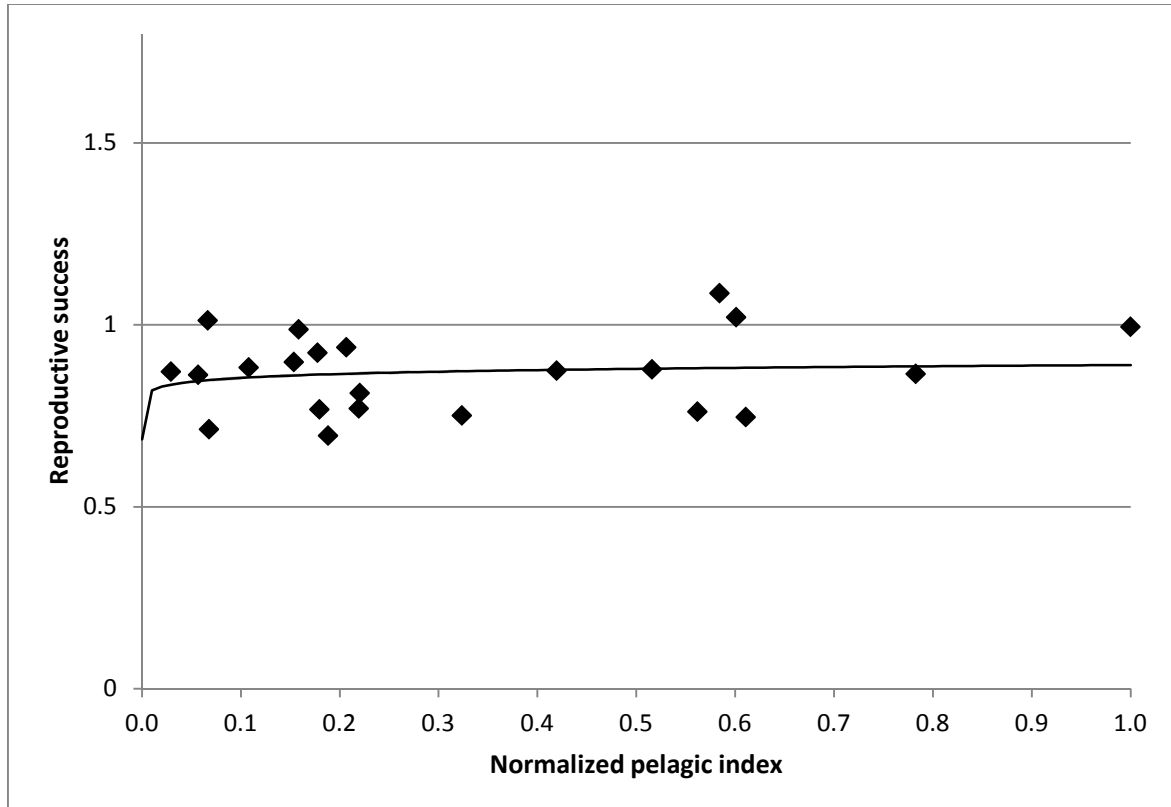


Figure 5: The thin line is the deterministic relationship between the normalized pelagic index (anchovy recruits west of Cape Infanta) and model estimated penguin reproductive success.

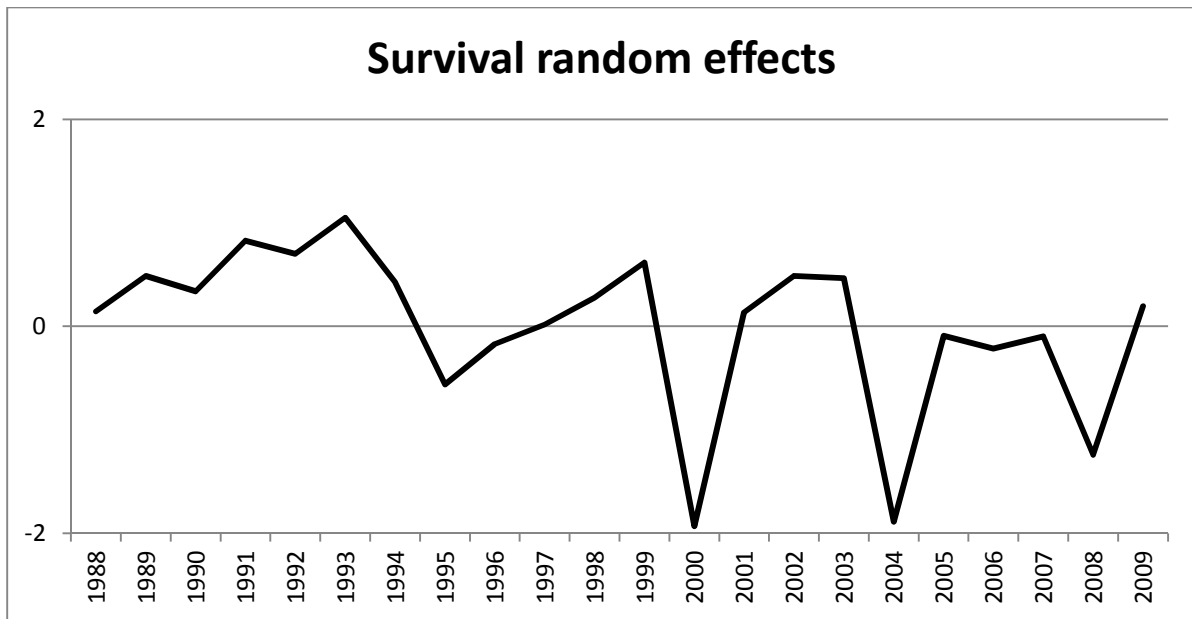


Figure 6: Random effects estimated for the adult annual survival rate.

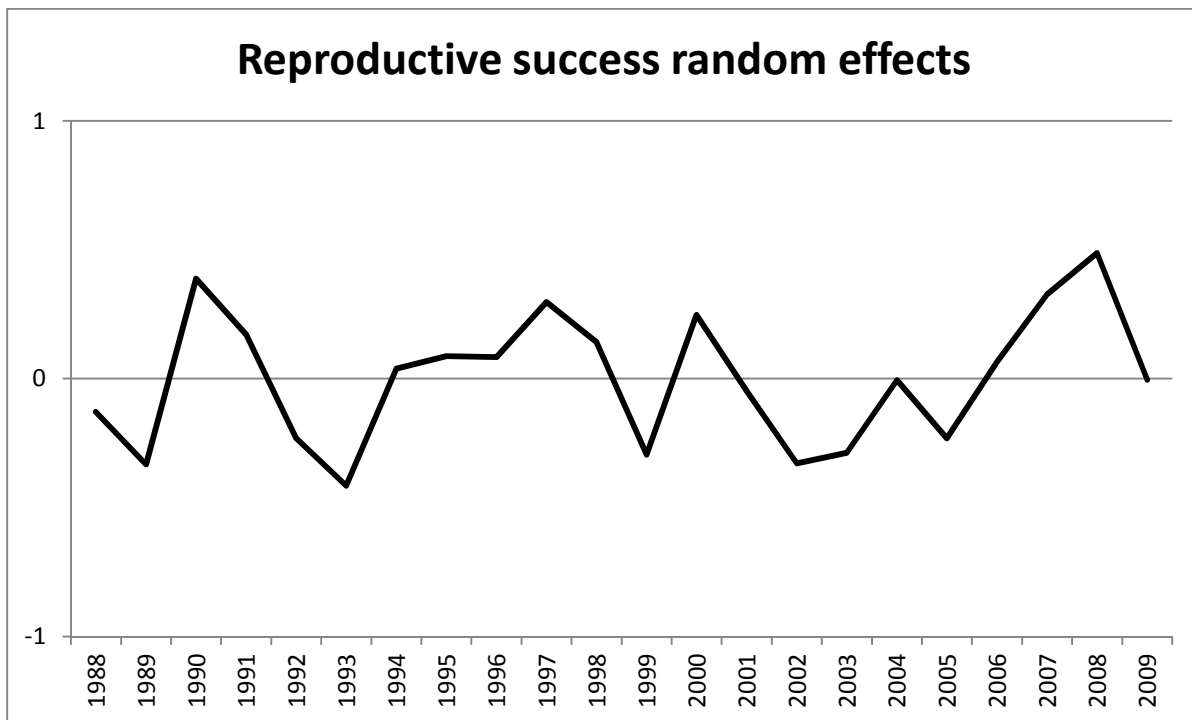


Figure 7: Random effects estimated for the annual reproductive success rate.

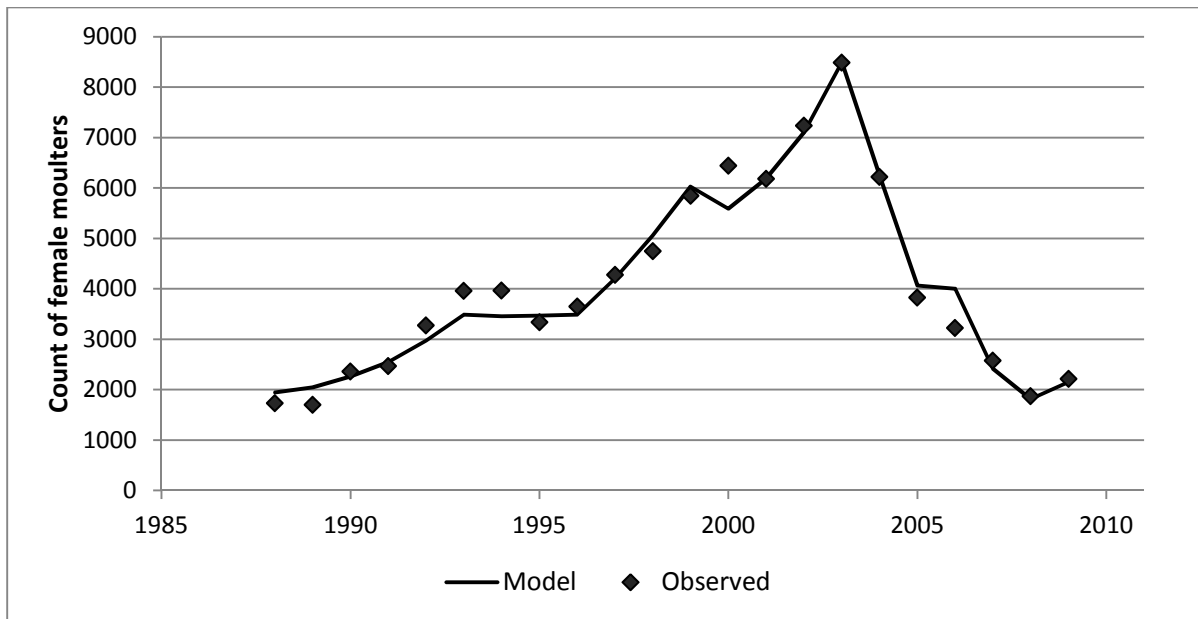


Figure 8: Observed female moult counts (diamonds) and penalized likelihood model estimates (line). Note that the model fits shown here and in the following plots correspond to Bayesian joint posterior modes.

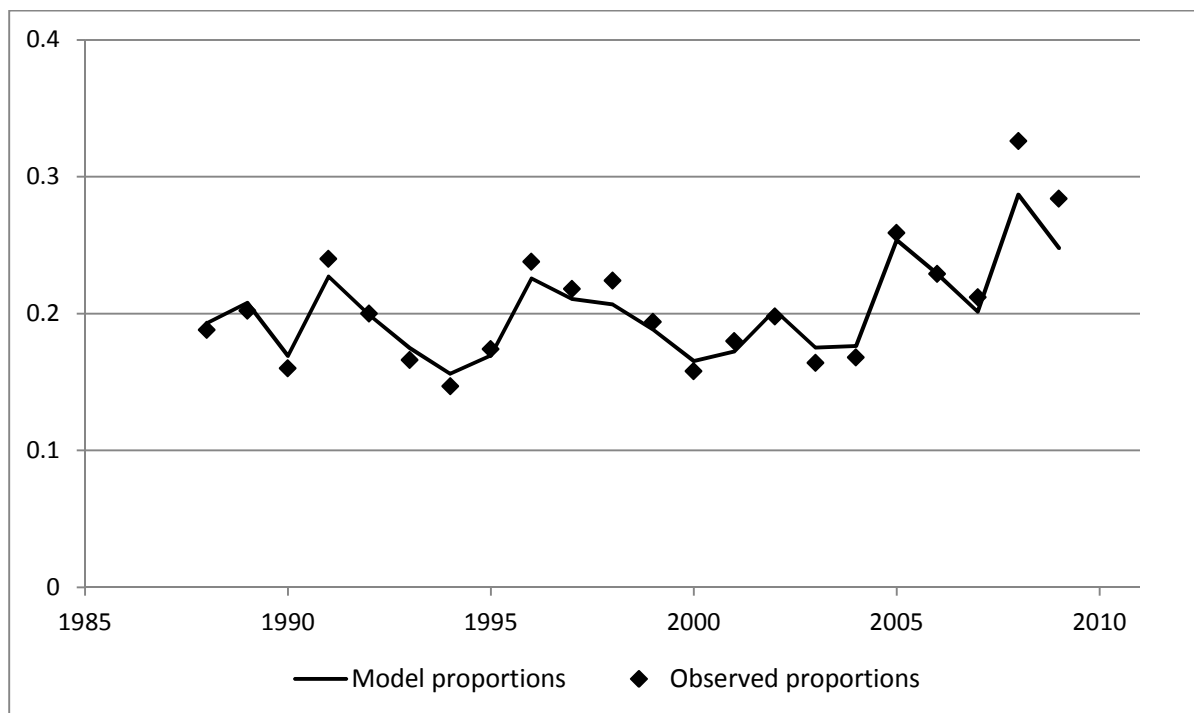


Figure 9: The annual proportion of juveniles as a fraction of the total number of moulters. Observed values are shown with diamonds and the penalized likelihood model estimates are shown with a line.

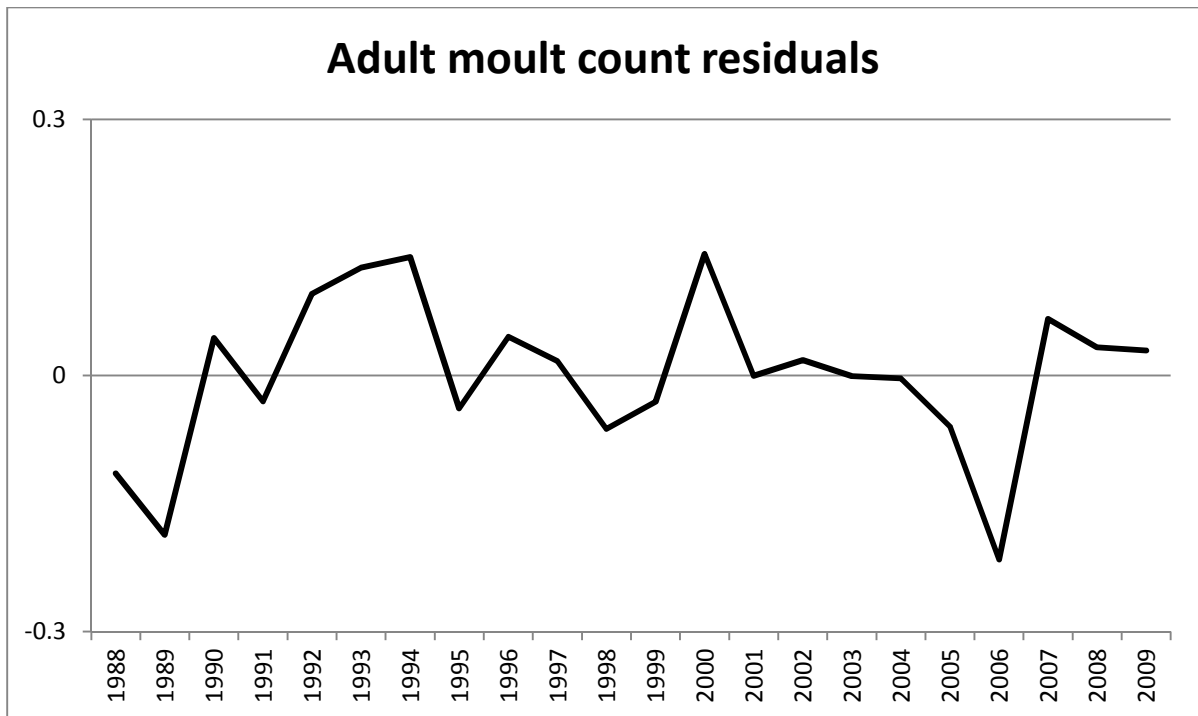


Figure 10: Differences between the logarithms of the observed moult counts and the penalized likelihood model estimated moult counts.

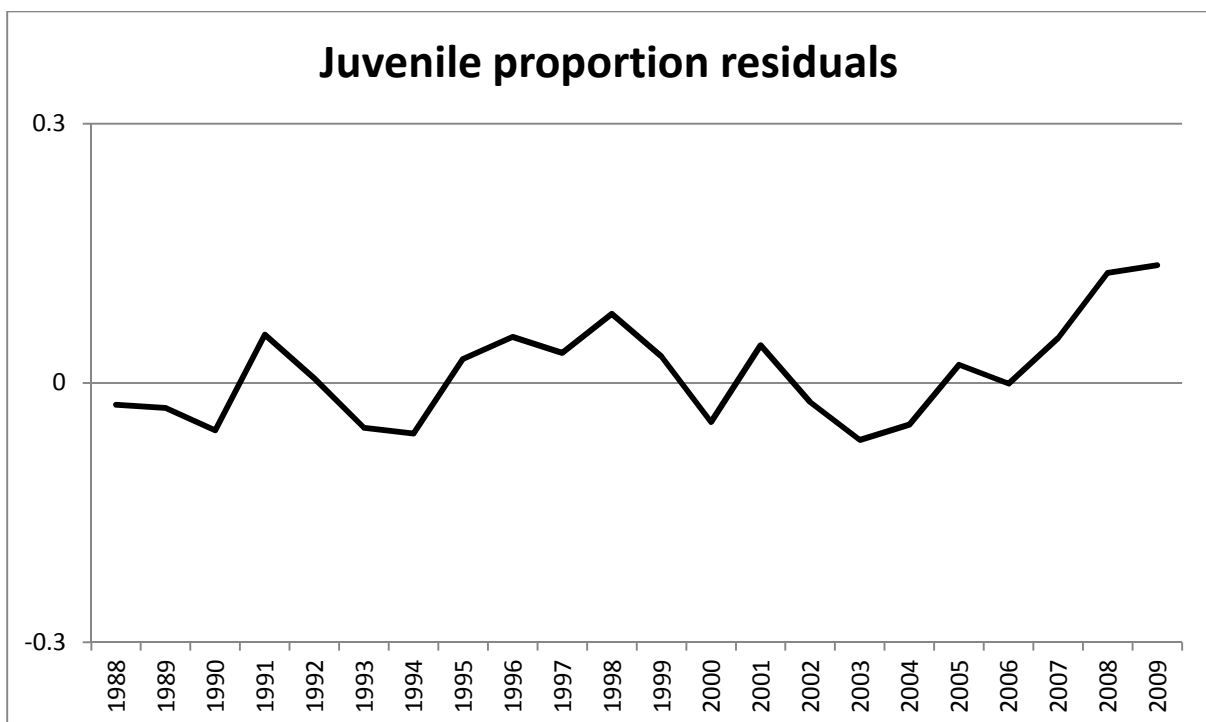


Figure 11: Differences between the logarithms of the observed proportions of immature birds in the moult counts and the penalized likelihood model estimated proportions.

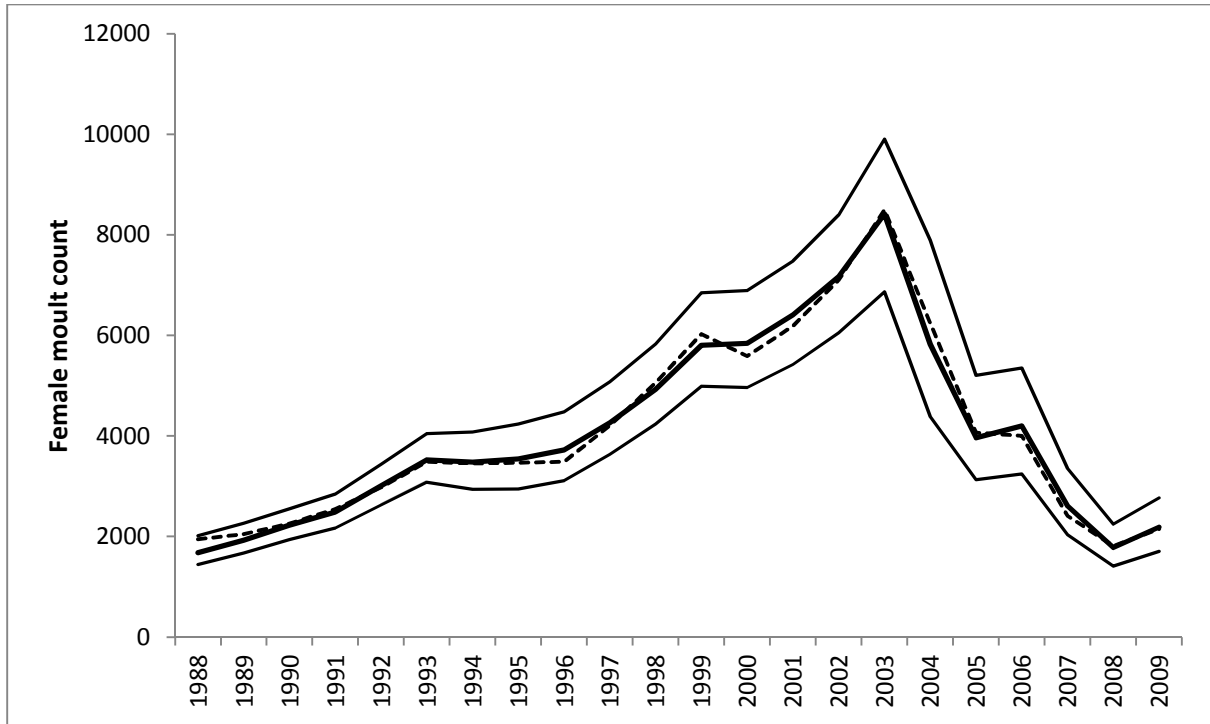


Figure 12: Time series of Bayesian posterior medians and 90% probability intervals for the modelled counts of female moulters. The joint posterior modes are indicated by the dashed line.



Figure 13: Time series of Bayesian posterior medians and 90% probability intervals for the modelled proportion of immature birds in the moult count each year. The joint posterior modes are indicated by the dashed line.

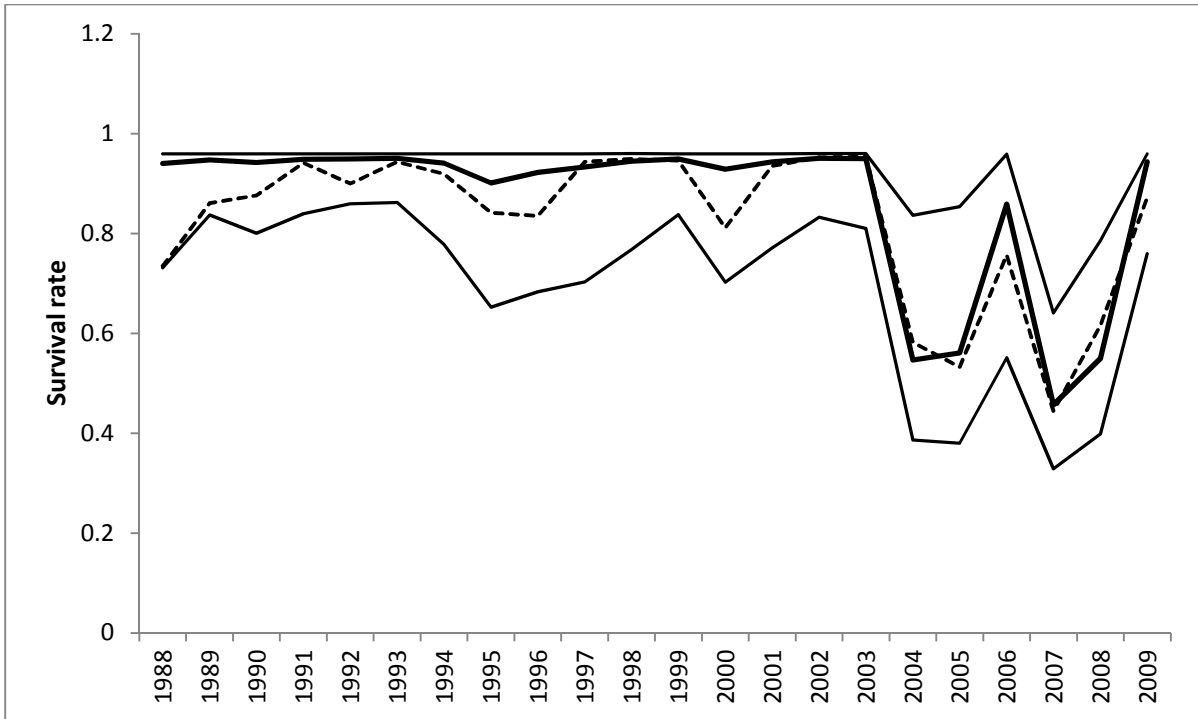


Figure 14: Time series of Bayesian posterior medians and 90% probability intervals for the annual survival rate of penguins. The joint posterior modes are indicated by the dashed line.

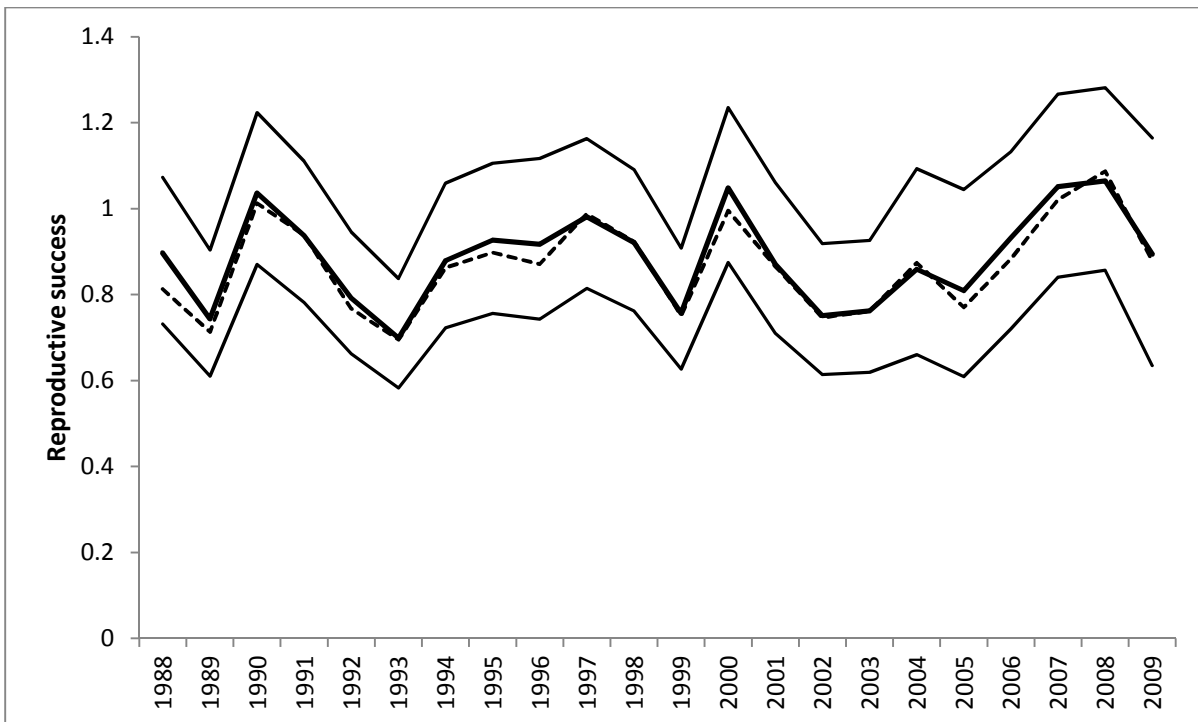


Figure 15: Time series of Bayesian posterior medians and 90% probability intervals for the annual reproductive success of penguins. The joint posterior modes are indicated by the dashed line.

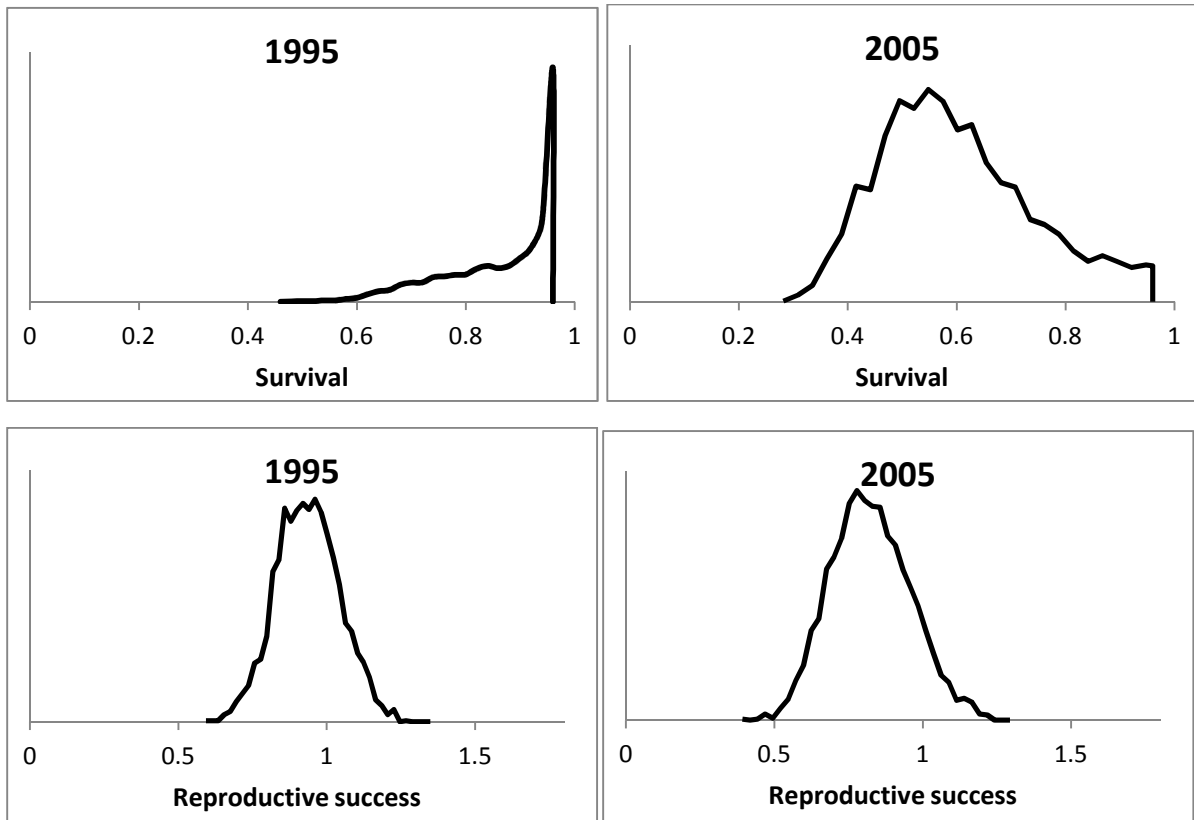


Figure 16: Bayesian marginal posterior distributions of the penguin survival and reproductive success for two years in the time series.