

Modelling the impact of the South African small pelagic fishery on African penguin dynamics

William Robinson¹

Marine Resource Assessment and Management group

Department of Mathematics and Applied Mathematics

University of Cape Town

1 Introduction

This paper incorporates a complete description of the penguin–fish interaction model for Robben Island. All data included are described in full in Section 2. The model equations are described in detail in Section 3. Base case results are presented in Section 4.

At the time of the International Fisheries Stock Assessment Review Workshop in December 2011, how to determine a suitable value or range for the $\tilde{\sigma}$ parameter remained uncertain. This parameter determines the degree of spread of estimated survival rates about those predicted by the relationship with pelagic fish biomass. Results indicated that different assumed values for $\tilde{\sigma}$ (0.05, 0.07, and 0.1) produced substantially different projections of penguin populations. A new modelling framework has since been implemented which has enabled a plausible maximum likelihood estimate of $\tilde{\sigma}$ to be determined.

2 Data

Teams of seabird biologists have been collecting African penguin data for several decades. Variables studied include population size, survival, and indices of reproductive success which include chick condition and growth rate, foraging trip distance and duration, and diet composition (Crawford *et al.* 2011a; Waller 2011; Bouwhuis *et al.* 2007; Pichegru *et al.* 2010). This section describes those data which are used as inputs to the penguin–fish interaction model described in Section 3. Counts of nests and moulting penguins enable estimation of colony population size (Section 2.1). Re-sighting records of tagged penguins enable the estimation of survival rates (Section 2.2). In Section 2.3, information regarding the numbers of penguins which died as a

¹william.robinson@uct.ac.za

direct result of two major oil spills is provided. Section 2.4 covers the time-series of sardine biomasses estimated from acoustic surveys.

2.1 Moult and breeder counts

2.1.1 Background

Various strategies have been attempted for obtaining total penguin abundances at colonies, including total counts of nests or adults and estimates from aerial photography. Estimates of numbers have also been obtained by extrapolation of penguin density in a sample to the area of the whole colony (Randall *et al.* 1986). All of these methods will give a negatively biased population estimate if penguins are absent from the colony at the time of the count. For this reason, counts are ideally made at the height of the egg-laying phase of the breeding season, as during incubation at least one parent must be present at its nest. However, the African penguin has an extended breeding season lasting from January to October. The time of the peak of the breeding season may vary amongst colonies and from year to year, and not all pairs necessarily lay at the same time. Indeed, the proportion of mature penguins which breed varies from year to year (Crawford and Dyer 1995).

Despite these drawbacks, nest counts are the most convenient method of tracking penguin population trends. Since 1987, counts have been made at most penguin colonies in South Africa in most years (Underhill *et al.* 2006). If possible, several counts are made at a colony during the breeding season, and the maximum of these is recorded. Of all the colonies, Robben Island is the best monitored, as it is easily accessible (Crawford *et al.* 1995). The full series of nest counts for Robben Island is given in Table 1.

In order to address the fact that the nest count tallies the breeding portion of the population only, an alternative method of abundance estimation was proposed by Randall *et al.* (1986) based on an annual count of moulting penguins. The premise is that every penguin must moult annually. Until recent observations provided some evidence to the contrary (Wolfaardt *et al.* 2009a), it was believed that African penguins always returned to their breeding colonies to moult. The conspicuous feather-shedding phase of the moult process has been reported by Randall (1983) to

have a mean duration of 12.7 days. Thus, an estimate for the population of a colony which is only slightly negatively biased should be obtainable by summing counts of moulting birds at fortnightly intervals throughout the year. These counts would include both young adults which had not yet commenced breeding, as well as older mature birds which may abstain from breeding in some years. This makes the moult count more robust than the nest count to variations in seasonal conditions.

One requirement for deriving the moult based abundance series is that colonies must be visited regularly, especially around the peak of the moult season. This can give rise to problems, as about 20 visits per year are necessary, which makes the method impractical for many colonies on account of limited resources. Thus, in the Western Cape, annual series of moult counts have been recorded only for Robben Island (since 1988), Boulders (since 1992) and Dassen Island (since 1994) (Crawford *et al.* 2000b; Wolfaardt *et al.* 2009a). Another problem is that not all moulting birds may be visible — some may moult in burrows, and birds moulting inland are usually not counted. As a result, the moult count at Dassen Island (where many birds moult inland) is known not to represent the whole population (Wolfaardt *et al.* 2009a).

In a study at Robben Island, Crawford and Boonstra (1994) found that the maximum nest count was usually obtained in May, which was identified as the peak of the breeding season. Also, just a few counts of moulting birds around the peak of the moult season at the beginning of December gave a good index for the total number of moulters. Furthermore, moult counts and nest counts were found to be highly correlated, and on average the number of moulting adults was 3.2 times greater than the maximum nest count. Thus, reasonable abundance indices could in principle be obtained with considerably less observer effort, but atypical conditions could cause a shift in the timing of the season peaks, resulting in high variance about the true numbers. Inter-annual variability in moult phenology is thought to be influenced by factors such as the success or failure of the previous breeding season, food availability, and major oil spills (Crawford *et al.* 2006).

A high degree of synchrony has been observed in the African penguin moult season, especially at Robben Island (Underhill and Crawford 1999; Wolfaardt *et al.* 2009a). As most penguins in the Western Cape moult between September and January, the annual moult season is defined to run

over the split year beginning on 1 July and lasting until 30 June the following year. Underhill and Crawford (1999) used linear interpolation to estimate the number of penguins moulting on each day between the days when actual counts were made. The resulting numbers could be summed and then divided by 12.7 (the mean duration of the feather-shedding phase) to obtain an aggregate count of moulters for the whole year. This method should give good results if intervals between counts are reasonably short, but the estimation becomes rather coarse if there are large gaps between counts, especially over the peak period (November and December).

The juvenile moult season is often characterized by two peak periods. The first peak occurs in late spring, coincident with the peak period of the adult moult at Robben Island, and the second peak occurs late in the summer (Underhill and Crawford 1999). Kemper and Roux (2005) and Wolfaardt *et al.* (2009a) suggest that the first peak may comprise birds which delay moulting into their adult plumage until their second summer when they moult with the majority of the colony (“skippers”), while birds which fledge early in the year may moult during late summer or autumn, soon after they reach 12 months of age (“squeezers”).

2.1.2 Double Gaussian method for interpolating moult counts

Because of the difficulties mentioned above associated with the linear interpolation method for obtaining annual moult count estimates, an alternative method of modelling the expected number of birds moulting on each day d as the sum of two Gaussian functions was developed:

$$\hat{M}_d = \frac{N_1}{\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{(d - \mu_1)^2}{2\sigma_1^2}\right] + \frac{N_2}{\sqrt{2\pi\sigma_2^2}} \exp\left[-\frac{(d - \mu_2)^2}{2\sigma_2^2}\right] \quad (1)$$

where:

$N_1 + N_2$ is the total number of moulting birds which would be counted if counts were made daily (if the moult season was “infinitely” long),

μ_1 and μ_2 are the days on which the two peaks of the moult season occur, and

σ_1 and σ_2 characterize the degree of synchrony within each moulting period.

The sum of two Gaussian curves was chosen because the juvenile penguin moulting season is known to have two peaks, and the additional flexibility is useful as it can take account of asymmetric shapes of the overall distribution.

Values for the six parameters in equation (1) are obtained by minimizing the negative log-likelihood for the Poisson process of sighting moulting birds given the data:

$$-\ln L = \sum_{i=1}^n \left(\hat{M}_i - M_i^{\text{obs}} \ln \hat{M}_i \right) \quad (2)$$

where i indexes each moulting count with a total of n such observations made during the season. This method seems likely to be more robust than the linear interpolation method, both in the case of noisy data and when the intervals between consecutive counts are several weeks or even months.

If σ_1 and/or σ_2 are large, the Gaussian functions have wide tails, resulting in values of $\hat{M}_d \gg 0$ for days outside of the split year in question. This has been taken into account by subtracting the tails T_1 and T_2 of each of the Gaussian functions:

$$T_1 = N_1 \left[\text{erfc} \left(\frac{\mu_1 - d_1}{\sqrt{2}\sigma_1} \right) + \text{erfc} \left(\frac{d_{L+1} - \mu_1}{\sqrt{2}\sigma_1} \right) \right] \quad (3)$$

$$T_2 = N_2 \left[\text{erfc} \left(\frac{\mu_2 - d_1}{\sqrt{2}\sigma_2} \right) + \text{erfc} \left(\frac{d_{L+1} - \mu_2}{\sqrt{2}\sigma_2} \right) \right] \quad (4)$$

where the subscripts 1 and L indicate the first and last days of the moulting year (1 July and 30 June respectively) and the complementary error function is defined as:

$$\text{erfc}(x) = \frac{1}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt \quad (5)$$

The total number of birds moulting in one year is then calculated as:

$$M = \frac{N_1 + N_2 - T_1 - T_2}{12.7} \quad (6)$$

By using the optimization program AD Model Builder (ADMB) (Fournier *et al.* 2011) to fit the double Gaussian curves to the annual moulting counts, the Hessian-based coefficient of variation

s is readily obtainable for each estimate under the assumption that model errors are Poisson distributed. However, the actual errors are greater than this assumption suggests, and hence account needs to be taken of overdispersion in calculating reliable coefficients of variation. After grouping observations so that each $\hat{M}_i > 5$, the overdispersion for each annual estimate is calculated as:

$$D = \sqrt{\frac{1}{n^* - p} \sum_{i=1}^{n^*} (M_i^{\text{obs}} - \hat{M}_i)^2} / \hat{M}_i \quad (7)$$

where n^* is the number of observations in that year (after grouping) and p is the number of parameters estimated. CV s are then calculated as:

$$CV = s\tilde{D} \quad (8)$$

where \tilde{D} is the median of all overdispersion parameters calculated for each series.

An alternative to the maximum likelihood approach is to use the Markov-chain Monte Carlo (MCMC) method to obtain a Bayes posterior distribution for M . This was attempted for a few years and gave very similar results to the Hessian based estimates of CV .

2.1.3 Results

Double Gaussian functions were fitted to all the available adult and juvenile moult count data for Robben Island. The number of observations, the estimated numbers of adult and juvenile moulters and the corresponding CV s are given for each year in Table 2. The differences between the estimates calculated following the double Gaussian method and the linear interpolation method are given in Table 3.

The only notable differences between the double Gaussian results and those published by Wolfaardt *et al.* (2009a) are for the seasons 1996/1997 and 2001/2002. The underlying reasons for these discrepancies are discussed below.

According to Underhill and Crawford (1999): “The abnormal pattern in 1996/97 was a result of errors in the November and December counts, which were undertaken by an inexperienced observer”. Consequently, three data points were excluded when fitting the double Gaussian

function for this season. The result is a 30% larger estimate for adult moulters and a 7% larger estimate for juvenile moulters for 1996/1997.

The 2001/2002 estimate for Robben Island is unreliable since only nine counts were taken during that year, with none before November. A decrease in moulters in the year following the *Treasure* oil spill would perhaps be expected, as was the case in 1995/1996 following the sinking of the *Apollo Sea*. Compared to the results of Wolfaardt *et al.* (2009a), the double Gaussian estimates given here are 25% and 31% lower for the adults and juveniles respectively.

2.1.4 Discussion

Results from deriving moult count estimates via the method of fitting a double Gaussian are in most cases similar to the results from linearly interpolating the count data. However, particularly in seasons where there are large gaps between some counts, the double Gaussian approach seems preferable. Accordingly the associated estimates are used in the modelling presented below.

The motivation for use of the double Gaussian was bimodality in the juvenile count distributions, and the corresponding estimates of μ_1 and μ_2 are broadly consistent with a first peak in early summer and a second peak in late summer (Table 4). The fits to the adult counts do not show a similar pattern (Table 5), and the justification of the use of two Gaussian curves there is more to provide greater overall shape flexibility. A more refined approach might consider fitting through the use of a Hermite polynomial series expansion, but this does not seem justified here as the double Gaussian approach achieves a sufficiently good fit to the data to provide estimates of adequate accuracy for population modelling purposes.

The estimates of overdispersion always exceed 1, but are also very variable and sometimes very large. The last category generally arises from instances where counts are relatively high at the very beginning and/or very end of the 12 month period considered, which are times when the double Gaussian predicts very low values. Although these differences are not large in the context of the overall count estimate M , they can influence the estimates of overdispersion D considerably. For this reason, CV s have been presented based on the median of the D estimates

for each season. This is of course a somewhat coarse approach, but seems useful to provide some broad indication of the precision of the estimates obtained.

A concern is that in recent years, numbers of adult moulters have been lower than expected based on the nest counts. While one possibility is that a higher proportion of birds is moulting inland in areas not checked by the counters, it appears that the main reason for this discrepancy is that substantial numbers of Robben Island breeders are not returning to the colony to moult. A rapidly increasing number of penguins has been observed moulting at the Stony Point colony at Betty's Bay, amounting to about 5 000 adults in the 2009/2010 moult season (Crawford *et al.* 2011a). Several of these birds are known to have bred previously at Robben Island and Dassen Island (Animal Demography Unit (ADU), unpublished data). The Stony Point nest count doubled in size from 2010 to 2011, which indicates that the colony is attracting large numbers of possibly permanent immigrants. If this trend continues, it would have problematic implications for the future application of the modelling process that follows.

2.2 Re-sightings of banded birds

Capture-mark-recapture analysis is a common method for tracking changes in avian life history parameters. In particular, the ability to track marked individuals through time enables researchers to obtain estimates for annual survival rates. This is possible even though the marked birds may not be seen every year and, indeed, their time of death is usually unknown (Lebreton *et al.* 1992). The calculations for these survival estimates are independent of abundance data (e.g. moult counts and nest counts), allowing for the estimation of immigration in an integrated population model.

African penguins have been tagged with individually identifiable flipper bands since 1947. Prior to 1972, penguins had to be caught in order to read the bands, but in that year stainless steel bands were introduced which could be read with the aid of binoculars (Cooper and Morant 1981; Randall *et al.* 1987). Re-sightings of banded African penguins have been recorded throughout their range, including at Robben Island since 1989 (Whittington *et al.* 2005b). A few hundred adult penguins in addition to some juvenile birds were fitted with bands annually. Re-sighting efforts were greatly increased following the *Apollo Sea* oil spill in 1994. During the massive

relocation and cleaning operation following the *Treasure* oil spill in 2000, many thousands of penguins were banded. As a result of this large tagging cohort, most of the re-sightings at Robben Island on record come from that group.

There has been much debate about the possibility that flipper bands may be harmful to penguins (Petersen *et al.* 2005). Studies on various species have shown possible negative effects of bands on swimming efficiency, survival, and breeding success. Factors thought to influence the effects include the material, size and colour of bands, environmental conditions, and the species concerned. Studies on African penguins have indicated that the birds are not negatively affected by the bands (Barham *et al.* 2008). Nevertheless, there is an interest in the development of alternative techniques for monitoring the African penguin, such as automated photo-recognition (Sherley *et al.* 2010) or subcutaneously implanted transponders (Jackson and Wilson 2002).

The database of African penguin re-sighting records is administered by the South African Bird Ringing Unit (SAFRING), and the number of records it contains is approaching a quarter of a million. There are some 76 367 records for birds seen at least once at Robben Island for the years 1970–2010. Of the 25 248 individuals banded, 15 603 were re-sighted at least once. The records indicate that 16 807 were banded as adults, 6 790 were banded as chicks, and the age group when banded for the remaining 1 651 penguins is unknown. The database includes 734 recoveries of bands from dead penguins, which is less than 1% of the total records.

The subset of records used in this analysis includes only the birds banded as adults, and excludes dead recoveries. Since the model presented in this work is for the Robben Island colony, an effort has been made to exclude the capture histories of penguins which were occasional visitors to the island. Hence, capture histories were rejected if more than half of the sighting locations for an individual were not at Robben Island.

A further consideration is that in an ideal capture-mark-recapture experiment, the extent of the capture season should be negligible compared to the interval for which survival rates are to be calculated (a year, in this case). However, penguin re-sightings are made throughout the year, meaning that the length of time between sightings of a bird recorded in consecutive years could in fact be anything from less than a month to as much as 23 months. In order to minimize any

possible bias which this might cause, a decision was made to include only records of sightings made during the months April–September, restricting the inter-sighting period to between 6 and 18 months. Nearly 80% of all sightings at Robben Island have taken place during this autumn–winter period.

The set of records consequently selected for use in this analysis comprises 12 207 capture histories for the period 1989–2010.

2.3 Major oil spills

The proportion of the African penguin population which has been affected by oil spills is considered to be among to highest of all seabird species globally (Wolfaardt *et al.* 2009b). The vast majority of oiled Robben Island penguins were contaminated as a result of the wrecks of the bulk ore carriers *Apollo Sea* on 20 June 1994 and *Treasure* on 23 June 2000 (Wolfaardt *et al.* 2008). Because such large numbers of penguins were affected by these two incidents during the relatively short existence of the Robben Island colony in recent times, special account is taken in the population model of the deaths of both adults and chicks as a direct result of these oil spills.

2.3.1 *Apollo Sea*

Shortly after sailing from Saldanha Bay in June 1994, the *Apollo Sea* sank near Dassen Island. Fuel oil washed ashore at both Dassen Island and Robben Island, and approximately 10 000 penguins were oiled. Underhill *et al.* (1999) suggest that nearly all penguins affected reached land alive, as during searches along the coastline very few dead penguins were found despite the prevailing onshore wind. This is consistent with observations following other spills (Wolfaardt *et al.* 2009b). Of the penguins rescued, about 2 400 came from Robben Island, half of which were successfully released, resulting in the loss of about 1 200 adult birds (Crawford *et al.* 1999).

The number of chicks which died as a result of the spill is more difficult to estimate. Wolfaardt *et al.* (2009b) give two methods of estimation. In the first method, an observed loss of 23 chicks from a sample of 38 nests is extrapolated over the whole colony which had 2 155 active nests at the time of the oil spill. The result is a projected loss of 1 680 chicks. The second estimate is based on the number of adults which were removed from the island, which could have represented

up to 1 500 nests. Multiplying by 0.47, which is the mean number of chicks fledged per nest at Robben Island during 1989–1995, gives a loss of 705 chicks due to oiling for the season. Note that the first estimate does not account for chicks which would not have fledged due to other factors, while the second estimate does not account for losses due to the disturbance to penguins which were not oiled.

As a rough estimate for the proportion of chicks lost due to the oil spill, the average of the two estimates given above, divided by the estimated total number of chicks which would have been present at the time of the spill is used. The observed mean clutch size in 1994 was 1.92 eggs per nest (Crawford *et al.* 1999). Thus the estimated proportion of chicks lost due to the oil spill is:

$$p_{1994}^{\text{oil}} = \frac{(1\,680 + 705)/2}{2\,155 \times 1.92} = 0.29 \quad (9)$$

2.3.2 *Treasure*

The bulk ore carrier *Treasure* sank between Robben Island and Dassen Island, and on 24 June 2000 spilt oil washed ashore on Robben Island, contaminating any penguin attempting to leave the island or to return after foraging. Crawford *et al.* (2000a) relate that of the 19 000 oiled penguins collected for cleaning, 14 825 were caught at Robben Island. By 18 July, about 1 659 of the oiled adult and juvenile penguins which had been captured had died. A total of 19 506 penguins which had not been oiled were captured, with 7 161 of these being removed from Robben Island, to safeguard them from oiling. Of these, 241 penguins died during or after transportation to Cape Recife (800 km to the East) where they were released to swim back so as to arrive after the oil had dissipated. During a search of about 250 nests at the Robben Island colony, only four oiled penguins were found dead, which can be extrapolated to the death of 90 adult penguins in the entire colony. As the number of penguins collected at Robben Island was several thousand more than the estimated population of the colony (18 000), a sizeable proportion of these penguins must have been from other colonies. A rough estimate for the total number of adults collected from the island plus the few thousand penguins which remained, is 26 500. The

number of Robben Island adults² which died due to oiling is thus estimated to be:

$$2N_y^{\text{oil}} = \left(1\,659 \times \frac{14\,825}{19\,000} + 241 \times \frac{7\,161}{19\,506} + 90 \right) \times \frac{18\,000}{26\,500} = 1\,000 \quad (10)$$

The number of chicks which died at Robben Island after the *Treasure* spill is estimated to be about 3 000 (Crawford *et al.* 2000a). Furthermore, 367 of the 3 350 chicks which were collected for captive rearing also died, 2 643 of which came from Robben Island. There were an estimated 6 000 chicks at the Robben Island colony at the time of the spill. The proportion of chicks lost at the time of the spill is thus estimated to be:

$$\left(3\,000 + 367 \times \frac{2\,643}{3\,350} \right) / 6\,000 = 0.55 \quad (11)$$

However, breeding attempts were likely successfully completed both before and after the spill, reducing the overall impact of the spill. Thus the proportion assumed for p_{2000}^{oil} in this analysis is reduced somewhat to $p_{2000}^{\text{oil}} = 0.4$.

The values for N_y^{oil} and p_y^{oil} are listed in Table 6. Note that for the model the numbers of adults given above are halved since only the female component of the population is considered in the model.

2.4 Pelagic biomass

Acoustic surveys have played a major role in the assessment of South African pelagic fish resources since 1984. The annual May survey estimates the strength of the recruitment to the sardine and the anchovy stock each year, while the annual November survey estimates the biomass of their 1+ populations. The early surveys established that during the 1980s anchovy was more abundant and widespread than had been previously thought, and the sardine resource was recovering strongly following heavy depletion (Hampton 1992). Additional data from subsequent years suggested relationships amongst trends in the biomass data and spatial distributions (Barange *et al.* 1999). After the upgrading of survey equipment in 1997, earlier data were revised to correct

²The factor of 2 appears on the left hand side of equation (10) because N_y refers only to the female component of the population in the model which follows.

for echosounder receiver saturation at high signal levels and attenuation effects in dense sardine shoals (Coetzee *et al.* 2008). With the incorporation of these revisions, comparable data series for sardine and anchovy now extend unbroken from 1984 to 2011 (de Moor *et al.* 2008; Shabangu *et al.* 2011).

The surveys are divided into strata which allows for comparisons of abundance in a specific region over time. Of interest for this study is the biomass series of sardine spawners west of Cape Agulhas (Moseley *et al.* 2011). In recent years, the divisions between the survey strata have remained fixed. The area west of Cape Agulhas corresponds to strata A–C (see Figure 1).

Table 7 lists the annual 1+ biomass of sardine observed in the standard November survey area, the area west of Cape Agulhas (strata A–C), and the proportion observed west of Cape Agulhas. Note that, in absolute terms, it is possible that the survey estimates may be biased. For example, there may be an error in the acoustic target strength used, or the survey may not cover the entire range of the resource. However, such bias is not of importance here as the model needs only a relative index of abundance.

3 Penguin–fish interaction model

3.1 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. The initial population size (at the start of year $y_0 = 1988$) N_0 and its age structure are:

$$N_{1988,a} = \begin{cases} N_0 e^{-a\lambda} & \text{for } 1 \leq a < A \\ \frac{N_0 e^{-a\lambda}}{1 - e^{-\lambda}} & \text{for } a = A \end{cases} \quad (12)$$

where A is the plus-group age, taken here to be $A = 5$. Both N_0 and λ are parameters whose values are estimated.

The following equations describe the population trajectory:

$$N_{y+1,a} = \begin{cases} \frac{1}{2}H_y \exp(-M_y^{1/2}) (1 - p_y^{\text{oil}}) \sum_{a=a^*}^A N_{y,a} & \text{if } a = 1 \\ N_{y,a-1}^* \exp(-M_y) & \text{if } a = 2 \\ N_{y,a-1}^* \exp(-M_y) + I_y & \text{if } a = 3 \\ N_{y,a-1}^* \exp(-M_y) & \text{if } 4 \leq a < A \\ (N_{y,a-1}^* + N_{y,a}^*) \exp(-M_y) & \text{if } a = A \end{cases} \quad (13)$$

where:

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),

M_y is the adult (post 1 January of first year of life) annual mortality rate in year y (taken to the power $1/2$ in the equation for $a = 1$ since the peak of the breeding season is assumed to be on 1 May),

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

a^* is the age at which the penguins first attempt to breed (taken here to be $a^* = 4$),

I_y is the number of penguins (all assumed to be of age 3) immigrating to Robben Island in year y (investigations allowing more flexible variation over time suggested that it is appropriate to estimate this as four constant levels for the periods 1989–1990, 1991–1993, 1994–1996, and 1997–1999),

$N_{y,a}^*$ is the number of penguins in year y of age a which survive the catastrophic oil spills:

$$N_{y,a}^* = N_{y,a} S_y^{\text{oil}} \quad (14)$$

where

$$S_y^{\text{oil}} = 1 - \frac{N_y^{\text{oil}}}{\exp(-M_y^{m_j/12}) \sum_{a=1}^A N_{y,a}} \quad (15)$$

N_y^{oil} is the number of juvenile and adult penguins estimated to have died as a result of oiling in year y , and

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

Table 6 lists model inputs related to the major oil spills. Parameter values which are fixed on input to the model are given in Table 8.

3.1.1 Annual adult mortality

Adult mortality is assumed to be age-independent. Note that the mortality rate M_y does not include loss of penguins as a result of the major oil spills, while the total survival rate S_y , defined as

$$S_y = \exp(-M_y) S_y^{\text{oil}} \quad (16)$$

includes both natural mortality and oil-induced mortality.

3.1.2 Age at first breeding

In a study at Robben Island over the period 1995–1999, Whittington *et al.* (2005a) recorded the ages at which penguins which had been banded as chicks were first observed breeding. Their results indicated that 32% were breeding by age 3, 57% by age 4, and 89% by age 5. It is possible that the first breeding attempt of some penguins may have been missed, so that the true percentages may be higher than those recorded. These figures are similar to those observed in an earlier Robben Island study (Crawford *et al.* 1999), at other African penguin colonies (Randall 1983), and in studies of other species of penguin.

The choice here for the age at which penguins enter the breeding population in the model $a^* = 4$ reflects the median of observed ages. Note that this is the age at which the penguins become *potential* breeders, and no assumptions are made about the proportion which actually attempt breeding. It is thought that this proportion may be related to pelagic biomass (Crawford *et al.* 1999).

3.2 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). For reproductive success, the estimates are drawn from a beta distribution which ensures that biologically plausible bounds are respected. For adult survival, the alternative approach below was developed and has been found to work satisfactorily.

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

$$B_{S,y} = I_{S,y} / \max\{I_S\} \quad (17)$$

Similarly, reproductive success depends on a similar index of biomass level:

$$B_{H,y} = I_{H,y} / \max\{I_H\} \quad (18)$$

In the base case model implemented here, I_S is taken to be the November survey sardine spawner biomass west of Cape Agulhas (Figure 2). The series I_H is taken to be the May survey anchovy recruit biomass west of Cape Infanta (Figure 3).

3.2.1 Adult survival

Annual adult penguin natural mortality M_y is assumed to depend on prey abundance $B_{S,y}$, modelled as follows:

$$M_y = M_{\min} + f_S(B_{S,y}) e^{X_y} \quad (19)$$

Note that M_y does not include any mortality due to the major oil spills (see equations (13)–(16)). The M_{\min} term is included to impose a biologically plausible lower bound on the mortality rate, set here to be $M_{\min} = 0.04$. The random effects X_y are distributed $N(0, \sigma_y^2)$ with:

$$\sigma_y = \sqrt{\exp[\tilde{\sigma}/f_S(B_{S,y})]^2 - 1} \quad (20)$$

Thus there is a lognormal random effect, but since the σ_y depend on the biomass $B_{S,y}$, the M_y distributions will all have exactly the same standard deviation. This is appropriate since then the data related to each year receive roughly equal weighting in the model fitting process, and, when projecting, high resource biomass does not force low penguin mortality.

The negative log-prior added to the negative log-likelihood for each year to reflect assumptions made above for the X_y random effects is:

$$P_X = \sum_y \left[\ln \sigma_y + \frac{1}{2} \left(\frac{X_y}{\sigma_y} \right)^2 \right] \quad (21)$$

An additional penalty term (“prior”) is added to aid estimation stability. The term chosen follows, and ensures that the annual mortality rates are evenly distributed about the curve relating mortality and biomass (i.e. the sum of the residuals is forced to zero):

$$P_{B-M} = 10^5 \left\{ \sum_y \left[M_y - (M_{\min} + f_S(B_{S,y})) \right] \right\}^2 \quad (22)$$

3.2.2 Biomass–mortality relationship

A variety of different functional forms have been implemented for the assumed relationship between pelagic biomass and penguin adult annual survival. A reason for the difficulty in choosing an appropriate function for this relationship is that the survival values are estimated within the model, and thus tend to depend on the relationship itself.

The first candidate function, which has an inverse power form, was defined as:

$$f_S(B_{S,y}) = M_{\min}^* + \left(\alpha + \frac{B_{S,y}}{\beta} \right)^{-n} \quad (23)$$

The motivation for including the M_{\min}^* term is that, when calculating M_y in equation (19), the possibility is excluded that the term additional to could go to zero when biomass is high. This would be problematic as then σ_y in equation (20) would go infinite, rendering the associated data point redundant because of its infinite variance. The lower bound on achieved M_y remains

M_{\min} , but the lower bound on the median of its distribution is $M_{\min} + M_{\min}^*$. For the base case model, the constant value used was $M_{\min}^* = 0.02$.

Instead of α and β , equation (23) can be re-parameterised in terms of the values of $f_S(B_{S,y})$ at $B_S = 0.1$ and $B_S = 1$ in order to introduce estimable parameters which are more orthogonal (i.e. the new parameters have low correlation) which assists with statistically stable estimation. If $U = f_S(B_S = 1)$ and $V = f_S(B_S = 0.1) - f_S(B_S = 1)$, then

$$\begin{aligned}\beta &= \frac{0.9}{(U - M_{\min}^*)^{-1/n} - (U + V - M_{\min}^*)^{-1/n}} \\ \alpha &= (U - M_{\min}^*)^{-1/n} - \frac{1}{\beta}\end{aligned}\tag{24}$$

The estimable parameters are then U , V , and n , where sensibly $V \geq 0$.

A concern arising from analyses with this functional form is that it turned out that the likelihood could always be increased by raising the value of the power parameter n , leading to difficulty in choosing an appropriate prior for n . This indicates that the functional form is perhaps not the most appropriate. An alternative function then considered is defined by connected straight lines between the points:

$$\begin{aligned}f_S(1.0) &= \alpha \\ f_S(B^*) &= \alpha \\ f_S(0.0) &= \alpha + \beta\end{aligned}\tag{25}$$

Here, the relationship is flat above a critical biomass level B^* . Examining the scatterplot of mortality estimates suggested a preference for the relationship to flatten as the biomass tends towards zero. This suggests that the logistic function may be suitable:

$$f_S(B_{S,y}) = \alpha + (\beta - \alpha) \frac{1}{1 + \exp[(B_{S,y} - \gamma) / \delta]}\tag{26}$$

This four-parameter function was found to perform better than the other forms. However, the data prefer a very steep gradient in increasing mortality as prey biomass drops below a critical value (i.e. an estimate for δ very close to zero whose biological realism seems questionable). For

this reason, a prior for the δ parameter was added, which itself has a logistic form to exclude very high gradients:

$$\text{prior}(\delta) = \frac{1}{1 + \exp[-(\delta - d^*)/\sigma_\delta]} \quad (27)$$

The effect of this prior is for a more gradual increase in penguin mortality as sardine biomass decreases. Values of the parameters of this logistic were chosen to be $d^* = 0.02$ and $\sigma_\delta = 0.005$, which give seemingly realistic results.

3.2.3 Reproductive success

The parameters $\bar{H}_y^* = f_H(B_{H,y})$ are estimated on the interval $[0, 1]$. These are transformed predicted annual reproductive success rates \bar{H}_y which fall in the interval $[0, H_{\max}]$:

$$\bar{H}_y = H_{\max} \bar{H}_y^* \quad (28)$$

The relationship between \bar{H}_y^* and the fish biomass index $B_{H,y}$ is assumed to be a constant in the base case given that analyses did not provide any indication of dependence:

$$f_H(B_{H,y}) = \eta \quad (29)$$

but alternative functional forms may readily be considered.

The parameters H_y^* are estimated for each year on the interval $[0, 1]$ and then transformed to the range $[0, H_{\max}]$ to obtain the annual reproductive success rates H_y as follows:

$$H_y = H_{\max} H_y^* \quad (30)$$

The H_y^* parameters are assumed to be beta-distributed about the predicted values \bar{H}_y^* . Setting $\kappa_{H,y} = \bar{H}_y^* (1 - \bar{H}_y^*) / \sigma_H^2 - 1$, the parameters of this beta distribution $\alpha_{H,y}$ and $\beta_{H,y}$ are:

$$\begin{aligned} \alpha_{H,y} &= \bar{H}_y^* \kappa_{H,y} \\ \beta_{H,y} &= \kappa_{H,y} (1 - \bar{H}_y^*) \end{aligned} \quad (31)$$

In order to obtain beta distributions which have a single mode and which vanish at the ends of the interval, the estimation procedure ensures that both $\alpha_{H,y}$ and $\beta_{H,y}$ are greater than 1.

The negative log-prior added to the negative log-likelihood, which reflects the assumption that H_y^* is beta-distributed, is:

$$P_H = \sum_y \left[\ln \frac{\Gamma(\alpha_{H,y} + \beta_{H,y})}{\Gamma(\alpha_{H,y}) \Gamma(\beta_{H,y})} - (\alpha_{H,y} - 1) \ln H_y^* - (\beta_{H,y} - 1) \ln (1 - H_y^*) \right] \quad (32)$$

An additional penalty term ensures that the annual reproductive success rates are evenly distributed about the assumed relationship, i.e. the sum of the residuals is forced to zero (this was found to aid estimation stability):

$$P_{B-H} = 10^5 \left[\sum_y (H_y - \bar{H}_y) \right]^2 \quad (33)$$

3.2.4 Immigration

The inclusion of tag-recapture information (see Section 3.3.2) allows, in principle, for the estimation of immigration in the model. All immigrants are assumed to be pre-breeders in adult plumage (see equation (13)). Initially, a separate estimate was made for the number of birds immigrating to Robben Island each year, but there was not support for the addition of so many parameters. Therefore, years were grouped together and an appropriately parsimonious model was selected using AIC. There was no support for immigration from the year 2000 onwards.

3.3 Model fitting

3.3.1 Fitting to moult counts

The population model is fitted to annual moult count data for both adult and juvenile birds. Assuming an equal sex ratio, the numbers of adult moulters in Table 2 are halved to correspond to females only. The counts are taken into account through the following negative log-likelihood

functions:

$$\begin{aligned} -\ln L_M &= \sum_y \left[\ln \sigma_{M,y} + \frac{1}{2\sigma_{M,y}^2} \left(\ln N_y^{\text{obs}} - \ln N_y^{\text{model}} \right)^2 \right] \\ -\ln L_J &= \sum_y \left[\ln \sigma_J + \frac{1}{2\sigma_{J,y}^2} \left(\ln J_y^{\text{prop}} - \ln J_y^{\text{model}} \right)^2 \right] \end{aligned} \quad (34)$$

$\sigma_{M,y}$ 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). The terms in the likelihood for the adult moult counts are weighted according to the coefficients of variation of the observations with:

$$\sigma_{M,y} = \sqrt{CV(N_y^{\text{obs}})^2 + 0.05 + \sigma_{\text{add}}^2} \quad (35)$$

where the additional variance σ_{add}^2 is an estimable parameter, with the true additional variance forced to be at least 0.05^2 . The value of the variance of the proportion of juveniles σ_J is fixed at $\sigma_J = 0.1$.

N_y^{model} is the number of female birds in adult plumage (aged 2 and over) counted in year y , calculated as:

$$N_y^{\text{model}} = q_M \sum_{a=2}^A N_{y,a} S_y^{1/12} \quad (36)$$

The proportion of these birds susceptible to observation is assumed here to be $q_M = 0.9$. The survival rate S_y is taken to the power of $1/12$ since the peak of the counting season is at the end of November.

J_y is the proportion of juvenile birds in the model at the time of the moult count, calculated as:

$$J_y = p_J N_{y,1} / \sum_{a=1}^A N_{y,a} \quad (37)$$

Note that $S_y^{1/12}$ factors in the numerator and denominator cancel. It has been suggested that some moulting juveniles are counted as adults, as distinguishing the different forms

can be difficult during the last few days of the moult. The factor p_J allows for this, but is assumed here to be $p_J = 1$.

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

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

3.3.2 Fitting to tag data

In addition to the moult counts, tag data for the period 1989–2010 have been incorporated into the analysis. This provides an independent estimate of annual survival rates and allows for the estimation of immigration. Based on the Cormack-Jolly-Seber (CJS) model (Cormack 1964; Jolly 1965; Seber 1965), the multinomial likelihood of the encounter histories is calculated through the estimation of re-sighting probabilities and survival rates for each year. Results from this implementation of the standard CJS model were checked against results produced by program MARK (White and Burnham 1999), and were found to be in agreement.

A variation on the standard Cormack-Jolly-Seber model which has been included in this analysis is the estimation of a transient factor (Pradel *et al.* 1997) relating to tags which are never re-sighted after deployment. For example, tags improperly attached could fall off, or animals might move to an area which is inaccessible or infrequently visited by tag spotters. In some cases, an improperly attached tag could contribute to a premature fatality. These short-term effects are modelled as additional mortality (which may in fact incorporate emigration) in the year that a penguin is banded. Note that this “transient” rate does not depend on prey abundance in any way.

Initially, a single estimated transient rate was applied to all years, except for the years of the two major oil spills, 1994 and 2000. However, a clear lack of fit to the tag-resighting data for some tag cohorts indicated that the assumption of this constant transient rate is inappropriate here. This was addressed by estimating a separate transient rate for each year. This improves the fit but it is questionable whether the addition of so many parameters is statistically justifiable.

Using AIC as the basis for model selection, various options of pooling of estimates for different years were explored.

The log-likelihood function for the multinomial distribution, ignoring the constant multinomial coefficient, is:

$$\ln L_{T-R} = \frac{1}{\hat{c}} \sum_{j=1}^m n_j \ln \hat{p}_j \quad (38)$$

where:

\hat{p}_j is the estimated probability of the j th encounter history occurring,

n_j is the number of times which that history has been observed,

m is the number of unique encounter histories, and

\hat{c} is a scalar included to account for overdispersion.

The value used for \hat{c} is obtained from analysing the data using MARK (see Table 8).

3.3.3 Objective function minimized

The overall negative log-posterior is obtained by summing penalties (negative log-priors) (21), (22), (32), and (33), and subtracting the log-likelihoods (34) and (38):

$$-\ln P_{\text{post}} = P_X + P_{B-M} + P_H + P_{B-H} - \ln L_M - \ln L_J - \ln L_{T-R} \quad (39)$$

4 Results

The model described above was implemented in ADMB (Fournier *et al.* 2011). Parameters listed in Table 8 were fixed on input.

A difficulty arises regarding the estimation of the parameter $\tilde{\sigma}$, which quantifies the degree of spread of survival estimates about the assumed relationship with pelagic fish abundance. When treating $\tilde{\sigma}$ as any other parameter in the maximum likelihood estimation, the program converges to a solution with $\tilde{\sigma}$ tending towards zero. This is because the likelihood term P_X (equation

(21)) is minimized when both the X_y parameters and σ_y (and hence $\tilde{\sigma}$ from equation (20)) tend to zero.

This “difficulty” is not, however, a real one. It arises from an incorrect statistical treatment of the X_y parameters in a frequentist context in circumstances where these residuals relate entirely to estimable model quantities rather than differences between model estimates and data. The appropriate approach is to treat the X_y parameters as random effects in a mixed effects model, and then integrate them out of the likelihood so that $\tilde{\sigma}$ is the only remaining estimable parameter from this term. In a full Bayesian context, integration over all X_y parameters is effected, so in theory no problem should occur.

However, the MCMC method for generating a Bayesian posterior distribution fails in this case. The reason for this is because the constraint in equation (22) is very tight, and hence the region within which the negative log-likelihood is near its minimum is very small. Jumps away from the starting value give such large changes in the negative log-likelihood that these jumps are always rejected in the MCMC algorithm. An attempt to attain the relative likelihood of different values of $\tilde{\sigma}$ using the Deviance Information Criterion (DIC) also failed for the same reason.

Instead, therefore, ADMB-RE was used. With this program, random effects are integrated out using the Laplace approximation (Skaug and Fournier 2006). Once a maximum likelihood estimate was obtained of $\tilde{\sigma} = 0.088$ (with an associated Hessian-based standard error of 0.013), MCMC was used to integrate over the other parameters to obtain posterior distributions.

Strictly the integration of the random effects should occur at each step of the MCMC process to obtain “step-specific” $\tilde{\sigma}$ estimates. This was attempted, but it proved beyond the capabilities of the software to obtain convergence. This does mean, however, that the results obtained are conditional on the ADMB-RE maximum likelihood estimate for $\tilde{\sigma}$, and make no allowance for the variance about this estimate. The implications of this for the results will be checked by numerically integrating MCMC results over a range of values for $\tilde{\sigma}$ with weighting proportionate to the likelihood profile for the $\tilde{\sigma}$ estimate.

Parameters estimated within the model are listed in Table 9. Fits to the data and residual patterns are plotted in Figures 4–6. Further plots show time-series of estimated adult annual

survival (Figure 7) and female moult counts (Figure 8), with 90% probability intervals. Figure 9 shows estimated immigration, and Figure 10 shows estimated “transient” rates associated with the tag-recapture data. Figure 11 compares projected penguin abundances for assumed future sardine biomass levels of $B_S = 0.2$ and $B_S = 0.3$.

References

- Barange M, Hampton I, and Roel BA.** 1999. *Trends in the abundance and distribution of anchovy and sardine on the South African continental shelf in the 1990s, deduced from acoustic surveys.* South African Journal of Marine Science, 21: 367–391.
- Barham PJ, Underhill LG, Crawford RJM, Leshoro TM, and Bolton DA.** 2008. *Impact of flipper-banding on breeding success of African penguins *Spheniscus demersus* at Robben Island: comparisons among silicone rubber bands, stainless-steel bands and no bands.* African Journal of Marine Science, 30: 595–602.
- Bouwhuis S, Visser GH, and Underhill LG.** 2007. *Energy budget of African Penguin *Spheniscus demersus* chicks.* In: SP Kirkman (editor), *Final Report of the BCLME (Benguela Current Large Marine Ecosystem) Project on Top Predators as Biological Indicators of Ecosystem Change in the BCLME*, chapter 15, 125–127. Avian Demography Unit, Cape Town.
- Coetzee JC, Merkle D, de Moor CL, Twatwa NM, Barange M, and Butterworth DS.** 2008. *Refined estimates of South African pelagic fish biomass from hydro-acoustic surveys: quantifying the effects of target strength, signal attenuation and receiver saturation.* African Journal of Marine Science, 30: 205–217.
- Cooper J and Morant PD.** 1981. *The design of stainless steel flipper bands for penguins.* Ostrich, 52: 119–123.
- Cormack RM.** 1964. *Estimates of survival from the sighting of marked animals.* Biometrika, 51: 429–438.
- Crawford RJM, Altwegg R, Barham BJ, Barham PJ, Durant JM, Dyer BM, Geldenhuys D, Makhado AB, Pichegru L, Ryan PG, Underhill LG, Upfold L, Visagie J, Waller LJ, and Whittington PA.** 2011a. *Collapse of South Africa’s penguins in the early 21st century.* African Journal of Marine Science, 33: 139–156.
- Crawford RJM and Boonstra H.** 1994. *Counts of moulting and breeding jackass penguins *Spheniscus demersus*: A comparison at Robben Island, 1988–1993.* Marine Ornithology, 22: 213–219.
- Crawford RJM, Davis SA, Harding RT, Jackson LF, Leshoro TM, Meÿer MA, Randall RM, Underhill LG, Upfold L, van Dalsen AP, van der Merwe E, Whittington**

- PA, Williams AJ, and Wolfaardt AC.** 2000a. *Initial impact of the Treasure oil spill on seabirds off western South Africa.* South African Journal of Marine Science, 22: 157–176.
- Crawford RJM and Dyer BM.** 1995. *Responses by four seabird species to a fluctuating availability of Cape Anchovy *Engraulis capensis* off South Africa.* Ibis, 137: 329–339.
- Crawford RJM, Hemming M, Kemper J, Klages NTW, Randall RM, Underhill LG, Venter AD, Ward VL, and Wolfaardt AC.** 2006. *Molt of the African penguin, *Spheniscus demersus*, in relation to its breeding season and food availability.* Acta Zoologica Sinica, 52: 444–447.
- Crawford RJM, Shannon LJ, and Whittington PA.** 1999. *Population dynamics of the African penguin *Spheniscus demersus* at Robben Island, South Africa.* Marine Ornithology, 27: 139–147.
- Crawford RJM, Shannon LJ, Whittington PA, and Murison G.** 2000b. *Factors influencing growth of the African penguin colony at Boulders, South Africa, 1985–1999.* South African Journal Of Marine Science, 22: 111–119.
- Crawford RJM, Waller LJ, Upfold L, Pichegru L, Makhado AB, Geldenhuys D, and Dyer BM.** 2011b. *Numbers of African Penguins breeding in South Africa in 2011.* Technical report, Document No. Fisheries/2011/SWG-PEL/Island Closure Task Team/01. Cape Town: DAFF, Branch: Fisheries Management.
- Crawford RJM, Williams AJ, Hofmeyr JH, Klages NTW, Randall RM, Cooper J, Dyer BM, and Chesselet YJ.** 1995. *Trends of African penguin *Spheniscus demersus* populations in the 20th century.* South African Journal Of Marine Science, 16: 101–118.
- de Moor CL, Butterworth DS, and Coetzee JC.** 2008. *Revised estimates of abundance of South African sardine and anchovy from acoustic surveys adjusting for echosounder saturation in earlier surveys and attenuation effects for sardine.* African Journal of Marine Science, 30: 219–232.
- Fournier DA, Skaug HJ, Ancheta J, Ianelli JN, Magnusson A, Maunder MN, Nielsen A, and Sibert J.** 2011. *AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models.* Optimization Methods and Software, 1–17.
- Hampton I.** 1992. *The role of acoustic surveys in the assessment of pelagic fish resources on the South African continental shelf.* South African Journal of Marine Science, 12: 1031–1050.
- Jackson S and Wilson RP.** 2002. *The potential costs of flipper-bands to penguins.* Functional Ecology, 16: 141–148.

- Jolly GM.** 1965. *Explicit estimates from capture-recapture data with both death and immigration – stochastic model.* Biometrika, 52: 225–247.
- Kemper J and Roux JP.** 2005. *Of squeezers and skippers: factors determining the age at moult of immature African Penguins *Spheniscus demersus* in Namibia.* Ibis, 147: 346–352.
- Lebreton JD, Burnham KP, Clobert J, and Anderson DR.** 1992. *Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies.* Ecological Monographs, 62: 67–118.
- Moseley C, Wanless RM, Altwegg R, Crawford RJM, and Badenhorst A.** 2011. *Hypotheses for relating anchovy and sardine abundance with African Penguin population processes.* Technical report, Document No. Fisheries/2011/SWG-PEL/03. Cape Town: DAFF, Branch: Fisheries Management.
- Petersen SL, Branch GM, Ainley DG, Boersma PD, Cooper J, and Woehler EJ.** 2005. *Is flipper banding of penguins a problem?* Marine Ornithology, 33: 75–79.
- Pichegru L, Grémillet D, Crawford RJM, and Ryan PG.** 2010. *Marine no-take zone rapidly benefits endangered penguin.* Biology letters, 6: 498–501.
- Pradel R, Hines JE, Lebreton JD, and Nichols JD.** 1997. *Capture-Recapture Survival Models Taking Account of Transients.* Biometrics, 53: 60–72.
- Randall RM.** 1983. *Biology of the Jackass Penguin *Spheniscus demersus* (L.) at St Croix Island, South Africa.* PhD thesis, University of Port Elizabeth.
- Randall RM, Randall BM, Cooper J, and Frost PGH.** 1986. *A new census method for penguins tested on jackass penguins *Spheniscus demersus*.* Ostrich, 57: 211–215.
- Randall RM, Randall BM, Cooper J, La Cock GD, and Ross GJ.** 1987. *Jackass Penguin *Spheniscus demersus* movements, inter-island visits, and settlement.* Journal of Field Ornithology, 58: 445–455.
- Seber GAF.** 1965. *A note on the multiple-recapture census.* Biometrika, 52: 249–259.
- Shabangu F, Coetzee JC, Merkle D, and Mushanganyisi K.** 2011. *Results of the 2011 spawner biomass survey.* Technical report, Document No. Fisheries/2011/SWG-PEL/92. Cape Town: DAFF, Branch: Fisheries Management.
- Sherley RB, Burghardt T, Barham PJ, Campbell N, and Cuthill IC.** 2010. *Spotting the difference: towards fully-automated population monitoring of African penguins *Spheniscus demersus*.* Endangered Species Research, 11: 101–111.
- Skaug HJ and Fournier DA.** 2006. *Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models.* Computational Statistics & Data Analysis, 51: 699–709.

- Underhill LG, Bartlett PA, Baumann L, Crawford RJM, Dyer BM, Gildenhuyst AM, Nel DC, Oatley TB, Thornton M, Upfold L, Williams AJ, Whittington PA, and Wolfaardt AC.** 1999. *Mortality and survival of African Penguins *Spheniscus demersus* involved in the Apollo Sea oil spill: an evaluation of rehabilitation efforts.* Ibis, 141: 29–37.
- Underhill LG and Crawford RJM.** 1999. *Season of moult of African penguins at Robben Island, South Africa, and its variation, 1988–1998.* South African Journal of Marine Science, 21: 437–441.
- Underhill LG, Crawford RJM, Wolfaardt AC, Whittington PA, Dyer BM, Leshoro TM, Ruthenberg M, and Upfold L.** 2006. *Regionally coherent trends in colonies of African penguins *Spheniscus demersus* in the Western Cape, South Africa, 1987–2005.* African Journal of Marine Science, 28: 697–704.
- Waller LJ.** 2011. *The African Penguin *Spheniscus demersus*: Conservation and Management Issues.* PhD thesis, University of Cape Town.
- White GC and Burnham KP.** 1999. *Program MARK: survival estimation from populations of marked animals.* Bird Study, 46: 120–139.
- Whittington PA, Klages NTW, Crawford RJM, Wolfaardt AC, and Kemper J.** 2005a. *Age at first breeding of the African Penguin.* Ostrich, 76: 14–20.
- Whittington PA, Randall RM, Crawford RJM, Wolfaardt AC, Klages NTW, Randall BM, Bartlett PA, Chesselet YJ, and Jones R.** 2005b. *Patterns of immigration to and emigration from breeding colonies by African penguins.* African Journal of Marine Science, 27: 205–213.
- Wolfaardt AC, Underhill LG, Altwegg R, and Visagie J.** 2008. *Restoration of oiled African penguins *Spheniscus demersus* a decade after the Apollo Sea spill.* African Journal of Marine Science, 30: 421–436.
- Wolfaardt AC, Underhill LG, and Crawford RJM.** 2009a. *Comparison of moult phenology of African penguins *Spheniscus demersus* at Robben and Dassen islands.* African Journal of Marine Science, 31: 19–29.
- Wolfaardt AC, Williams AJ, Underhill LG, Crawford RJM, and Whittington PA.** 2009b. *Review of the rescue, rehabilitation and restoration of oiled seabirds in South Africa, especially African penguins *Spheniscus demersus* and Cape gannets *Morus capensis*, 1983–2005.* African Journal of Marine Science, 31: 31–54.

Table 1: Total penguin nest counts at Robben Island since the island was recolonized in 1983. The maximum nest count for each year is indicated. Subsets of this series were published by Crawford *et al.* (1995) for the years 1983–1994 and Underhill *et al.* (2006) for the years 1987–2005. Note, however, that this table includes subsequent revisions for 2000–2004. These revisions, as well as counts for 2006–2008 were supplied by RJM Crawford (pers. comm.). Nest counts for the most recent years (2009–2011) were reported by Crawford *et al.* (2011b).

Year	Number of breeding pairs
1983	9
1984	24
1985	103
1986	227
1987	476
1988	849
1989	829
1990	1 278
1991	1 879
1992	2 027
1993	2 176
1994	2 799
1995	2 279
1996	3 097
1997	3 336
1998	3 467
1999	4 399
2000	5 407
2001	6 427
2002	7 099
2003	5 968
2004	7 798
2005	7 152
2006	3 697
2007	5 935
2008	2 234
2009	2 415
2010	2 628
2011	1 817

Table 2: Numbers of penguins moulting at Robben Island as estimated using the double Gaussian approach described in Section 2.1. The column for the number of adults is labelled $2N_y^{\text{obs}}$ because these estimates must be halved before being input to the population model to represent the female component only.

Season	Counts	Adults		Juveniles		Proportion
	n_y	$2N_y^{\text{obs}}$	$CV(\%)$	J_y^{obs}	$CV(\%)$	J_y^{prop}
1988/1989	18	3 466	7.9	800	10.7	0.188
1989/1990	25	3 397	7.4	851	10.7	0.200
1990/1991	24	4 724	6.1	915	10.1	0.162
1991/1992	26	4 939	6.0	1 603	7.5	0.245
1992/1993	26	6 544	5.3	1 631	7.6	0.199
1993/1994	24	7 915	4.8	1 574	7.6	0.166
1994/1995	25	7 932	4.8	1 368	8.3	0.147
1995/1996	25	6 674	5.4	1 406	8.3	0.174
1996/1997	23	7 300	8.0	2 292	9.5	0.239
1997/1998	25	8 556	4.4	2 381	6.0	0.218
1998/1999	20	9 125	4.3	2 738	5.9	0.231
1999/2000	21	11 694	4.0	2 811	5.8	0.194
2000/2001	15	12 883	3.8	2 417	8.5	0.158
2001/2002	9	12 361	8.6	2 712	8.5	0.180
2002/2003	16	14 521	3.9	3 430	6.4	0.191
2003/2004	14	16 975	3.6	3 326	5.8	0.164
2004/2005	15	12 442	4.0	2 508	5.5	0.168
2005/2006	20	7 660	4.9	2 614	6.0	0.254
2006/2007	17	6 453	6.0	1 889	8.9	0.226
2007/2008	19	5 157	6.4	1 393	10.7	0.213
2008/2009	26	3 745	7.2	1 799	7.2	0.325
2009/2010	26	4 209	19.1	1 730	8.1	0.291
2010/2011	20	3 000	9.9	1 199	13.9	0.286

Table 3: Numbers of penguins moulting at Robben Island as estimated by the linear interpolation approach published by Wolfaardt *et al.* (2009a) and by the double Gaussian approach described in Section 2.1.2.

Season	Adults			Juveniles		
	Linear interpolation	Double Gaussian	Difference (%)	Linear interpolation	Double Gaussian	Difference (%)
1988/1989	3 459	3 466	0.2	842	800	-4.9
1989/1990	3 392	3 397	0.1	866	851	-1.7
1990/1991	4 730	4 724	-0.1	911	915	0.4
1991/1992	4 915	4 939	0.5	1 598	1 603	0.3
1992/1993	6 538	6 544	0.1	1 597	1 631	2.1
1993/1994	8 002	7 915	-1.1	1 585	1 574	-0.7
1994/1995	7 948	7 932	-0.2	1 373	1 368	-0.3
1995/1996	6 563	6 674	1.7	1 403	1 406	0.2
1996/1997	5 608	7 300	30.2	2 138	2 292	7.2
1997/1998	8 696	8 556	-1.6	2 351	2 381	1.3
1998/1999	9 397	9 125	-2.9	2 834	2 738	-3.4
1999/2000	11 765	11 694	-0.6	2 803	2 811	0.3
2000/2001	13 362	12 883	-3.6	2 565	2 417	-5.8
2001/2002	16 439	12 361	-24.8	3 921	2 712	-30.8
2002/2003	14 737	14 521	-1.5	3 330	3 430	3.0
2003/2004	17 424	16 975	-2.6	3 440	3 326	-3.3
2004/2005	12 871	12 442	-3.3	2 617	2 508	-4.2
2005/2006	7 769	7 660	-1.4	2 654	2 614	-1.5

Table 4: Estimates of the parameters in equation (1) which best fit the counts of moulting juvenile penguins at Robben Island.

Season	N_1	μ_1	σ_1	N_2	μ_2	σ_2
1988/1989	6 654	1 Dec	11.4	3 516	23 Feb	42.1
1989/1990	2 986	3 Dec	6.0	7 859	16 Jan	62.7
1990/1991	5 322	4 Dec	14.9	6 419	17 Feb	64.6
1991/1992	10 597	10 Dec	14.1	9 846	6 Jan	68.6
1992/1993	8 746	7 Dec	7.6	11 986	24 Dec	59.0
1993/1994	12 337	3 Dec	13.4	7 690	28 Jan	58.9
1994/1995	6 721	27 Dec	14.0	10 681	16 Jan	57.6
1995/1996	8 924	28 Nov	31.5	8 934	2 Mar	32.0
1996/1997	23 105	19 Dec	22.9	6 013	24 Mar	31.7
1997/1998	17 602	24 Nov	25.3	12 716	20 Jan	63.5
1998/1999	18 893	28 Nov	26.2	15 954	21 Jan	60.2
1999/2000	24 837	30 Nov	20.8	10 857	3 Mar	26.9
2000/2001	22 681	15 Nov	27.6	8 017	16 Mar	33.8
2001/2002	21 819	26 Nov	12.3	12 701	26 Feb	49.5
2002/2003	24 436	4 Dec	12.9	19 999	11 Jan	89.4
2003/2004	29 857	25 Nov	26.8	12 386	1 Mar	17.8
2004/2005	20 654	2 Dec	17.5	11 256	5 Jan	64.5
2005/2006	25 724	29 Nov	19.6	7 644	14 Jan	77.8
2006/2007	16 736	19 Nov	18.9	8 033	22 Nov	83.8
2007/2008	6 331	5 Dec	13.6	11 547	20 Dec	75.4
2008/2009	7 275	30 Nov	17.7	15 586	14 Jan	52.7
2009/2010	8 475	1 Dec	12.4	13 536	6 Jan	60.8
2010/2011	12 289	4 Nov	36.2	2 941	7 Mar	30.4

Table 5: Estimates of the parameters in equation (1) which best fit the counts of moulting adult penguins at Robben Island.

Season	N_1	μ_1	σ_1	N_2	μ_2	σ_2
1988/1989	39 248	13 Dec	19.3	4 828	19 Jan	66.2
1989/1990	33 982	13 Dec	16.3	9 164	7 Jan	54.7
1990/1991	48 582	12 Dec	19.1	11 425	11 Jan	56.0
1991/1992	46 954	12 Dec	20.4	15 806	11 Jan	57.8
1992/1993	71 918	10 Dec	17.0	11 210	4 Jan	56.8
1993/1994	91 001	12 Dec	21.9	9 547	16 Jan	58.2
1994/1995	32 724	12 Nov	13.6	68 015	23 Dec	41.0
1995/1996	64 385	7 Dec	19.5	20 401	11 Jan	56.9
1996/1997	72 095	17 Dec	22.9	20 668	20 Jan	57.3
1997/1998	83 273	5 Dec	22.4	25 436	8 Jan	58.1
1998/1999	98 177	14 Dec	25.6	17 742	14 Jan	57.1
1999/2000	117 280	6 Dec	18.6	31 249	12 Jan	51.3
2000/2001	113 540	22 Nov	17.1	50 074	22 Dec	42.1
2001/2002	101 810	27 Nov	19.6	55 507	6 Jan	65.8
2002/2003	169 441	4 Dec	26.2	14 992	19 Mar	34.5
2003/2004	111 400	30 Nov	18.2	104 350	11 Dec	54.9
2004/2005	116 200	30 Nov	20.2	41 817	1 Jan	48.2
2005/2006	80 941	1 Dec	28.9	16 646	29 Dec	77.1
2006/2007	66 480	29 Nov	31.1	15 793	15 Dec	77.1
2007/2008	52 628	4 Dec	22.5	13 021	2 Jan	72.3
2008/2009	39 798	11 Dec	28.7	7 821	6 Jan	67.9
2009/2010	10 484	11 Dec	7.7	42 988	16 Dec	49.7
2010/2011	22 367	7 Dec	24.1	15 759	19 Dec	55.9

Table 6: Model inputs related to oiling.

Quantity	Description	Value
N_{1994}^{oil}	Number of adult females which died as a result of the <i>Apollo Sea</i> oil spill	600
N_{2000}^{oil}	Number of adult females which died as a result of the <i>Treasure</i> oil spill	500
p_{1994}^{oil}	Proportion of chicks which died due to the <i>Apollo Sea</i> oil spill	0.29
p_{2000}^{oil}	Proportion of chicks which died due to the <i>Treasure</i> oil spill	0.40
m_{1994}	Time of year of the <i>Apollo Sea</i> oil spill (months)	5.7
m_{2000}	Time of year of the <i>Treasure</i> oil spill (months)	5.8

Table 7: Sardine November survey 1+ biomass (in tonnes) observed in the whole assessment survey area, the biomass observed west of Cape Agulhas (strata A–C), the proportion of the assessment total which was observed in strata A–C, and the average inter-transect sampling *CV*.

Year	Total (t)	Strata A–C (t)	Proportion	<i>CV</i>
1984	48 378	48 009	0.992	0.972
1985	45 013	25 457	0.566	0.449
1986	299 797	238 230	0.795	0.696
1987	111 285	94 165	0.846	0.402
1988	134 362	128 043	0.953	0.715
1989	256 655	198 328	0.773	0.225
1990	289 876	248 855	0.858	0.276
1991	597 858	517 180	0.865	0.308
1992	494 157	247 756	0.501	0.488
1993	560 019	480 822	0.859	0.316
1994	518 354	389 730	0.752	0.280
1995	844 727	348 832	0.413	0.515
1996	529 456	257 763	0.487	0.359
1997	1 224 632	964 835	0.788	0.244
1998	1 607 328	1 082 547	0.674	0.251
1999	1 635 410	708 029	0.433	0.212
2000	2 292 380	726 230	0.317	0.500
2001	2 309 600	669 617	0.290	0.142
2002	4 206 250	1 184 713	0.282	0.227
2003	3 564 171	1 343 118	0.377	0.197
2004	2 615 715	292 522	0.112	0.334
2005	1 048 991	75 604	0.072	0.300
2006	712 553	177 885	0.250	0.346
2007	256 727	57 666	0.225	0.345
2008	384 080	211 871	0.552	0.422
2009	502 254	262 853	0.523	0.271
2010	508 392	309 465	0.609	0.235

Table 8: Parameter values fixed on input to the model.

Parameter	Description	Value
A	Plus group age	5
p_J	Detectability of juveniles relative to adults in the moult count	1.0
a^*	Age of first breeding attempt	4
σ_J	Standard errors of the logarithms of the juvenile proportions	0.1
σ_H	Standard deviation of reproductive success	0.1
H_{\max}	Maximum allowed reproductive success	1.8
q_M	Proportion of moulters susceptible to observation	0.9
M_{\min}	Minimum allowed mortality rate	0.04
d^*	Parameter of logistic prior on δ in the biomass–mortality relationship	0.02
σ_δ	Parameter of logistic prior on δ in the biomass–mortality relationship	0.005
\hat{c}	Overdispersion factor for the tag-recapture likelihood	5.57

Table 9: Parameters estimated within the model; where the parameter is estimated as the log of an abundance, the corresponding abundance value follows in parenthesis.

Parameter	Description	Prior	Posterior mode	
$\ln N_0$	Log of initial population size	$U[1, 10]$	6.788	(887)
λ	Initial population profile parameter	$U[0, 3]$	0.205	
σ_{add}	Additional variance in moult counts	$U[0, 1]$	0.000	
$\tilde{\sigma}$	Parameter related to variability about the assumed biomass–mortality relationship	$U[0.001, 0.5]$	0.088	
η	Reproductive success relationship	$U[0, 1]$	0.510	
X_y	Adult mortality random effects	$U[-4.5, 4.5]$		
H_y	Reproductive success	$U[0.001, 0.999]$		
p_y	Tag re-sighting probabilities	$U[0, 1]$		
α			0.087	
β	Parameters of the biomass–mortality relationship (equation (26))	$U[0, 1]$	0.592	
γ			0.171	
δ			0.043	
$\ln I_{1989-1990}$	Log of the number of three-year-old penguins immigrating to Robben Island in the years indicated	$U[-4, 10]$	6.675	(792)
$\ln I_{1991-1993}$			6.030	(415)
$\ln I_{1994-1996}$			-3.999	(0)
$\ln I_{1997-1999}$			5.673	(290)
	Transient-related apparent mortality rates			
t_1	1989, 1993, 1995–1999, 2004–2009		0.109	
t_2	1990		1.262	
t_3	1991		0.634	
t_4	1992	$U[0, 5]$	1.959	
t_5	1994		0.000	
t_6	2000		0.000	
t_7	2001–2003		0.684	

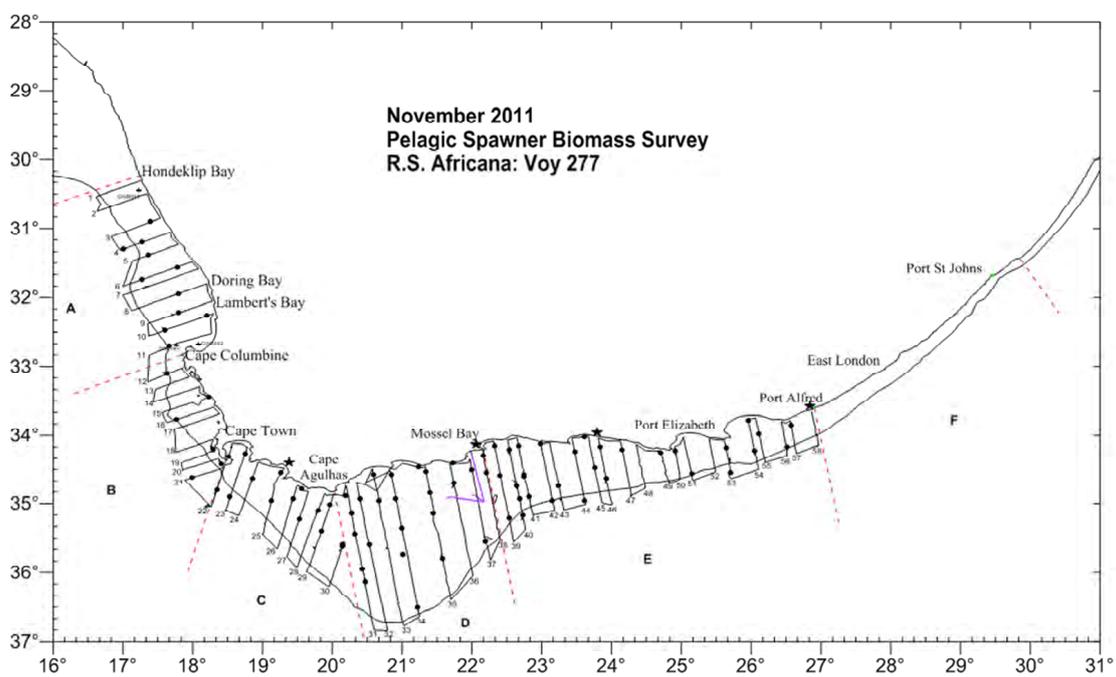


Figure 1: Cruise track of the November 2011 pelagic biomass survey (Shabangu *et al.* 2011).

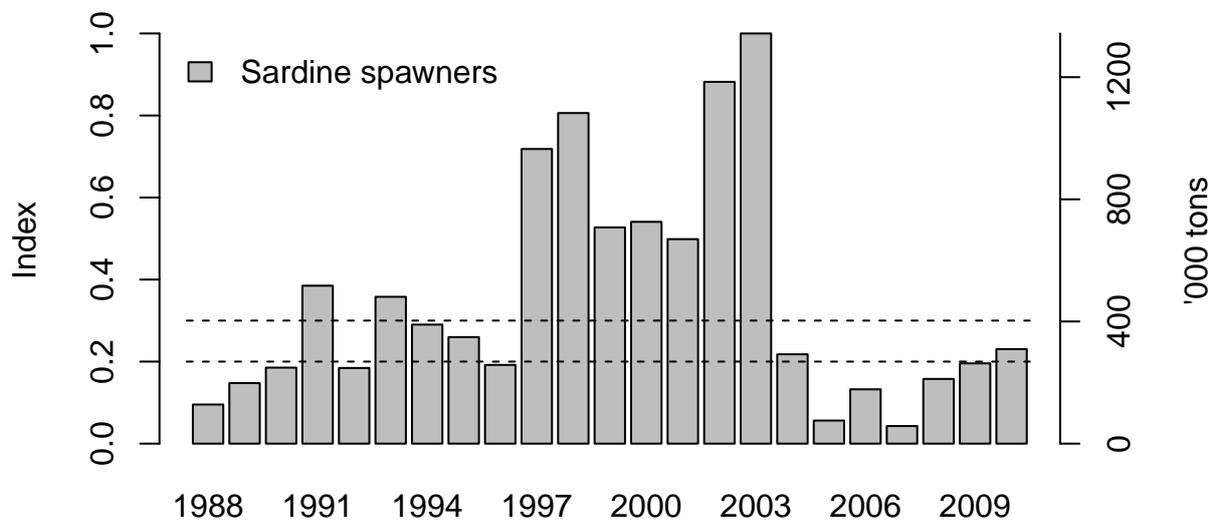


Figure 2: The index for the November survey sardine spawner biomass west of Cape Agulhas (strata A–C) is expressed relative to the maximum (observed in 2003) on the left hand axis. Dashed horizontal lines at index values 0.2 and 0.3 indicate the range of biomass levels used for the penguin population projections shown subsequently. The right-hand axis shows the corresponding biomass in thousands of tonnes.

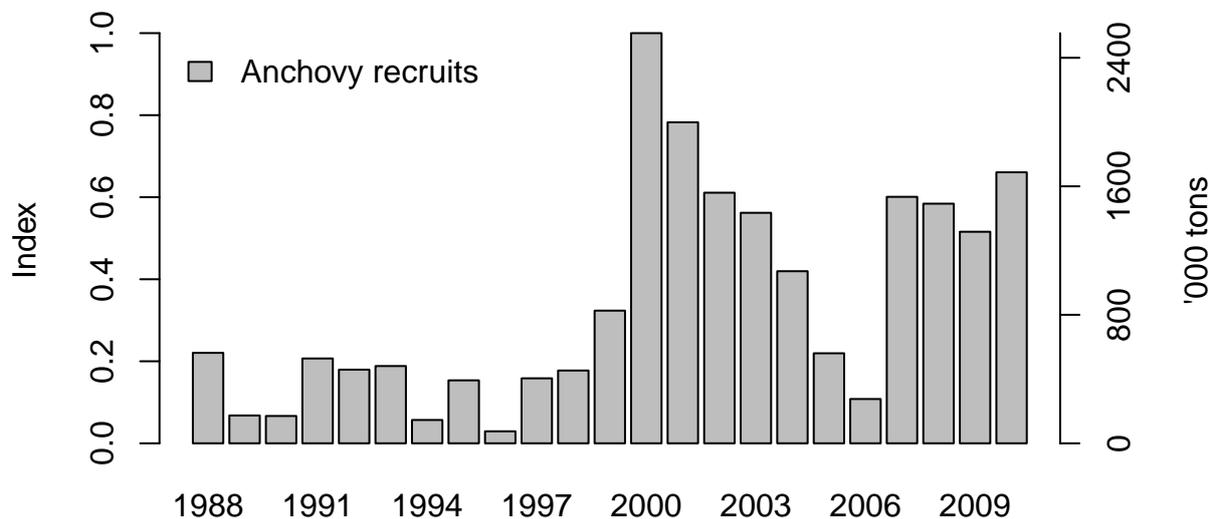


Figure 3: The index for the May survey anchovy recruit biomass west of Cape Infanta is expressed relative to the maximum (observed in 2000). The right-hand axis shows the corresponding biomass in thousands of tonnes.

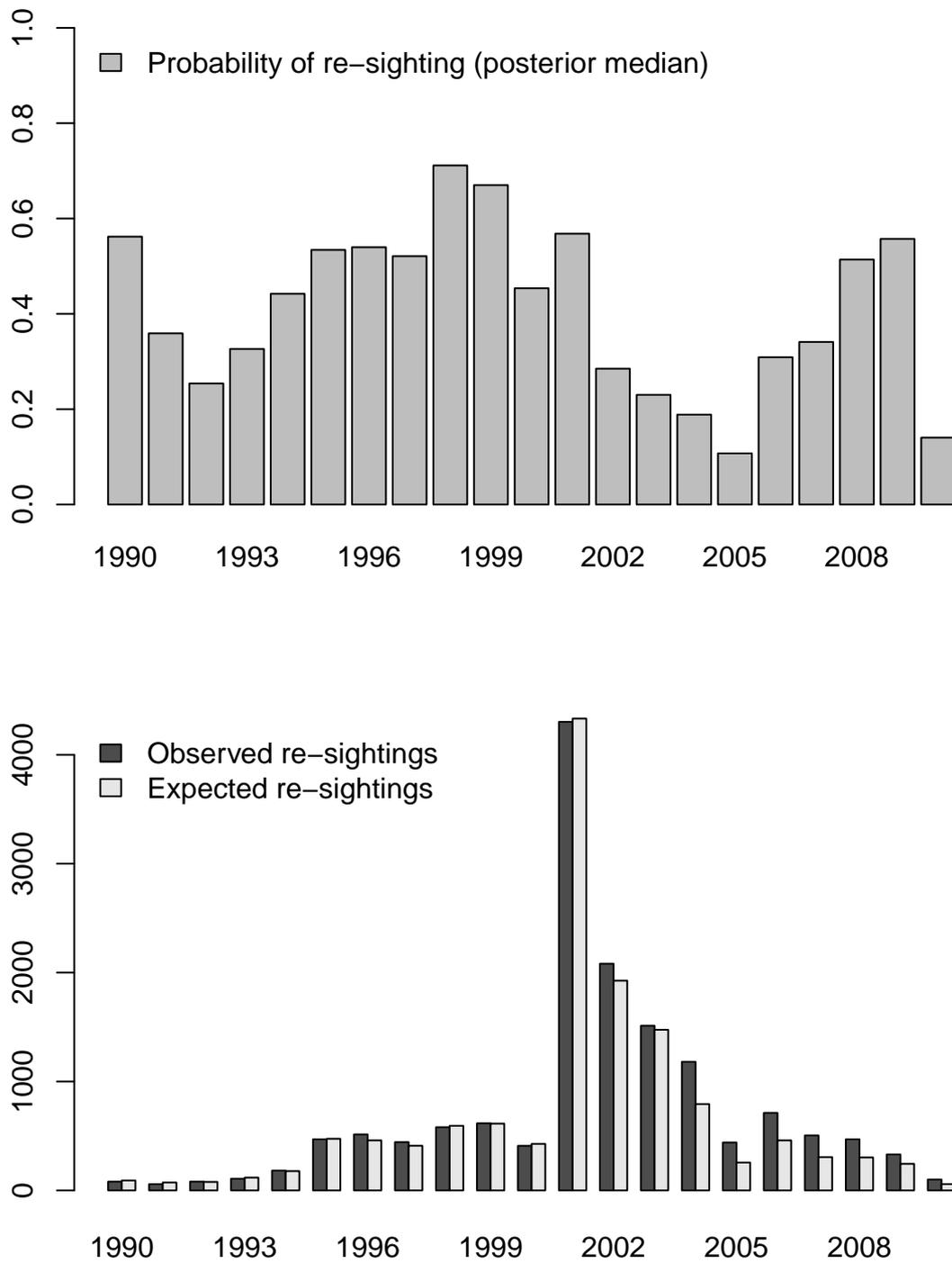


Figure 4: Results of fitting to tag-recapture data. Top: annual probability of re-sighting banded birds. Bottom: comparison of observed and expected numbers of banded penguins re-sighted each year for the joint posterior mode.

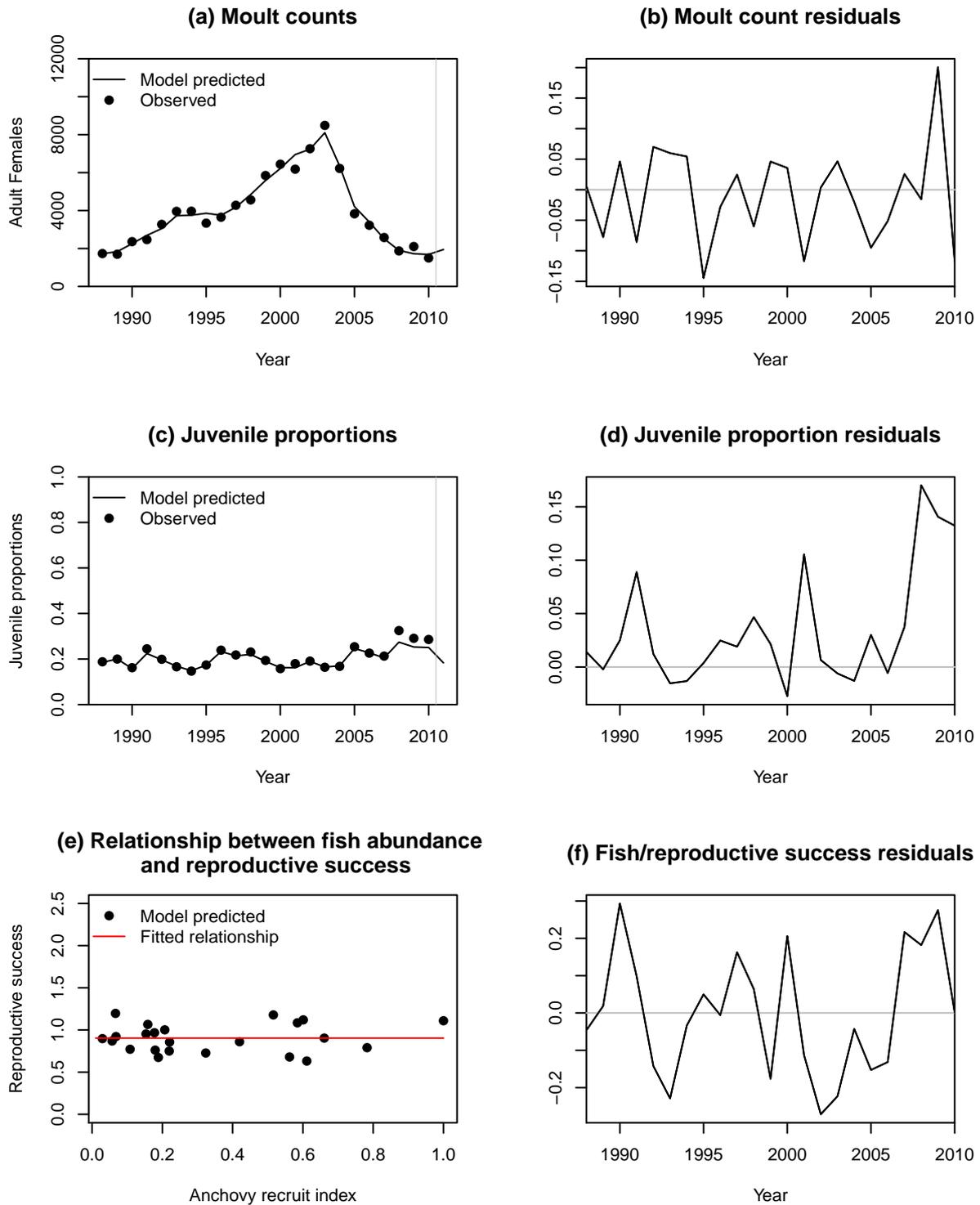


Figure 5: Fits to data/relationships and associated residuals at the joint posterior mode for the base case. The residuals in (b) and (d) are the differences between the logarithms of the observations and the model predicted values. The residuals in (f) are the differences between the estimated reproductive success rates and the assumed relationship.

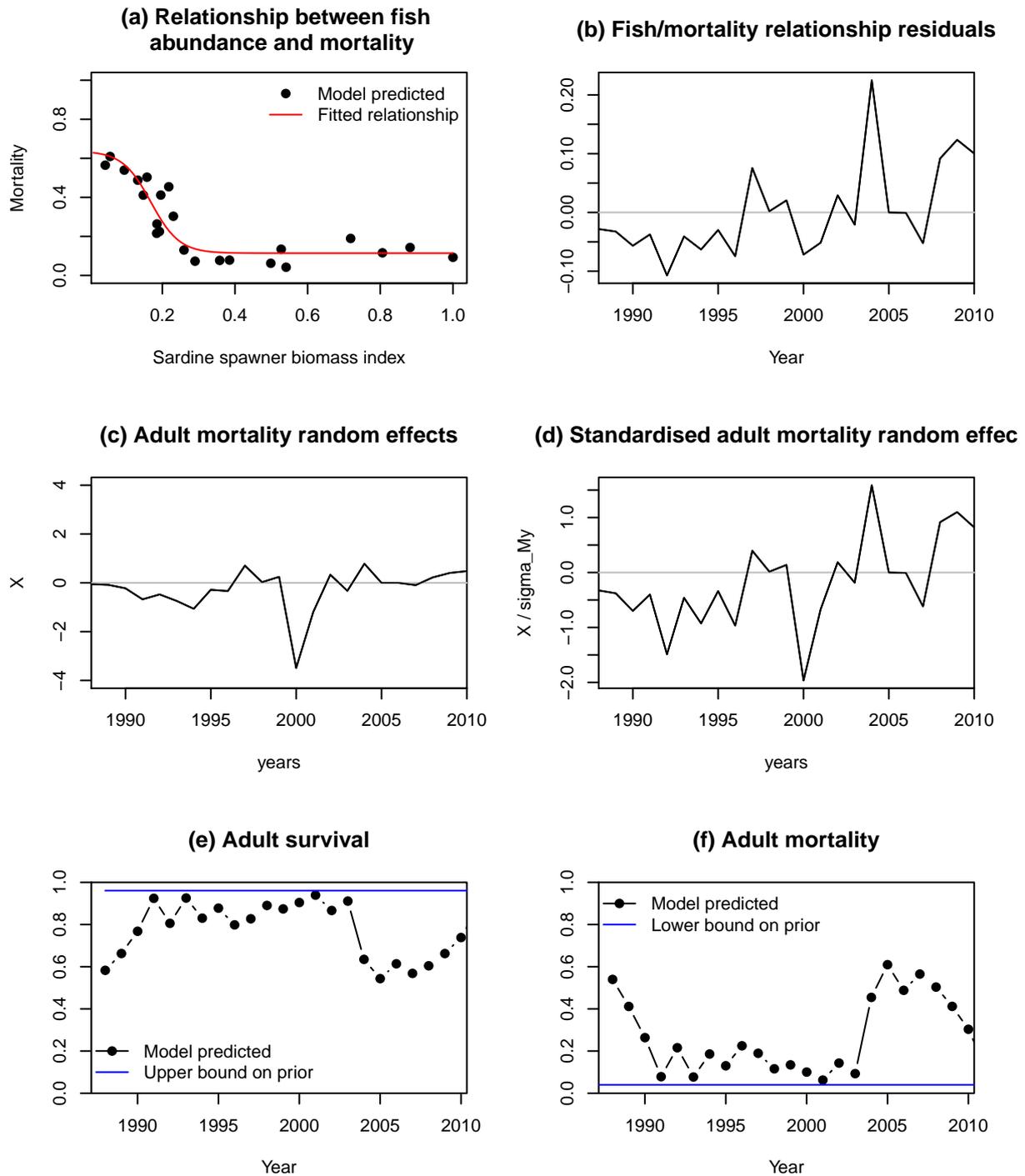


Figure 6: Further fits to data/relationships and associated residuals for the joint posterior mode for the base case. Residuals in (b) are the differences between the estimated annual mortality rates and those predicted by the relationship with fish abundance. The random effects X_y are in (c), and are standardised in (d) by dividing by σ_y . Plot (e) shows the time series of the estimates for the adult survival rates, and (f) shows the corresponding mortality rates.

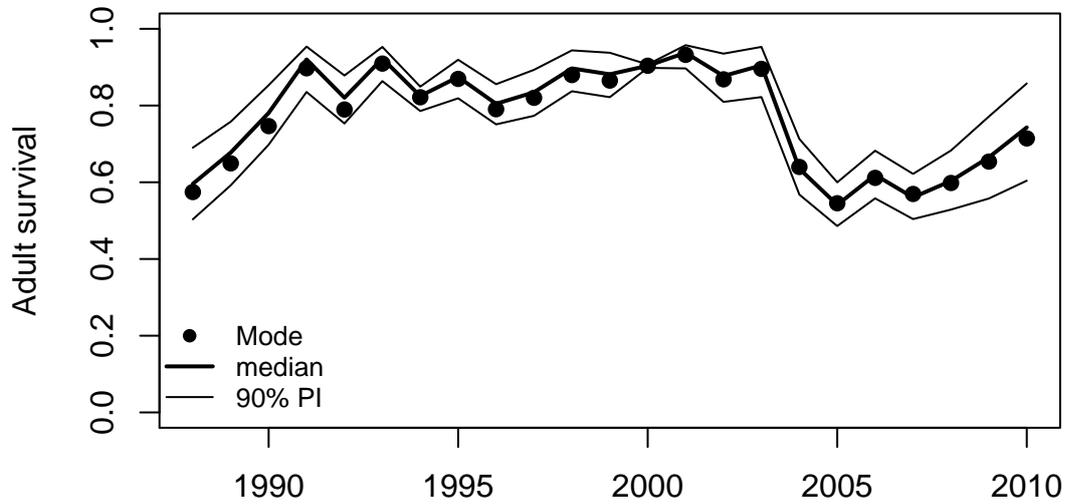


Figure 7: Time-series of estimated adult penguin annual survival rates at the joint posterior mode (black circles). The median and 90% probability interval of the Bayesian posterior distribution are indicated with lines.

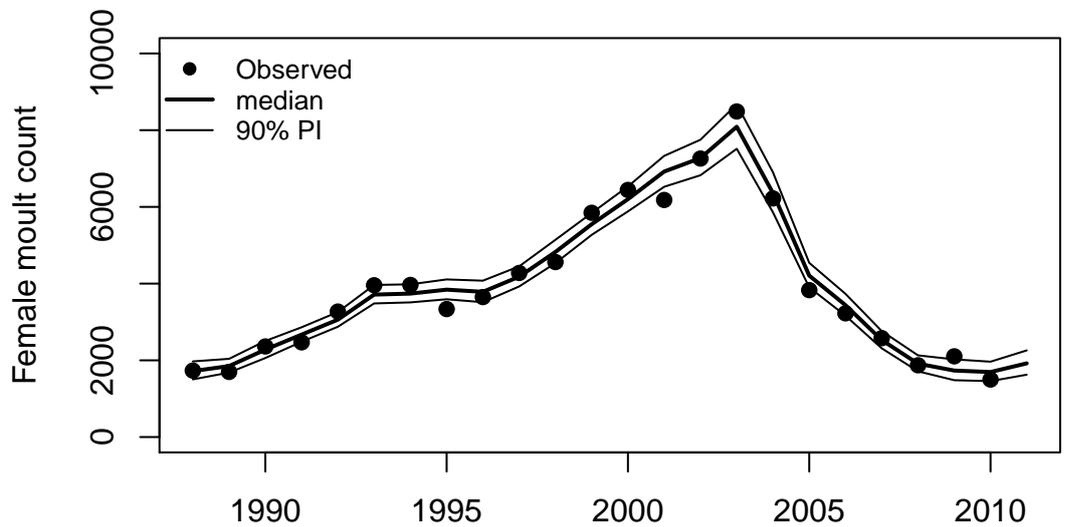


Figure 8: Time-series of female moulting penguins from observations (black circles). The median and 90% probability interval of the Bayesian posterior distribution of the model-predicted moult counts are indicated with lines.

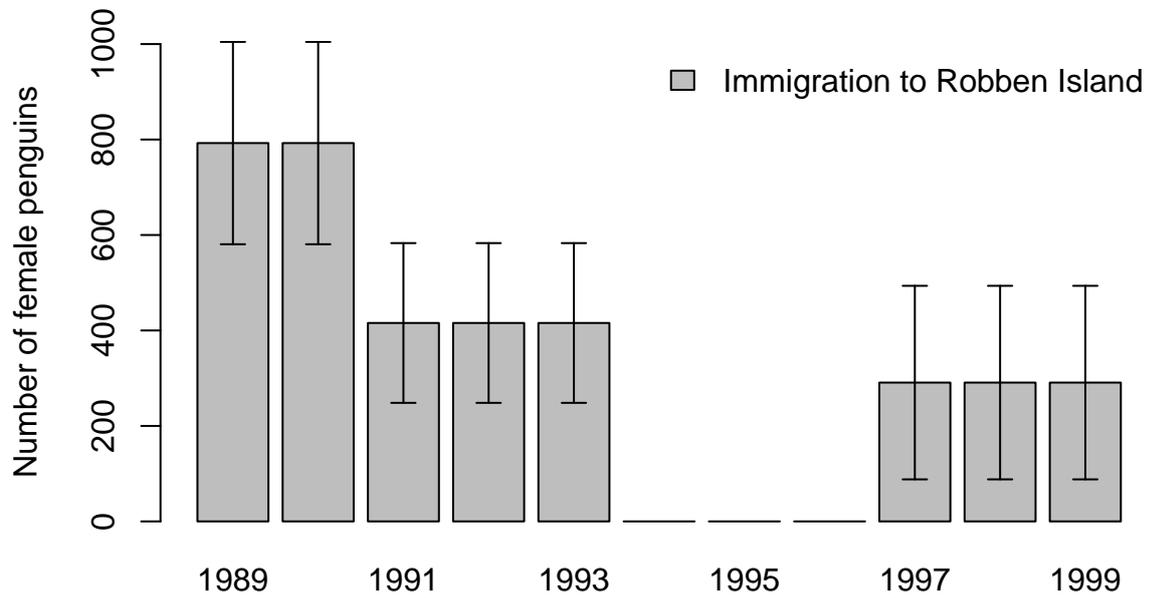


Figure 9: Number of three-year-old female penguins estimated to immigrate to Robben Island each year, with associated standard errors.

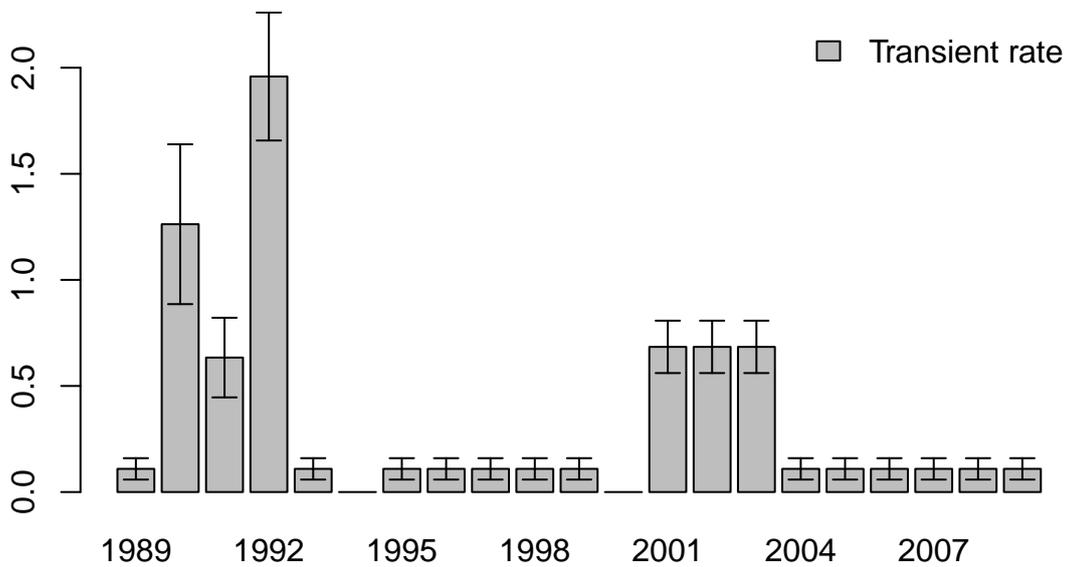


Figure 10: Estimated transient-related apparent mortality rates, with associated standard errors.

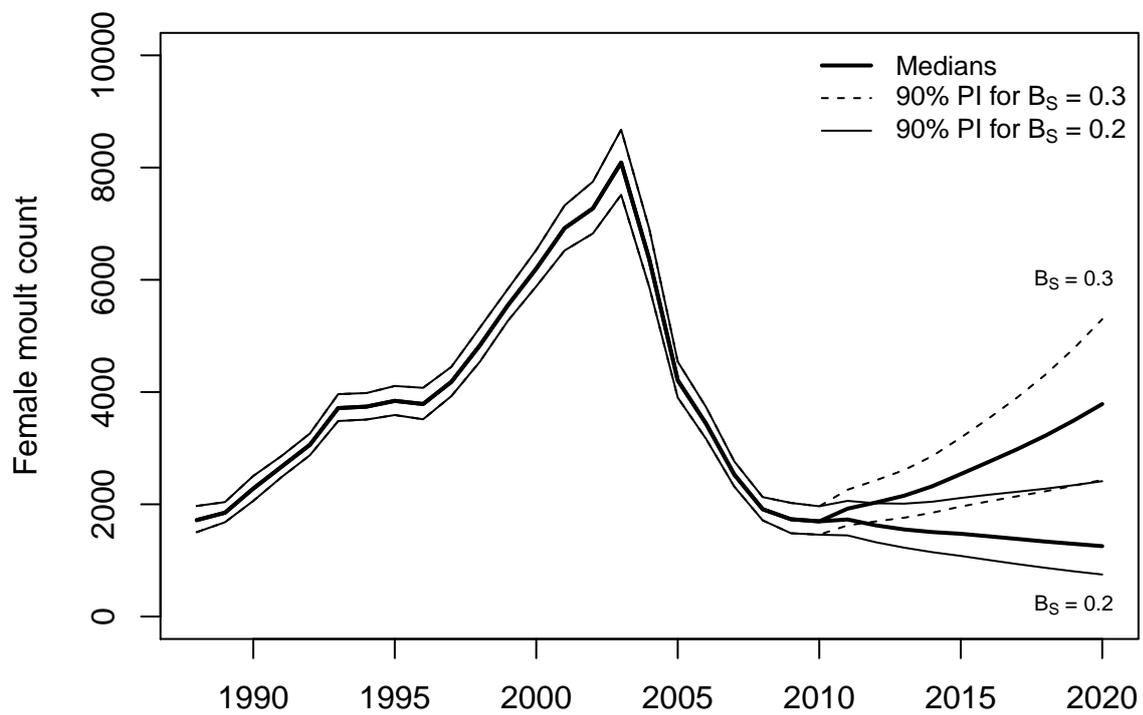


Figure 11: Time-series of model-predicted numbers of female moulting penguins with projections from 2010 from the joint posterior mode for the base case model. Two future sardine spawner biomass levels west of Cape Agulhas are assumed: $B_S = 0.2$ and $B_S = 0.3$.