Illustrative projections of Robben Island penguin numbers

William Robinson[1] and Doug Butterworth

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

November 2011

Introduction

A primary goal of the Robben Island penguin model is to link it with the pelagic OMP in order to give projected penguin numbers for various candidate management procedures. Given that neither OMP–13 nor updated population models for sardine and anchovy are finalised as yet, in this paper illustrative graphs have been produced using output for sardine and anchovy abundance from the previous version of the management procedure (OMP–08).

Methods

A set of 1000 plausible trajectories for observed November sardine survey biomass were generated under OMP–08 (Figure 1). Linking to this OMP output, the Robben Island penguin model (Robinson and Butterworth, 2011) is used to project penguin abundances forward to 2028. As the model relates penguin annual survival to sardine abundance in pelagic survey strata A–C, each OMP-generated biomass value is scaled by multiplying by an historically observed proportion of sardine in the area concerned (Figure 2). Two scenarios are considered regarding the future distribution of sardine:

1. in future the sardine distribution is similar to that seen from 1984 to 1999, and
2. future sardine distributions continue along the lines seen from 2000 to 2010.

We remind that the rates of annual adult survival S_y and annual natural mortality M_y are related as:

$$S_y = e^{-M_y} \tag{1}$$

where M_y is modelled as follows:

$$M_y = M_{\min} + f_S\left(B_{S,y}\right)e^{X_y} \tag{2}$$

and X_y is distributed $N\left(0,\sigma_y^2\right)$ with

$$\sigma_y = \sqrt{e^{\left[\tilde{\sigma}/f_S\left(B_{s,y}\right)\right]^2} - 1} \tag{3}$$

[1]william.robinson@uct.ac.za

Thus, when projecting, values for X_y are generated by drawing normally distributed random numbers with variance σ_y^2. Results indicate that the parameter $\tilde{\sigma}$ has a strong influence (see the Appendix). In order to determine a suitable range of $\tilde{\sigma}$ values to use, plots of the various components of the likelihood function were examined (Figure 3). It is evident that there is a trade-off between reasonable fits to the moult counts and tag data on one hand and the value of the penalty function ("log prior") for the residuals of the mortality–biomass relationship on the other.

Under a Bayesian approach, with integration over a prior for $\tilde{\sigma}$ as will occur for final computations under this model, results would be dominated by very low $\tilde{\sigma}$ values given the rapidity with which the total negative log likelihood decreases with $\tilde{\sigma}$ (see Figure 3(a)). However, this takes one into territory where fits to the moult counts and particularly the tag-recapture data are likely mis-specified, assuming that the plots in Figure 3(b) and (d) respectively are reflections of such at low $\tilde{\sigma}$. For a choice of $\tilde{\sigma} = 0.05$, the first of these contributions to the overall likelihood hardly increases for yet higher $\tilde{\sigma}$, and further increases for the second are limited. Thus, $\tilde{\sigma} = 0.05$, or more "conservatively" 0.07, would seem a defensible lower bound for any Bayesian prior, and then the further decrease in L to $\tilde{\sigma} = 0.10$ would see contributions from yet higher values to the Bayesian integral to be negligible. Accordingly, the illustrative plots shown below, for $\tilde{\sigma}$ values of 0.05, 0.07 and 0.1, likely bound the results which would follow from a reasonable choice for a Bayesian prior for $\tilde{\sigma}$.

For each year of projections, the median of the 1000 penguin trajectories was recorded, as well as the 10th and 90th percentiles.

Results

Various penguin abundance trajectories are shown in Figure 4 to Figure 6. The three aspects which vary are:

1. value of $\tilde{\sigma}$ (0.05, 0.07 or 0.1),
2. future distribution of sardine (1984–1999 proportions or 2000–2010 proportions), and
3. fishing level (OMP–08 or no fishing).

In Figure 7, projected penguin abundances in 2028 are plotted for the various cases considered.

Discussion

It is clear that these analyses indicate that the factors with the greatest influence on penguin numbers are the future distribution of sardine and the value assumed for $\tilde{\sigma}$, i.e. the level of variability allowed about the curve relating adult penguin mortality to sardine abundance. More variability predicts lower penguin abundances. These results suggest that fishing (at least to the levels permitted under OMP–08) has a relatively minor effect.

Reference

Robinson WML and Butterworth DS. 2011. Full description of the Robben Island Penguin–Fish interaction model. Document MARAM IWS/DEC11/P/PENG/P1

Appendix

Choosing a value or range for $\tilde{\sigma}$

The term in the negative log likelihood that includes $\tilde{\sigma}$ is of the form:

$$-\ln L = \sum_y \left(\ln\tilde{\sigma} + \frac{1}{2}\frac{X_y^2}{\tilde{\sigma}^2} \right)$$

where $X_y =$ "observed" value – model predicted value.

Note that as $\tilde{\sigma}$ is varied, the contributions of the terms on the right hand side are of the form shown in Figure A1 below.

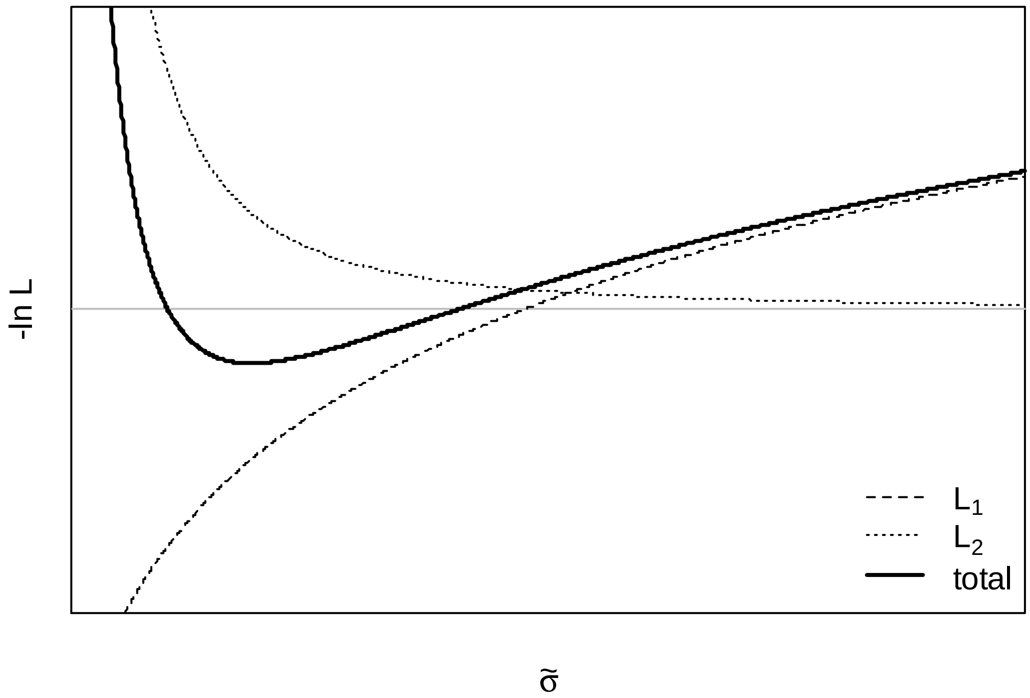


Figure A1: Comparison of terms in the negative log likelihood function as a function of $\tilde{\sigma}$. The likelihood components are $L_1 = \sum \ln\tilde{\sigma}$ and $L_2 = \sum X_y^2 / 2\tilde{\sigma}^2$.

As $\tilde{\sigma} \to 0$, the sum of the two terms $\to +\infty$, because $1/\tilde{\sigma}^2$ approaches $+\infty$ more strongly than $\ln\tilde{\sigma}$ approaches $-\infty$ in this limit. Thus the maximum likelihood criterion will always produce a positive estimate for $\tilde{\sigma}$.

However, that holds only when the "observed values" are data, which do not change. In some cases, as here and also when fitting stock-recruitment curves in population models, these are model estimates which can change as $\tilde{\sigma}$ changes. In particular, as $\tilde{\sigma}$ is decreased, the associated relationship becomes so heavily weighted that all the $X_y \to 0$, so that the minimum value of

$-\ln L$ is obtained as $\tilde{\sigma} \rightarrow 0$ irrespective of deterioration in the fits by the model to other time series of genuine data.

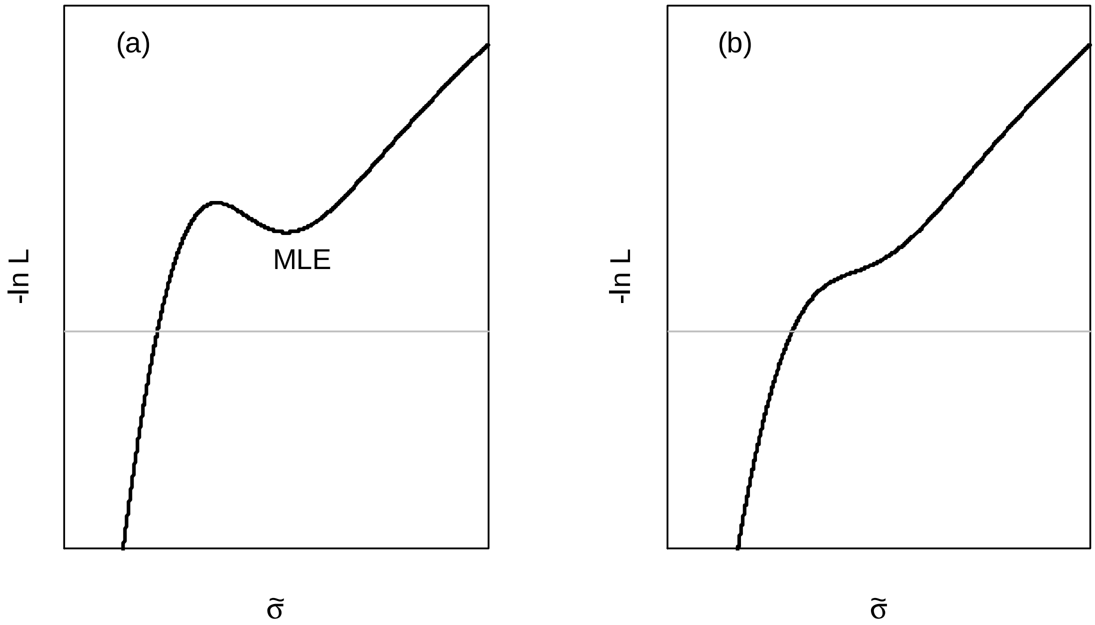


Figure A2: (a) The negative log likelihood function has a local minimum. (b) The negative log likelihood function is decreasing monotonically as $\tilde{\sigma}$ decreases.

In some cases $-\ln L$ is multimodal as in Figure A2(a), so that the value shown can be taken as the best estimate. However, in others $-\ln L$ decreases monotonically with $\tilde{\sigma}$, so that there is no clear choice.

The approach suggested therefore when cases such as Figure A2(b) occur is to choose the lowest value of $\tilde{\sigma}$ for which the model fit to other data series input does not reflect mis-specification, e.g. manifest systematic trends in residuals. This approach can also be used to set the lower bound for a prior for $\tilde{\sigma}$ for a Bayesian approach which integrates the likelihood over this prior.

Figures

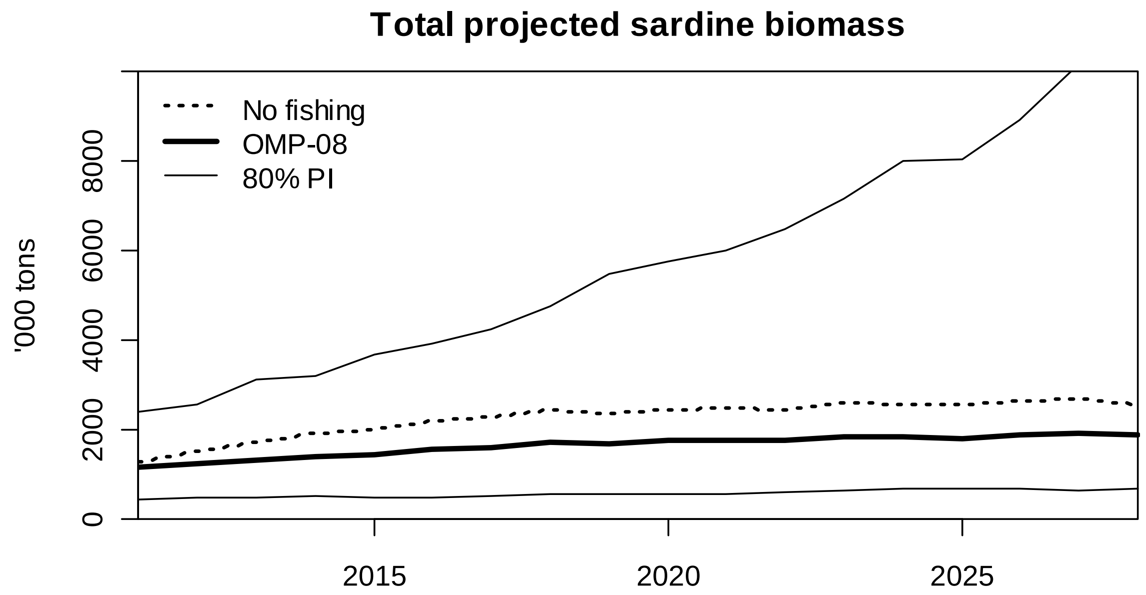


Figure 1: Median total observed sardine November survey abundance projected under OMP–08 and in the absence of fishing, used for the illustrative projections of penguin numbers. The 80% probability interval is for the projection under OMP–08.

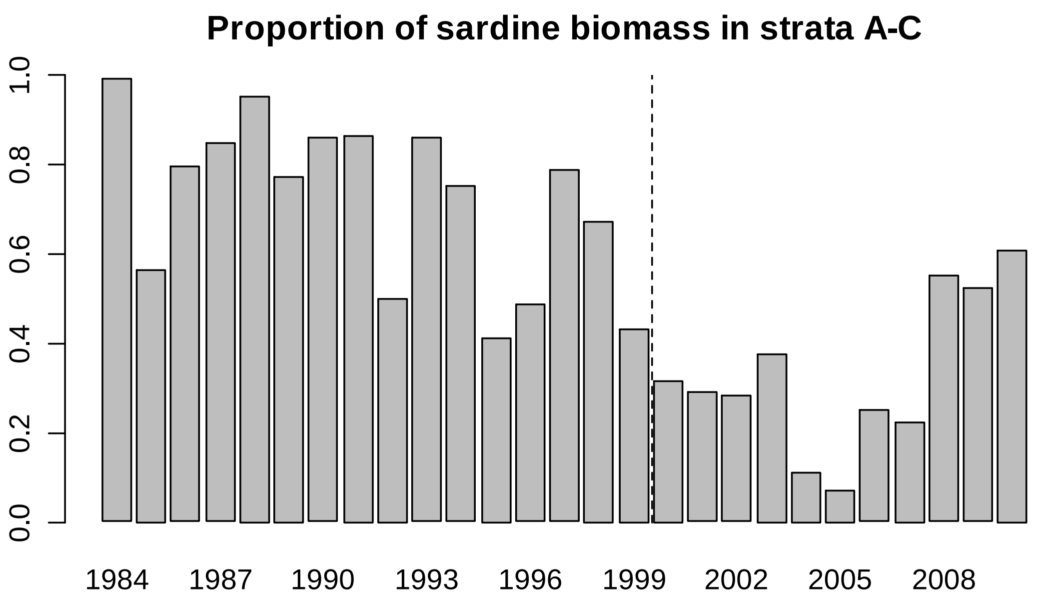


Figure 2: Time series of proportions of sardine spawner biomass observed annually in the November survey in strata A–C (west of Cape Agulhas). The dashed line indicates the division between the two pools of values used in penguin abundance projections.

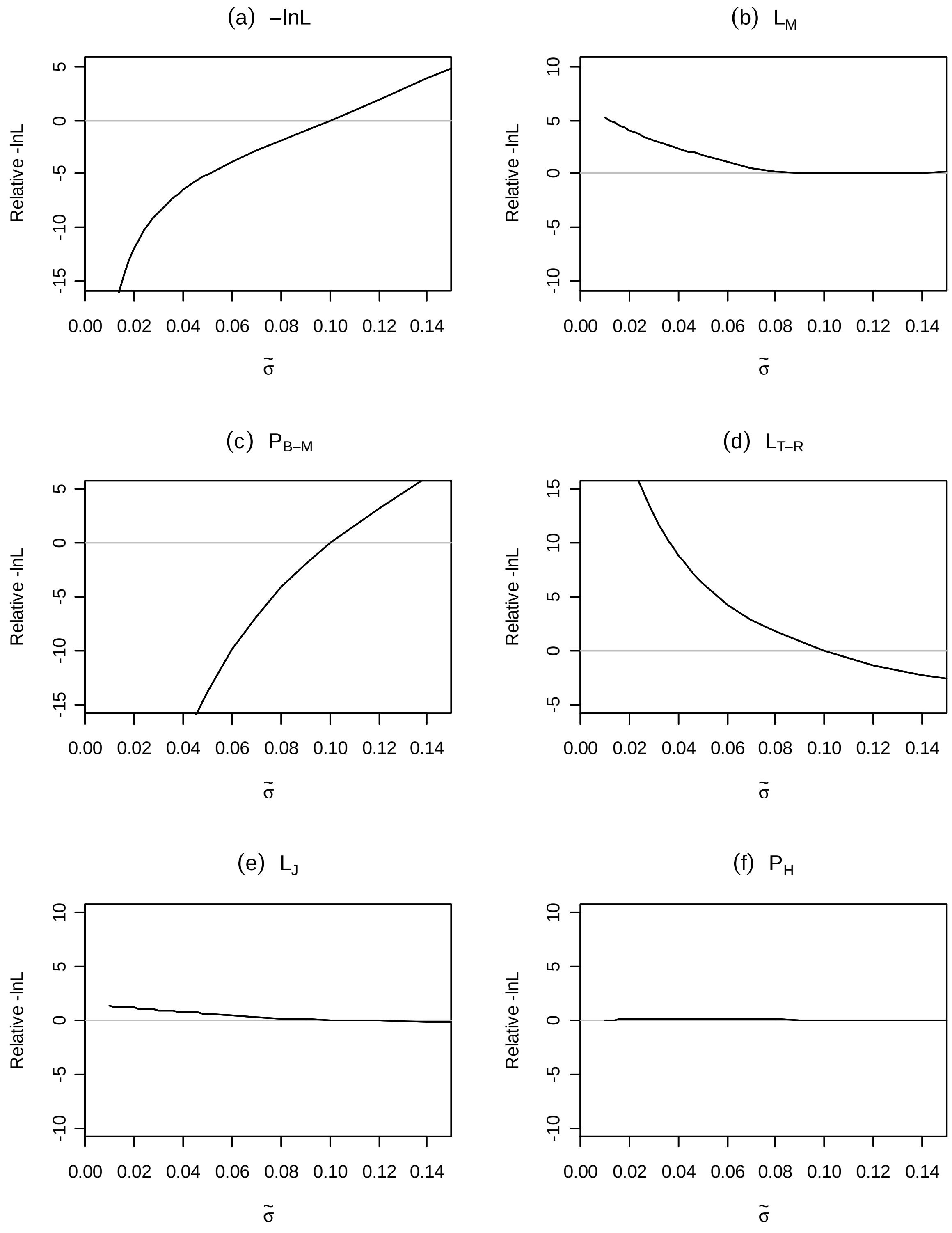


Figure 3: Relative values of the negative log likelihood and its component terms for a range of values of $\tilde{\sigma}$ between 0.01 and 0.15. In all cases, the vertical axis has been shifted so that the value at $\tilde{\sigma}=0.1$ is zero. (a) Total negative log likelihood, (b) likelihood of moult counts, (c) penalty on the biomass–mortality relationship, (d) likelihood of tag data, (e) likelihood of proportion of juveniles in the moult count data, (f) penalty on the biomass–reproductive success relationship.

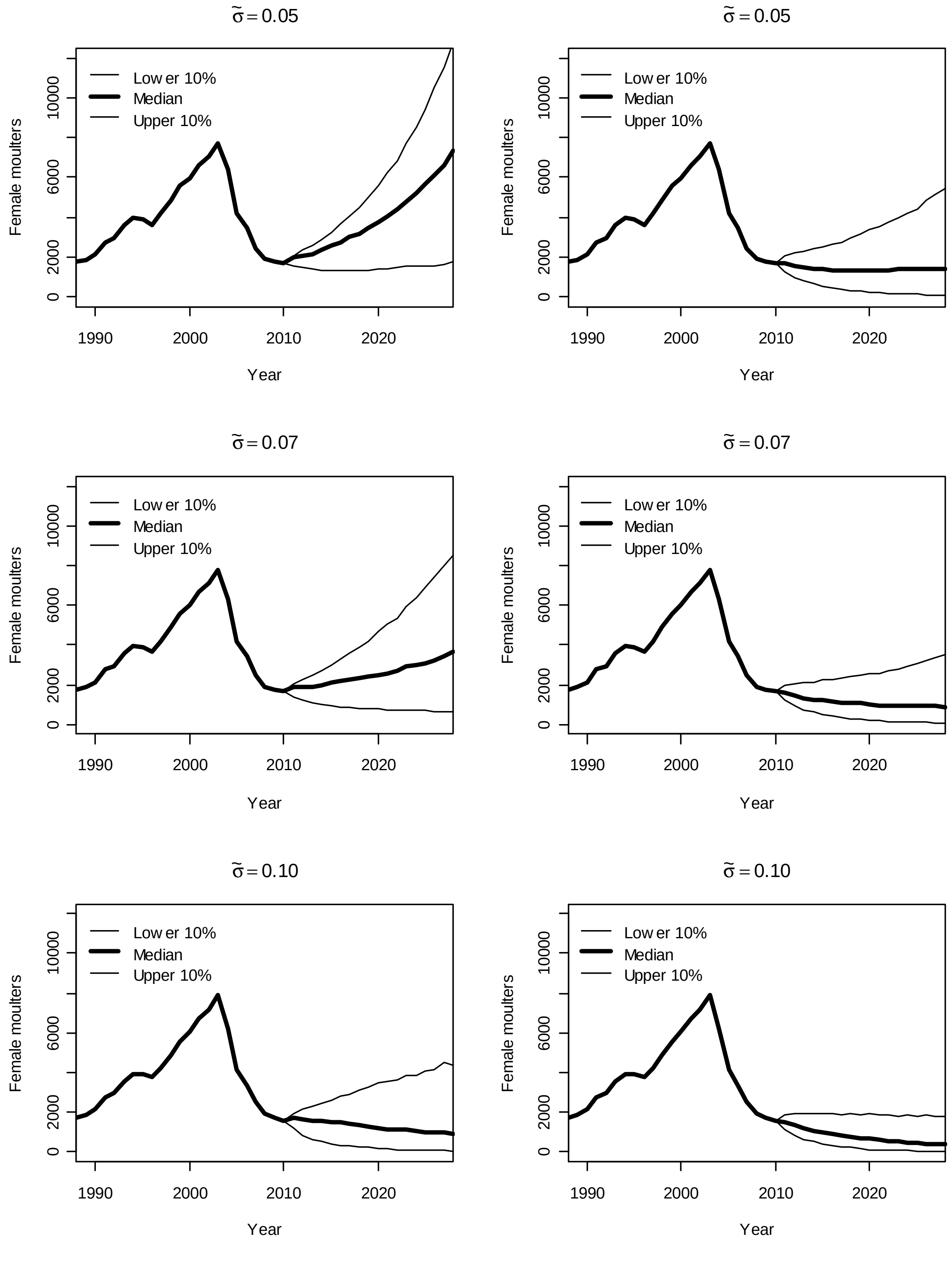


Figure 4: Trajectories of numbers of female penguins fitted to Robben Island moult count data with projections of median values and 80% confidence intervals for different values of $\tilde{\sigma}$. In the left hand column it is assumed that in future sardine will return to their pre-2000 spatial distribution, while in the right hand column values from the period 2000-2010 are used for future sardine proportions in strata A–C. The future annual fishing levels are assumed to be the median values under OMP–08.

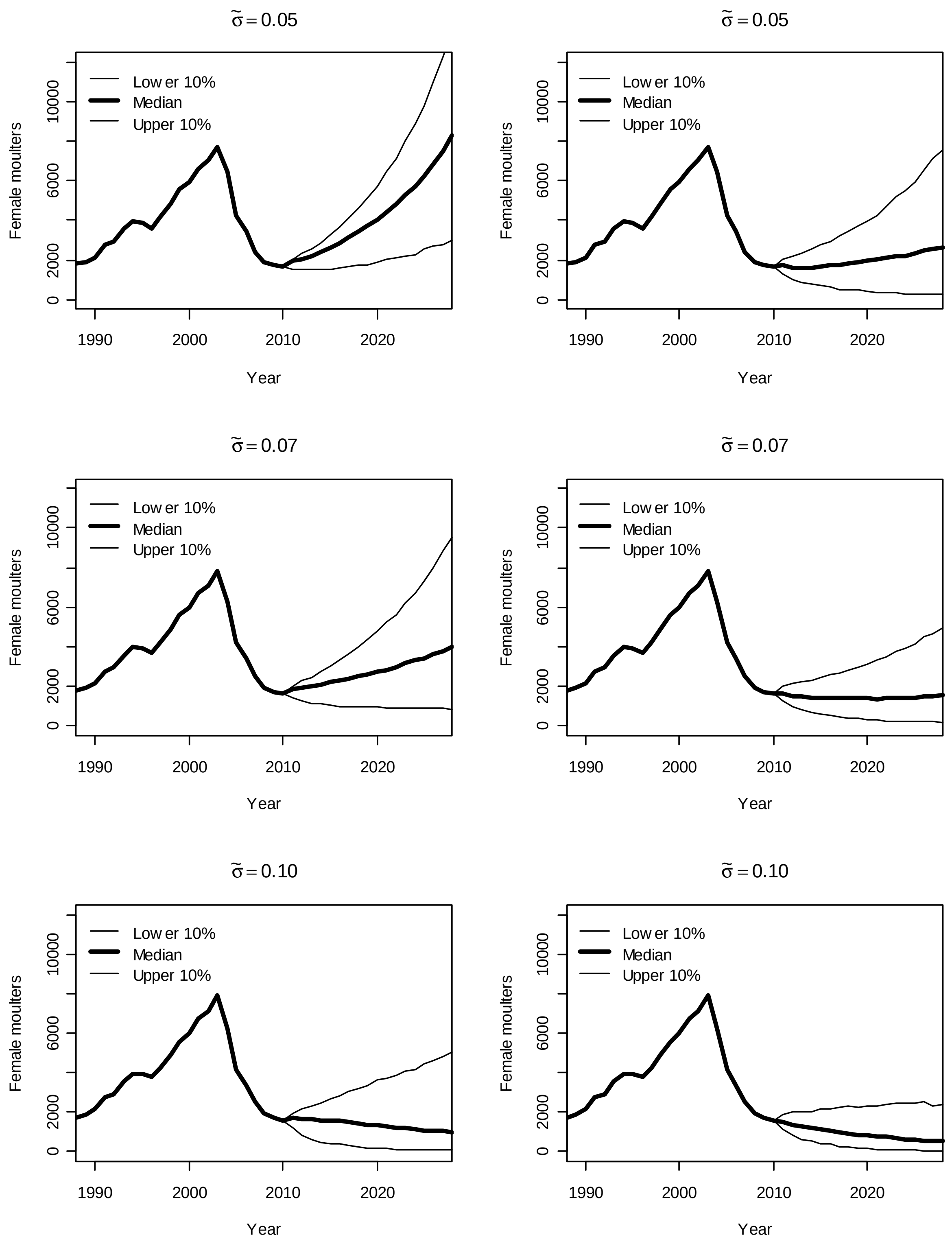
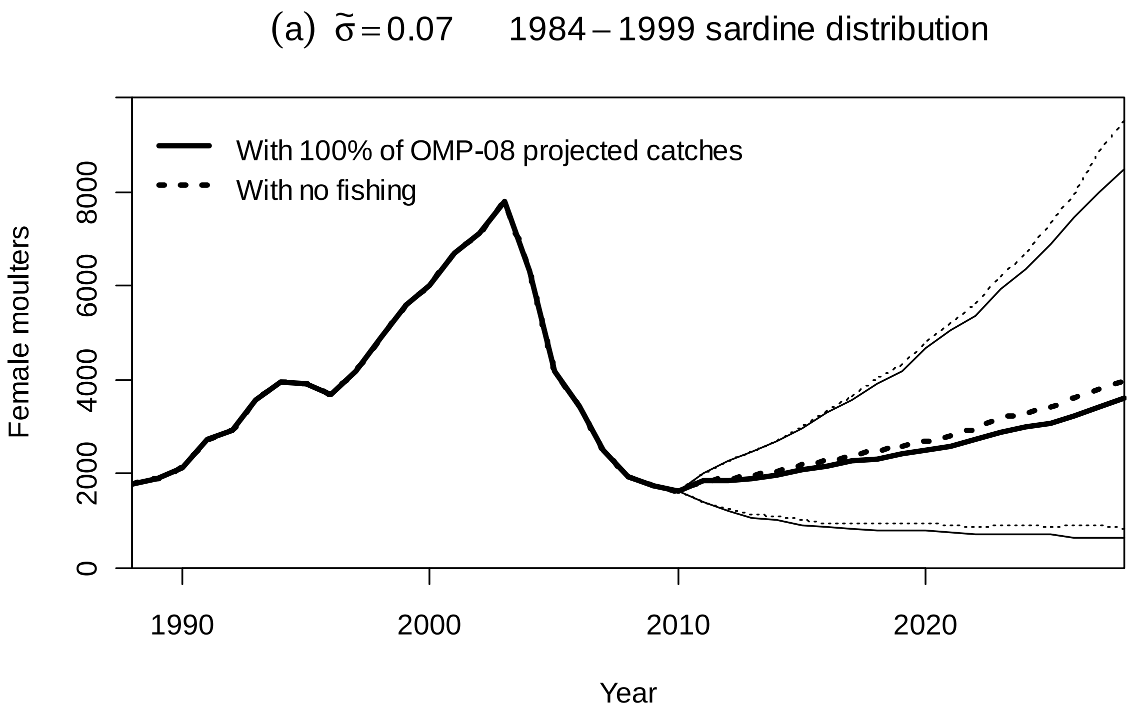


Figure 5: For these projections, it is assumed that no fishing of sardine and anchovy takes place in the future. As in Figure 4, the left hand column corresponds to projections with the 1984–1999 spatial distribution of sardine, while the right hand column assumes that the distribution of sardine will remain as it has been in the period 2000–2010.



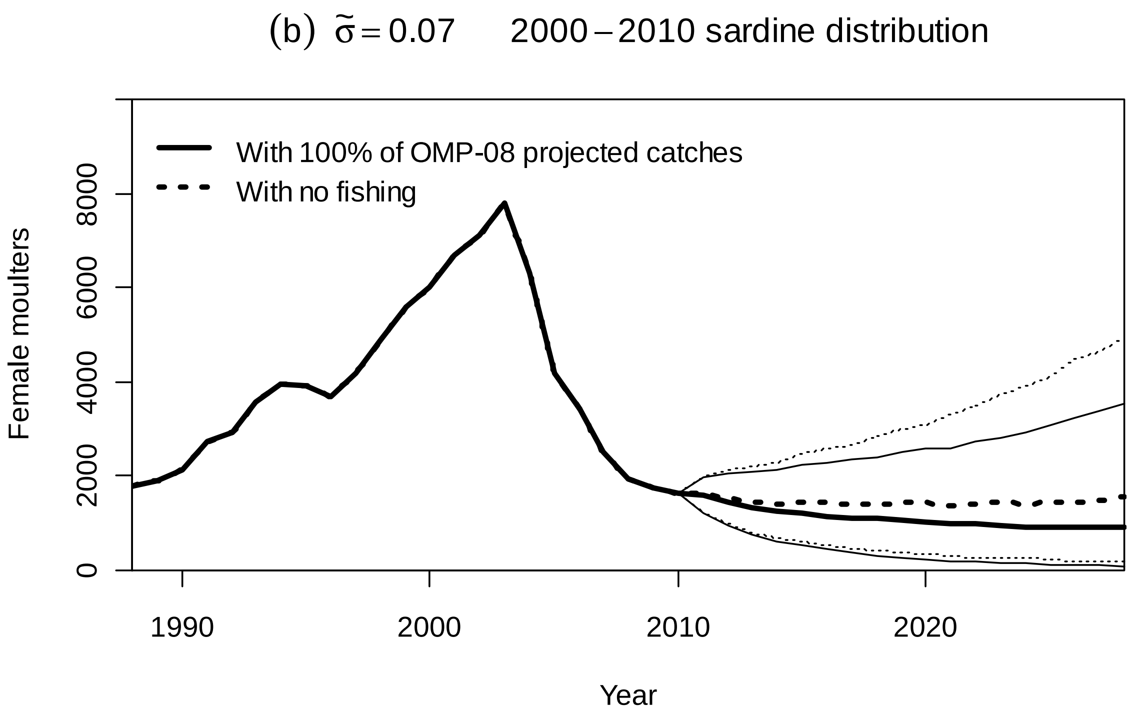


Figure 6: The median trajectories and 80% confidence intervals are compared for scenarios with and without future sardine and anchovy fishing. In plot (a) proportions of future sardine biomass assumed to be in strata A–C are drawn from the 1984–1999 values, while in plot (b) the proportions are drawn from the recent (2000–2010) values, when a greater fraction of sardine has been observed to the east of Cape Agulhas.

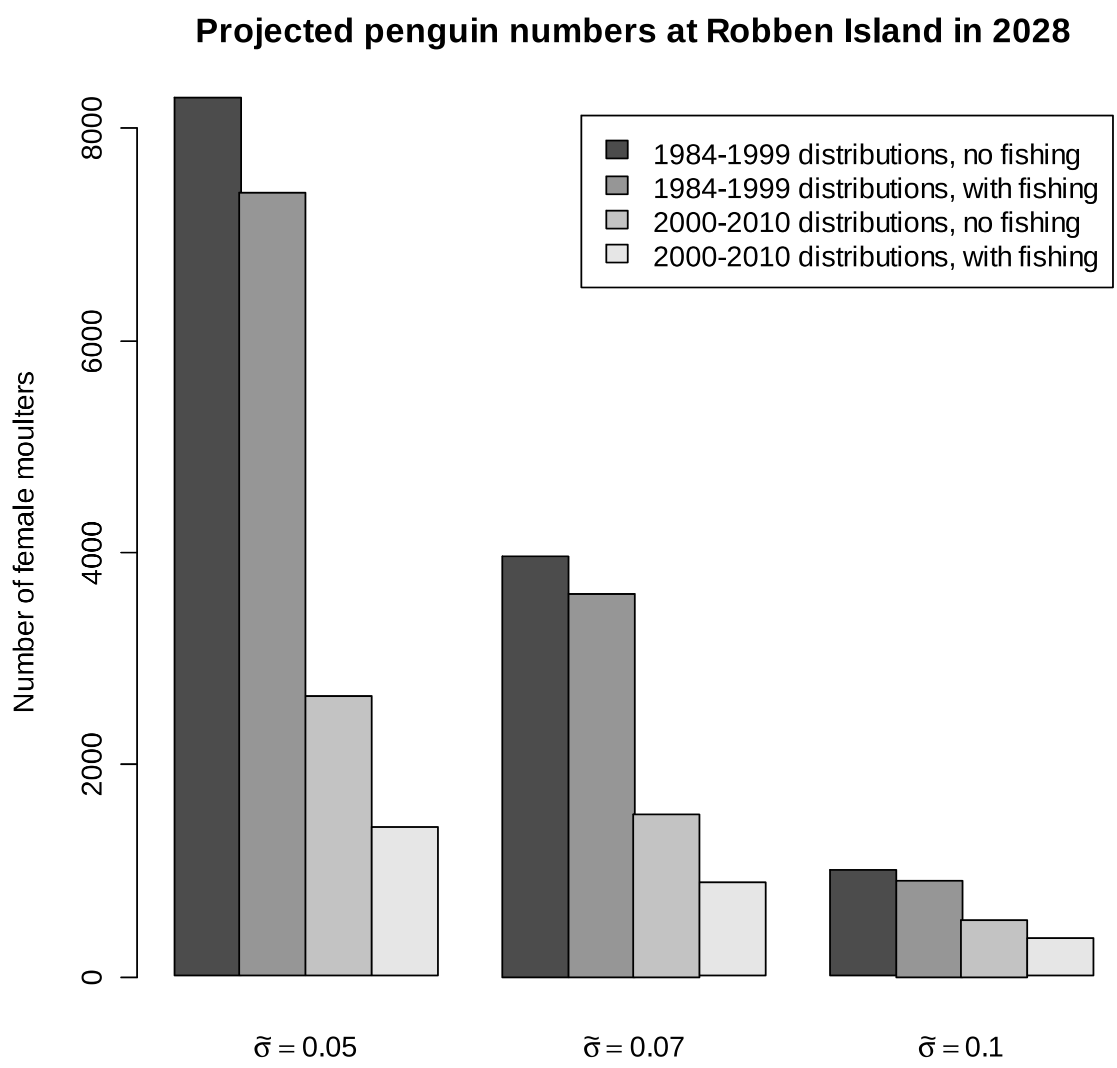


Figure 7: Projected median numbers of female penguins at Robben Island which would be expected to be observed in the moult count in the year 2028 under different assumptions for the future spatial distribution of sardine, and under fisheries either taking the median catch levels of OMP–08 or nothing. Results are shown for different assumed values of $\tilde{\sigma}$.