Summary of Progress and Plans for Penguin-related Analyses, with Reference to Comments made on Previous Submissions

D.S. Butterworth and W. Robinson

MARAM (Marine Resource Assessment and Management Group)
Department of Mathematics and Applied Mathematics
University of Cape Town
August 2010

Introduction

Modelling of penguin dynamics, including of their interaction with pelagic fish abundance, by MARAM scientists is taking place on three fronts:

I) Models of the possible impact of fishing in the vicinity of penguin colonies
II) Models of the impact of the total pelagic catch on penguin dynamics
III) Models of Robben Island penguin dynamics using tag-recapture data.

This paper summarises the current situation, offering an understanding of recent debates on certain aspects of earlier analyses, and indicates plans for future analyses on all three fronts. It should be noted that progress on I) is linked to the initiatives of the ICTT (Island Closure Task Team) of the PWG. Furthermore note that it is planned to submit results from the analyses planned for review at the annual international assessment review workshop scheduled to be held over 29 November to 3 December later this year, so that results therefrom can inform the PWG deliberations later in December from which recommendations for management of the Pelagic Fishery for 2011 will be developed.

I) Models of the possible impact of fishing in the vicinity of penguin colonies

a) GLM analyses

Related work was first detailed in EAFWG/OCT/OCT2007/STG/04, which considered estimating of the power of experimental open/close programmes at Dassen and Robben Islands to detect the effect of fishing in the neighbourhood of the penguin colonies at these islands, based on existing time series of fledging success and breeders per adult moulter indices. In addition that paper performed an initial GLM analysis to determine the direction and magnitude of the effect of pelagic catches on penguin reproductive output in terms of these two indices.

Some extensions of that work are reported in MCM/2010/SWG-PEL/27. At the recent ICTT meeting, 12 penguin data series were identified for further analyses along those lines.

Commentary

In MCM/2010/SWG-PEL/23, it was argued that the conclusion of EAFWG/OCT/OCT2007/STG/04 of positive correlations between fish catches and penguin reproduction was erroneous because “a much more likely explanation is that fisheries and penguins benefited in similar ways from variability

in the environment: a good year for fisheries would have been a good year for penguins". However, as explained in MCM/2010/SWG-PEL/27, the basic model of EAFWG/OCT/OCT2007/STG/04:

$$\ln(F^{h}_{y,i}) = \alpha_i + \beta_y + \lambda_i \frac{C_{y,i}}{\overline{C}_i} + \varepsilon_{y,i} \qquad (1)$$

where:

$F^{h}_{y,i}$ is a measure of penguin reproductive success at island i in year y,

$C_{y,i}$ is the total catch of sardine and anchovy taken within 20 km of island i in year y,

$\overline{C}_i$ is the mean annual catch over the period considered taken within 20 km of island i, and

λ_i is the parameter relating the effect of the extent of this catch around island i to the penguin response,

already takes such common influences into account by estimating an annual β_y factor which reflects common influences at the two colonies that may be good or bad for penguin reproduction in a particular year, thus avoiding such effects confounding the estimates of the λ_i parameters related to the possible impacts of fishing.

This counter to earlier queries would seem to have been understood and accepted by the ICTT.

Plans

GLM analyses of similar form to those of equation (1) above, and also considering survey estimates of recruitment as co-variates in some cases, will be pursued for the 12 series identified by the ICTT.

b) River model

A further approach to estimating this potential impact at the Robben and Dassen Island colonies is the simple application of a "River model" (see MCM/2010/SWG_PEL/Island Closure Task Team/10) of the movement of the anchovy recruits of the year down the west coast, to estimate the extents to which fishing over a series of past years has reduced the density of these fish below the levels that would have pertained in the absence of any fishing.

Commentary

Discussion in the ICTT led to the request that these analyses be extended to take account of the within-season variability in the flux of anchovy recruits down the west coast

Plans

Some measure is needed of the extent of anchovy recruitment flux variability. The ICTT suggested use of the short-time-scale variability in west coast anchovy catches, but this can be confounded by other factors such as avoidance of areas of high bycatch of juvenile sardine or horse mackerel. The immediate plan therefore is to instead first attempt to estimate this variability from the series of hydroacoustic series that took place from April to August of 2009 season in the neighbourhood of Robben Island (MCM/2010/SWG-PEL/05).

II) Models of the impact of the total pelagic catch on penguin dynamics

The decision of the PWG at its December 2008 meeting not to recommend a change in the sardine-anchovy OMP to reduce catch at lower pelagic abundance levels to possibly promote enhanced penguin recovery rates was based on a series of papers produced that year which led eventually to estimates of trade-offs between reductions in expected catch and increased rates of penguin recovery at Robben and Dassen Islands.

Following requests, that series of papers has recently been consolidated into a single document (MCM/2010/SWG-PEL/35).

It needs to be born in mind that limitations on time at that stage of those analyses meant that intended extensions of this work from the baseline ML(Penalised)E results presented then to Bayesian estimation and a fuller set of robustness tests had yet to be undertaken.

Commentary

Discussions concerning the penguin dynamics model used here have taken place, in part, in relation to the model used for investigations under III) on which further elaboration is provided below. However since some of the considerations raised are common to both modelling initiatives, it is more straightforward to deal with them here, noting also that discussions on the model for III) took place in relation to extensions of that model to incorporate Bayesian estimation.

First there were suggestions that such Bayesian results would depend on the functional form and bounds for the priors assumed for certain parameters such as penguin (median) survival rate in the analyses, but a number of computations showed that such sensitivity was very small (see e.g. MCM/2010/SWG-PEL/09).

Then concerns were expressed that the joint posterior mode estimates of a number of parameters fell on constraint boundaries (particularly that for the maximum penguin survival rate). First, having such estimates fall on boundaries can be indicative of model mis-specification, which in turn would render conclusions drawn from the model of questionable reliability. Secondly, however, it is also true that the large observation errors in the data typical in this field can readily lead to such modal estimates lying on demographic constraint boundaries for models of the type in use, and that this is therefore not of itself a sufficient reason to disregard model results. The extent to which the first of these considerations is offset by the second remains under debate.

One possible effect investigated to check whether it would reduce the number of instances of posterior mode estimates on constraint boundaries was undercounting of juvenile relative to adult penguins in moult counts. The associated analyses did not lead to such a reduction, but did suggest that there was a relative undercount whose extent was in the 60% to 80% range (MCM/2010/SWG-PEL/27). Comments on the plausibility of this result by penguin biologists in the PWG familiar with the moult count exercise would be appreciated.

Associated with the issue of some posterior mode estimates for survival rate on a demographic constraint boundary, the concern has been raised of whether this could in turn lead to bias (and in particular negative bias) under the approach used in 2008 (see summary in MCM/2010/SWG-PEL/35) to estimate the impact of reduced abundance of pelagic fish on penguin reproduction and survival rate parameters.

There has also been debate on the most appropriate way to display the results of Bayesian analyses for posterior distributions of certain quantities when the joint posterior mode lies on a boundary of that distribution, given that while ideally full distributions should be plotted, in practice that would render results too voluminous. Thus summary statistics of these distributions need to be provided instead. There are two candidates for 90% probability/credibility intervals:

a) the symmetric approach of stating the 5 and 95%iles, which might exclude the joint posterior mode; and

b) the selection of two percentiles that minimise the width of the 90% interval; under this approach the joint posterior mode is less likely to be excluded.

The second of these approaches seems to enjoy wider preference in the PWG. As far as a statistic characterising the centre rather than the width of posterior distributions is concerned, the median seems preferable to the joint posterior mode, both because of greater stability in a time series context and also because the final results of interest (the impact of low pelagic abundances on penguin population trends) involve integration over such posterior distributions, for which the medians provide a better characterisation of likely impacts on results.

It is important, in considering these matters, to realise that it is results for the likely impact of reduced pelagic fish abundance on penguin dynamics that is the point of focus. In that context, parameters such as penguin survival rates can be considered as "nuisance parameters", so that possible bias in their estimation would not be a concern if the impacts on penguin dynamics themselves were being reliably estimated.

Plans

First the 2008 analysis approach described in MCM/2010/SWG-PEL/35 will be extended to a fully Bayesian form.

The robustness of the results of that approach to alternative assumptions will be thoroughly explored. This will include consideration of:

- Different assumptions for the components (recruit or spawner biomass survey, which area) of the sardine and anchovy abundances to which penguin demographic parameters are most closely linked
- Alternative functional forms for that dependence
- Alternative models for the error structure of that relationship
- The impact of possible immigration on the dynamics of penguins in the model.

Thus, for example, one possible improvement to the existing model that has been suggested (A E Punt, pers. commn) is to replace the current error model for the survival rate *vs* fish abundance relationship:

$$S_{y,i} = S_{\max} \frac{\exp\left(\alpha_i B_y + \beta_i + \eta_{y,i}\right)}{1+\exp\left(\alpha_i B_y + \beta_i + \eta_{y,i}\right)} \qquad \eta_{y,i} \sim N\left(0, \sigma_\eta^2\right) \tag{2}$$

by a beta distribution whose expectation is given by the multiplier of S_{max} equation (2) with η set to zero, so that the variance of $S_{y,i}$ is unrelated to its mean.

A demonstration that results for projected penguin population growth rates are indeed robust to such variations would considerably enhance the appropriate level of faith to be placed in them, but would still not completely address the issue of the possible problems raised above in relation to joint posterior mode estimates of a number of annual penguin survival rates falling on demographic boundaries. To deal with that, simulation testing of the Bayesian estimation method is planned to be undertaken, with pseudo-data sets for such testing purposes being generated from a model whose parameters are fixed at the posterior medians of the marginal posteriors for a baseline fit of the model to (now slightly updated) data. The parameters of the error distributions used to generate those pseudo-data will be as estimated in the original fit of the model. This would indicate whether there are any serious biases in the estimates from the model of the impact of pelagic fish abundance on future penguin population trends, as well as throw light on whether having joint posterior mode survival rate estimates on constraint boundaries does adversely impact the reliability of results.

III) Models of Robben Island penguin dynamics using tag-recapture data

The first model of this type is detailed in MCM/2009/SWG-PEL/33. It treats Robben Island penguins as a closed population, with input data including in particular tag-recapture data to inform on distributions for non-juvenile survival rates and their temporal changes. Its primary intent was to estimate the extent of over-dispersion in the tag-recapture data with a view to subsequently estimate the power of possible future tagging programmes to assess the impact of fishing near colonies on penguin survival rates through an experimental open/close regime.

Estimates were originally presented on a Maximum Penalised Likelihood Basis, but this was later extended to a full Bayesian approach to provide posterior distributions in subsequent documents (e.g. MCM/2010/SWG-PEL/07).

Commentary

A concern expressed about this model is that the tag-recapture process is modelled by an over-dispersed Poisson distribution instead of the more conventional multinomial distribution approach of standard mark-recapture analysis packages such as MARK.

Plans

The model will be rerun with the Poisson likelihood replaced by the multinomial form used in MARK. Success has now been achieved in adapting the MARK software to provide likelihood values for specific parameter value inputs (WITHOUT optimising!), and work in currently in progress to link this software to the remainder of the estimation code for the model.

References

Altwegg R. 2010. Remaining problems with penguin analysis and the interpretation of evidence for fisheries benefitting penguins. Document MCM/2010/SWG-PEL/23.

Brandão A, Butterworth DS. 2007. An initial analysis of the power of monitoring certain indices to determine the effect of fishing on penguin reproductive success from an experiment where pelagic fishing is prohibited in the neighbourhood of Robben Island, but continues around Dassen Island. Document EAFWG/OCT2007/STG/04.

Butterworth DS, de Moor CL. 2010. A very simple implementation of the "River Model" to estimate the impact of fishing on the amount of anchovy available to West Coast penguin colonies. Document MCM/2010/SWG_PEL/Island Closure Task Team/10.

Butterworth DS, Robinson W. 2010. Response to Comments on : Robben Island penguin survival rates from Bayesian analysis by William Robinson and Doug Butterworth (Altwegg: MCM/2010/SWG-PEL/08). Document MCM/2010/SWG-PEL/09.

Butterworth DS, Robinson W. 2010. Further responses to comments in Altwegg MCM/2010/SWG-PEL/23 regarding penguin analyses. Document MCM/2010/SWG-PEL/27.

Coetzee J, Merkle D, Shabangu F, Rademan J. 2010. Updates and corrections to inshore survey biomass estimates. Document MCM/2010/SWG-PEL/05.

Robinson W, Butterworth D. 2009. Fitting both moult counts and tagging data to a population model for Robben Island penguins. Document MCM/2009/SWG-PEL/33.

Robinson W, Butterworth D. 2010. Robben Island penguin survival rates from Bayesian analysis. Document MCM/2010/SWG-PEL/07.

Robinson W, Plagányi É, Butterworth D, de Moor C. 2010. Summary of penguin–pelagic fish interaction modelling during 2008. Document MCM/2010/SWG-PEL/35.