Updated GLMM standardisation of the commercial abalone CPUE for Zone F

Anabela Brandão and Doug S. Butterworth

Marine Resource Assessment & Management Group (MARAM)
Department of Mathematics and Applied Mathematics
University of Cape Town
Rondebosch, 7701, South Africa

August 2015

Abstract

This paper presents an update of the standardisation of the abalone CPUE for Zone F using a GLMM approach proposed by Brandão and Butterworth (2012), which adds new data for the 2011/12 to 2013/2014 fishing "years" for that Zone. The standardised CPUE indices have been fairly stable since 2007/08 fishing "year", with a slight drop in the latest year.

Introduction

Similar catch-per-unit-effort (CPUE) General Linear Mixed Model (GLMM) standardisation procedure described in Brandão and Butterworth (2012) for Zones A to D and have been applied to the commercial abalone data for Zone F. This procedure was first reported in Brandão and Butterworth (2012), and is updated on this paper to take into account further data that are now available. The principle objective of the GLMM analyses is to obtain series of relative abundance indices that have been standardised by incorporating important covariates in the explanation of abalone CPUE variation.

The data

Commercial catch data (as kg whole mass), and effort data (as total duration of dives in minutes for each day dived) are available for the period 1981 to 2014. A year in this paper refers to a Model-year, where a Model-year y runs from October of year y-1 to September of year y. The covariates included in the GLM analysis include the date (in terms of Model-year and season (3-monthly periods)), and the divers. Records with a dive time less than 10 minutes were excluded. A total of 1 833 data points remained for the analysis. Table 1 gives the number of records used in the GLMM analysis per Model-year.

General Linear Mixed Model (GLMM) to standardise the CPUE

The GLM (General Linear Model) used to standardise commercial CPUE indices assumes that all factors in the model are fixed effects with the variance of the response values being that of the error term ε. In a GLM analysis we model only the mean (i.e. the fixed effects) of the data. A GLMM has the ability to model not only the mean of the data but also its variance. In fact, a GLMM also allows for the presence of random variables (called random effects) which describe additional variability in the data apart from that reflected by the error term of equation (1). One of the covariates used in the GLM by Plagányi and Edwards (2007) is "divers" with 442 different levels (in the present analysis) associated with different divers, with some of the divers in the fishery having very few dives. The alternative approach proposed by Brandão and Butterworth (2012) is to treat "divers" as a random effect in a GLMM.

The GLMM applied to the abalone commercial CPUE data is of the form:

$$\ln\left(CPUE\right) = \mathbf{X}\alpha + \mathbf{Z}\beta + \varepsilon, \quad (1)$$

where :

CPUE is the catch-per-unit-effort defined as catch (kg) divided by dive time (minutes),

α is the unknown vector of fixed effects parameters which includes:

$\mu + \alpha_{year} + \beta_{season} + \eta_{year \times season}$, where

μ is the intercept,

year is a factor with 30 levels associated with the Model-years 1981–2014 (excluding the years 1984, 1985, 1999 and 2009. In 2009 the fishery was closed),

season is a factor with 4 levels associated with the season effect (1 = Jan-Mar; 2 = Apr-Jun; 3 = Jul-Sep; 4 = Oct-Dec), and

year×*season* is the interaction between year and season,

X is the design matrix for the fixed effects,

β is the unknown vector of random effects parameters (here diver which is a factor with 442 levels associated with the diver code, which includes both the entitlement holders coded in the database as well as "divers",

Z is the design matrix for the random effects,

ε is an error term assumed to be normally distributed and independent of the random effects.

This approach assumes that both the random effects and the error term have zero mean, i.e. E(β) = E(ε) = 0, so that $\mathrm{E}\left(\ln\left(CPUE\right)\right) = X\alpha$. The variance-covariance matrix for the residual errors (ε) is denoted by **R** and that for the random effects (β) by **G**. The analyses undertaken here assume that the residual errors as well as the random effects are homoscedastic and are uncorrelated, so that both **R** and **G** are diagonal matrices given by:

$$\mathbf{R} = \sigma_{\varepsilon}^{2}\mathbf{I}$$
$$\mathbf{G} = \sigma_{\beta}^{2}\mathbf{I}$$

where **I** denotes an identity matrix. Thus, in the mixed model, the variance-covariance matrix (**V**) for the response variable is given by:

$$\mathrm{Cov}\left(\ln\left(CPUE\right)\right) = \mathbf{V} = \mathbf{ZGZ}^{\mathrm{T}} + \mathbf{R}\,,$$

where $\mathbf{Z}^{\mathrm{T}}$ denotes the transpose of the matrix **Z**.

The estimation of the variance components (**R** and **G**), the fixed effects (α) and the random effects (β) parameters in GLMM requires two steps. First the variance components are estimated by the method of residual maximum likelihood (REML), which produces unbiased estimates for the variance components as it takes into account the degrees of freedom used in estimating the fixed effects. Once estimates of **R** and **G** have been obtained, estimates for the fixed effects parameters (α) can be obtained as well as predictors for the random effects parameters (β).

For this model, because of interactions with year (which imply changing spatio-temporal distribution patterns), the standardised CPUE series is obtained from:

$$CPUE_{year} = \left[\sum\nolimits_{season}\left(\exp\left(\mu + \alpha_{year} + \beta_{season} + \varphi_{diver} + \eta_{year \times season}\right)\right)\right] / 4 \qquad (2)$$

where the standardisation is with respect to a diver code = 110, which contained the most observations.

The reason for standardising in this way when year interactions are present is that the standardised CPUE is to be used as an index of relative abundance when input to assessment models. CPUE itself is assumed to be proportional to local density, so that averaging over season is necessary to provide a quantity representative of a consistently calculated average over each year. This averaging is unnecessary in the absence of such interactions, because then the $\exp\left(\alpha_{year}\right)$ term alone would then be proportional to abundance.

Results and Discussion

Table 2 lists the nominal and the GLMM-standardised CPUE indices provided by the model and Figure 1 shows these graphically. There is not much difference between the GLMM standardised CPUE and the nominal series. Both drop slightly for the latest year.

Reference

Brandão, A. and Butterworth, D.S. 2012. GLM and GLMM standardization of the commercial abalone CPUE for Zones A-D. FISHERIES/2012/AUG/SWG-AB/04.

Brandão, A. and Butterworth, D.S. 2012. GLM and GLMM standardization of the commercial abalone CPUE for Zone F. FISHERIES/2012/AUG/SWG-AB/09.

Plagányi, É.E. and Butterworth, D.S. 2010. A spatial- and age-structured assessment model to estimate the impact of illegal fishing and ecosystem change on the South African abalone *Haliotis midae* resource. African Journal of Marine Science, 32(2):207-236.

Table 1. The number of data entries per Model year available for the final GLM analysis to standardise the commercial abalone CPUE series for Zone F are shown. The abalone fishery was closed in February 2008 and reopened in 2010. Model-years are defined as the period October of the preceding year to September of the year indicated.

Model year	Number of records
1981	24
1982	16
1983	31
1984	
1985	
1986	65
1987	48
1988	48
1989	63
1990	61
1991	42
1992	63
1993	23
1994	32
1995	17
1996	16
1997	31
1998	36
1999	
2000	39
2001	40
2002	56
2003	58
2004	26
2005	24
2006	22
2007	89
2008	269
2009	
2010	72
2011	126
2012	103
2013	87
2014	206

Table 2. Nominal, GLM and GLMM-standardised commercial CPUE series for abalone for Model-years (October of the preceding year to September of the year indicated) 1981 to 2014 for Zone F. The nominal and both sets of standardised values have been divided by the mean value of the respective series.

Model year	Nominal CPUE	GLMM standardised CPUE
1981	1.499	1.212
1982	1.518	1.129
1983	1.433	1.204
1984		
1985		
1986	1.420	1.276
1987	1.183	1.166
1988	1.617	1.816
1989	0.782	0.854
1990	1.548	1.627
1991	1.202	1.164
1992	0.844	0.888
1993	1.030	1.144
1994	1.983	2.159
1995	1.374	1.578
1996	1.596	1.759
1997	1.346	1.313
1998	0.984	1.065
1999		
2000	0.520	0.467
2001	0.770	0.687
2002	0.759	0.695
2003	0.663	0.603
2004	0.740	0.771
2005	1.104	1.178
2006	1.182	1.290
2007	0.319	0.351
2008	0.528	0.446
2009		
2010	0.418	0.437
2011	0.409	0.438
2012	0.424	0.424
2013	0.447	0.519
2014	0.357	0.340

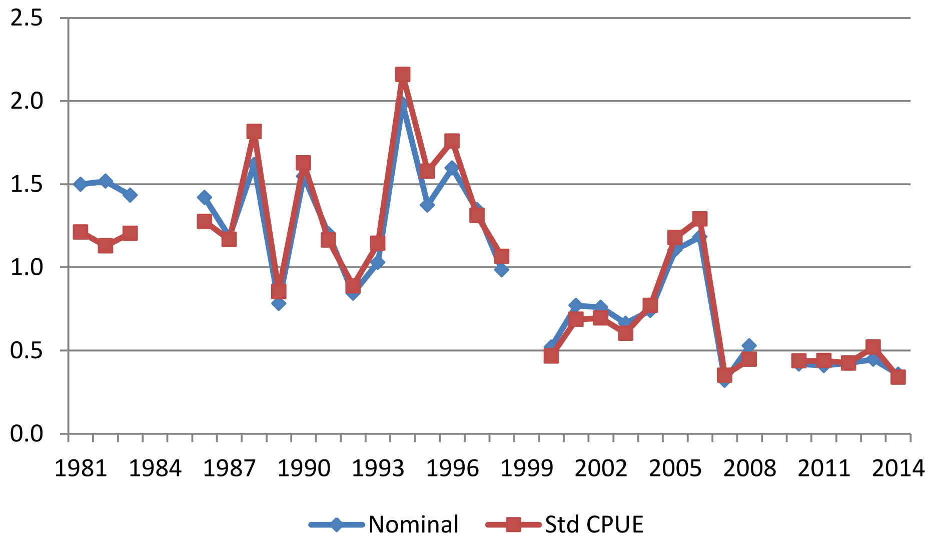

Figure 1. GLMM-standardised CPUE trends (normalised to their means over the 30 year period) for Zone F. For comparison, the nominal series (also normalised to its mean over the 30 year period) is also shown.