GLMM standardisation of the commercial abalone CPUE for Zones A-D over the period 1980–2012

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 2013

Abstract

This paper presents an update of the standardisation of the abalone CPUE using a GLMM approach proposed by Brandão and Butterworth (2012) which adds new data for the 2011/2012 fishing in Zones A and B. The standardised values of CPUE for 2011/2012 for each of these Zones are very similar to the corresponding values from one year earlier.

Introduction

Brandão and Butterworth (2012) compared the standardised catch-per-unit-effort (CPUE) series obtained from a General Linear Model (GLM) procedure described in Plagányi and Edwards (2007) and Brandão and Butterworth (2009) with the standardised CPUE series obtained from a General Linear Mixed Model (GLMM) approach. Very little difference between the results for the two different methods of standardisation was found and it was thus decided that in the future a GLMM approach would be applied to the commercial abalone data to obtain a standardised CPUE series. In this paper the GLMM described in Brandão and Butterworth (2012) has been applied to the commercial abalone data for Zones A-D to incorporate the further data now available for Zones A and B for the 2012 Model-year, where a Model-year *y* runs from October of year *y*-1 to September of year *y*, as well as some further data which have become available for 2011. The principle objective of the GLMM analysis 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 1980 to 2012. The covariates included in the GLMM analysis include the date (in terms of Model-year and season (3-monthly periods)), the divers, and the Zones that were dived. Zone C is split into subareas CNP (nonpoached) and CP (poached). Records with a dive time less than 10 minutes were excluded as well as outliers based upon observations with large residuals (> 6 standard deviations) in an initial GLM fit (see Plagányi and Edwards, 2007 and Brandão and Butterworth, 2009). Years which had too few records (less than eight) in a Zone/subarea were also excluded as were records for divers that had less than eight dives in the whole database. A total of 41 575 data points remained for the analysis. Table 1 gives the number of records used in the final GLMM analysis per Model-year and per Zone/subarea.

General Linear Mixed Model (GLMM) to standardise the CPUE

The GLM used by Plagányi and Edwards (2007) and Brandão and Butterworth (2009) 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 ε. One of the covariates that was used in the GLM by Plagányi and Edwards (2007) is “divers” with 296 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 the GLMM.

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

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

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:

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

μ is the intercept,

year is a factor with 32 levels associated with the Model-years 1980–2012 (excluding 2009 during which 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), *zone* is a factor with 5 levels associated with the different zones/subareas (A, B, CNP, CP and D),

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

year×zone is the interaction between year and zones/subareas, and

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 296 levels associated with the diver code, which includes both the entitlement holders coded in the database as well as "divers". Some divers not yet allocated a code were given a temporary code of 555 for the purposes of this analysis[1]),

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).

[1] For the Model years 2006 to 2012 over which this code was used, such records comprise 0.04% of the total.

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

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

where the standardisation is with respect to a fisher code = 8, which contained the most observations as well as the longest period in operation in the fishery.

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 will 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 graphical comparisons of the same. Broadly speaking, the standardisation makes relatively little difference to the nominal trends. Table 3 shows the parameter estimates, together with standard errors, obtained for the single fixed factors included in the GLMM model. The standardised CPUE values for Zones A and B for the 2012 model-year are very similar to the corresponding values for 2011.

Reference

Brandão, A. and Butterworth, D.S. 2009. A summary of the General Linear Model analyses applied to the commercial abalone CPUE data for Zones A-D over the period 1980-2008. MCM/2009/OCT/SWG-AB/06.

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

Plagányi, É. and Edwards, C. 2007. Summary of the GLM used to standardise abalone catch-per-unit-effort data for Zones A-D over the period 1980-2006. Marine and Coastal Management document: WG/AB/07/Aug/19.

Table 1. The number of data entries per Zone available for the final GLMM analysis to standardise the commercial abalone CPUE series are shown. Subarea CNP was closed during the 2001 fishing season and subarea CP during both the 2001, 2002 and 2003 fishing seasons. The abalone fishery was closed in February 2008 and reopened in 2010. Some sample sizes were considered too small and were not included in the analysis (see text). Model-years are defined as the period October of the preceding year to September of the year indicated.

Model year	**Zone/subarea**				
	A	**B**	**CNP**	**CP**	**D**
1980	257	555	73	753	535
1981	192	578	147	622	383
1982	311	610	109	594	608
1983	327	690	144	466	301
1984	334	696	274	364	373
1985	359	619	158	366	583
1986	340	763	222	445	205
1987	443	585	105	494	144
1988	457	434	96	498	147
1989	447	414	91	504	184
1990	525	410	138	458	140
1991	446	403	161	539	167
1992	347	302	98	396	142
1993	299	237	110	333	75
1994	345	290	155	287	162
1995	441	238	137	333	170
1996	506	324	401	427	206
1997	720	248	249	116	194
1998	599	472	205	71	290
1999	667	412	56	8	295
2000	447	319	23		302
2001	387	286			133
2002	284	226	96		94
2003	412	128	54		25
2004	97	566	155		60
2005	63	595	167		52
2006	41	671	164		48
2007		482			
2008		291			
2009					
2010	175	229			
2011	363	381			
2012	279	331			

Table 2. Nominal and GLMM-standardised commercial CPUE series for abalone for Model-years (October of the preceding year to September of the year indicated) 1980 to 2012 and Zones/subareas A, B, CNP, CP and D. Both the nominal and the standardised values have been divided by the mean value of the respective series.

a) Nominal CPUE series

Model year	Zone/subarea				
	A	B	CNP	CP	D
1980	1.081	0.798	0.873	0.840	0.907
1981	1.055	0.800	0.897	0.834	0.833
1982	0.926	0.809	0.881	0.834	0.803
1983	0.913	0.790	0.939	0.875	0.723
1984	0.991	0.843	0.962	0.891	0.796
1985	0.930	0.854	0.919	0.965	0.809
1986	1.030	0.931	1.023	1.103	0.770
1987	1.059	0.913	1.144	1.057	0.868
1988	1.137	1.001	1.185	1.148	1.033
1989	1.040	1.018	1.154	1.116	0.894
1990	1.193	1.241	1.417	1.215	1.245
1991	1.192	1.286	1.222	1.102	1.231
1992	1.317	1.350	1.264	1.234	1.163
1993	1.425	1.604	1.089	1.295	1.909
1994	1.353	1.387	1.219	1.324	1.709
1995	1.259	1.495	1.252	1.130	1.476
1996	1.249	1.413	0.979	0.904	1.439
1997	1.156	1.503	0.892	0.727	1.495
1998	1.195	1.350	0.983	0.737	1.548
1999	1.031	1.187	0.989	0.672	1.035
2000	1.064	1.199	1.092		0.950
2001	1.057	1.106			0.848
2002	1.059	1.135	1.278		0.769
2003	0.861	1.078	0.803		0.489
2004	0.826	0.795	0.568		0.449
2005	0.522	0.714	0.519		0.391
2006	0.554	0.601	0.458		0.418
2007		0.518			
2008		0.511			
2009					
2010	0.641	0.765			
2011	0.459	0.501			
2012	0.424	0.509			

b) GLMM-standardised CPUE series

Model year	Zone/subarea				
	A	B	CNP	CP	D
1980	1.092	0.852	0.942	0.915	0.947
1981	1.092	0.861	1.011	0.924	0.882
1982	0.950	0.887	1.016	0.913	0.864
1983	0.925	0.838	0.986	0.929	0.753
1984	0.988	0.872	0.978	0.939	0.827
1985	0.938	0.884	0.933	0.977	0.843
1986	0.964	0.956	1.040	1.122	0.905
1987	1.013	0.906	1.110	1.025	0.963
1988	1.032	0.971	1.107	1.080	1.077
1989	1.059	1.038	1.135	1.127	1.008
1990	1.107	1.143	1.271	1.136	1.227
1991	1.123	1.081	1.148	1.075	1.057
1992	1.277	1.309	1.277	1.231	1.214
1993	1.199	1.428	1.091	1.276	1.673
1994	1.129	1.067	1.186	1.203	1.267
1995	1.211	1.263	1.203	1.118	1.218
1996	1.260	1.375	1.056	1.002	1.343
1997	1.193	1.503	0.865	0.728	1.431
1998	1.180	1.377	0.970	0.694	1.477
1999	1.080	1.303	1.019	0.590	1.082
2000	1.065	1.169	1.053		0.958
2001	1.081	1.126			0.860
2002	1.140	1.117	1.145		0.799
2003	0.939	1.069	0.782		0.564
2004	0.915	0.881	0.631		0.660
2005	0.633	0.751	0.558		0.501
2006	0.708	0.698	0.488		0.600
2007		0.627			
2008		0.590			
2009					
2010	0.713	0.878			
2011	0.515	0.582			
2012	0.477	0.599			

Table 3. Parameters estimates and standard errors for the single fixed factors *Year*, *Season* and *Zone* included in the GLMM to obtain standardised indices of abundance for abalone.

	Parameter estimate	**Standard error**
Year		
1980	0.000	—
1981	0.034	0.026
1982	0.043	0.026
1983	0.062	0.027
1984	0.075	0.026
1985	0.112	0.028
1986	0.100	0.027
1987	0.085	0.029
1988	0.175	0.031
1989	0.173	0.030
1990	0.410	0.030
1991	0.321	0.034
1992	0.500	0.041
1993	0.536	0.050
1994	0.349	0.042
1995	0.344	0.044
1996	0.558	0.031
1997	0.545	0.046
1998	0.600	0.027
1999	0.471	0.030
2000	0.439	0.034
2001	0.409	0.034
2002	0.371	0.035
2003	0.308	0.046
2004	0.317	0.047
2005	0.069	0.030
2006	-0.109	0.030
2007	-0.207	0.047
2008	-0.468	0.034
2009	—	—
2010	-0.051	0.036
2011	-0.408	0.033
2012	-0.349	0.036
Season		
Jan-Mar	0.000	—
Apr-Jun	-0.016	0.021
Jul-Sep	0.115	0.021
Oct-Nov	0.114	0.059
Zone		
A	0.393	0.030
B	0.000	—
CNP	-0.034	0.050
CP	-0.056	0.023
D	0.156	0.025

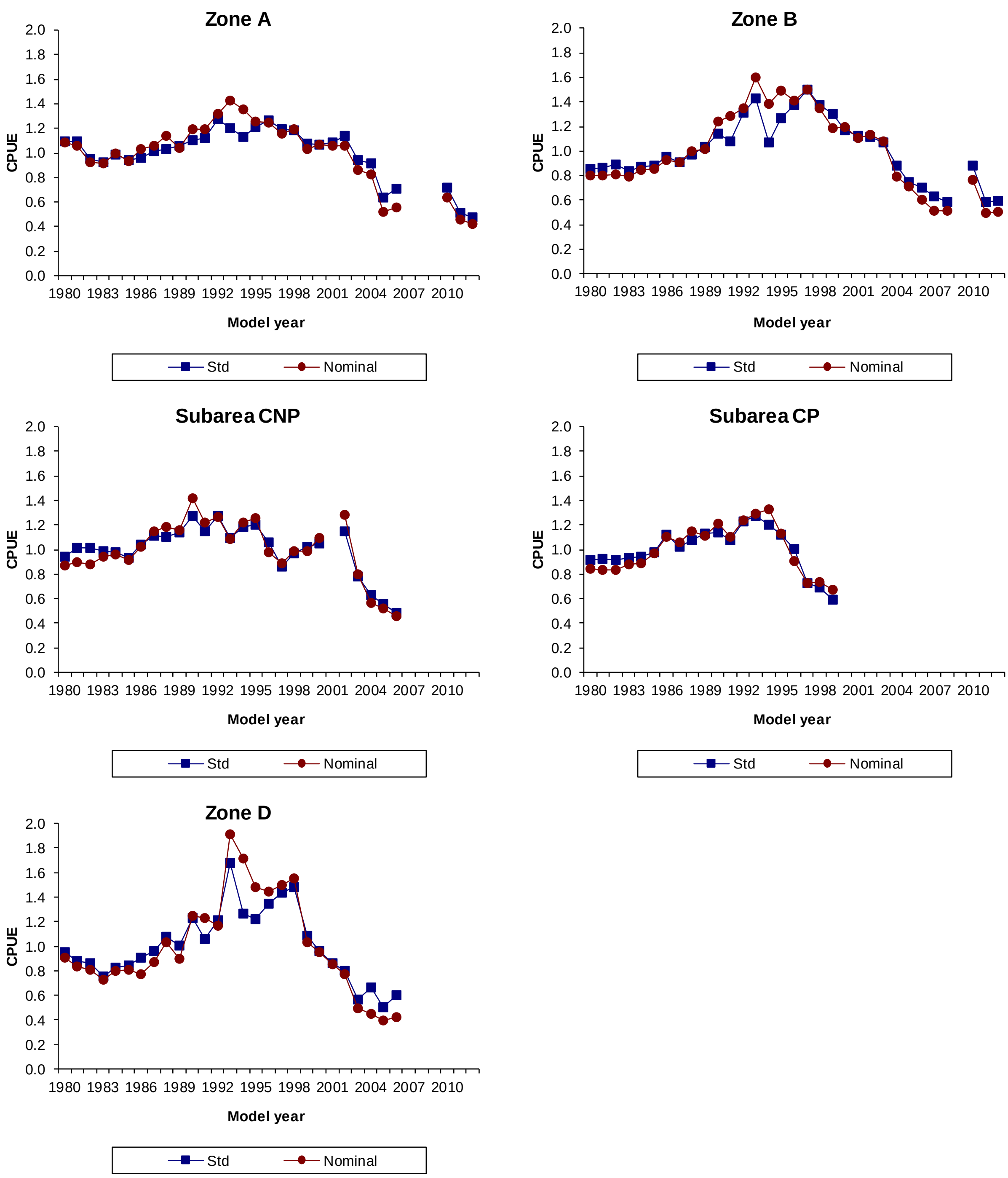

Figure 1. GLMM-standardised CPUE trends (normalised to their means over the 32 year period) for Zones/subareas A, B, CNP, CP and D. For comparison, the nominal series (also normalised to their means over that 32 year period) is also shown.