A summary of the General Linear Model analyses applied to the commercial abalone CPUE data for Zones A-D over the period 1980-2008

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

October 2009

Abstract

This paper updates earlier GLM standardisations of abalone CPUE data by Plagányi and Edwards (2007) to incorporate further data for recent years now available.

Introduction

The General Linear Model (GLM) analysis described in Plagányi and Edwards (2007) has been applied to the commercial abalone catch-per-unit-effort (CPUE) data for Zones A-D to incorporate further available data for the 2007 and 2008 Model years (where a Model year *y* runs from October of year *y*-1 to September of year *y*). The principle objective of the GLM analysis is to obtain a GLM model that incorporates important co-variates in the explanation of abalone CPUE variation to obtain standardised CPUE series that indexes resource abundance.

This paper describes the GLM model that has been used previously (Plagányi and Edwards, 2007). The method described by Plagányi and Edwards (2007) has been replicated as far as possible.

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 2008. The GLM is applied to data on a Model year basis, where a standard Model year *y* is taken to run from October of year *y*-1 to September of year *y.* Additional information in the database pertains to the dates, the divers, and

the areas and Zones that were dived. All data for Zones A, B, C and D are included in the analysis, with Zone C being split into subareas CNP (nonpoached) and CP (poached). The number of records available per Model year and by Zone is given in Table 1. A total of 41 378 data points is available for the GLM analysis. Records with a dive time less than 10 minutes were excluded as well as outliers (approximately 1% of total) based upon observations with large residuals (> 6 standard deviations) in an initial GLM fit. Years which had too few records in a Zone/subarea were also excluded (numbers shown in brackets in Table 1). A total of 40 988 data points remained for the analysis.

The General Linear Model (GLM) to standardise the CPUE

The following GLM model, which allows for possible annual differences in abalone spatial and temporal distribution, is used to standardise the commercial abalone CPUE data:

$$\ln(CPUE) = \mu + \alpha_{year} + \beta_{season} + \gamma_{zone} + \varphi_{fisher} + \eta_{year\times season} + \delta_{year\times zone} + \varepsilon \qquad (1)$$

where:

CPUE	is the catch-per-unit-effort defined as catch (kg) divided by dive time (minutes),
μ	is the intercept,
year	is a factor with 29 levels associated with the Model years 1980–2008,
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),
fisher	is a factor with 349 levels associated with the fisher code, which includes both the entitlement holders coded in the database as well as "divers". Some recent divers not yet allocated a code were given a temporary code of 555 for the purposes of this analysis,
year×*season*	is the interaction between year and season,
year×*zone*	is the interaction between year and zones/subareas, and
ε	is the error term assumed to be normally distributed.

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:

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

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 proportional to overall abundance. This averaging is unnecessary in the absence of such interactions, because then the $\exp(\alpha_{year})$ term alone will then be proportional to abundance. This averaging was not done performed in the analysis of Plagányi and Edwards (2007). Instead the standardised CPUE indices they reported corresponded to one particular season.

GLM Results and Discussion

Following an initial GLM fit, the residuals were examined to check, in particular, for evidence of heteroscedasticity. There was evidence of larger residuals associated with lower and larger effort. To account for this heteroscedasticity, the model was refit to the data using the iterative, inverse-weighting procedure applied by Plagányi and Edwards (2007), in which reduced weight is given to the data points with the largest variance in the model. To replicate this method, weighting factors were calculated by grouping the data into 12 effort categories, and estimating the mean variance of the residuals (the ε's) in each category, such that category 1 represents effort of <50 minutes dive duration, category 2 is 50-100 minutes and so on (Fig. 1). The weighting factors are simply the inverse of the variance for each category. The residuals were iteratively re-weighted until the model results converged.

The GLM model accounts for 45.7% of the total variation of abalone CPUE. Table 2 provides nominal and standardised CPUE indices provided by the model and Figures 2a–e show graphical comparisons of the same. For comparison, the standardised CPUE indices obtained by Plagányi and Edwards (2007) are also shown. Table 3 shows the parameter estimates, together with standard errors, obtained for the single factors included in the GLM model.

Figure 3 shows the comparison between nominal and standardised CPUE series for Zone A, where the standardised CPUE series where obtained using equation (2) and when using the method of Plagányi and Edwards (2007) (i.e. no averaging over seasons).

The GLM-standardised CPUE series presented in Table 2 will be used in the current 2009 abalone assessment. There are differences in the standardised CPUE abundance indices obtained in the

present analysis and those obtained by Plagányi and Edwards (2007), especially in the later years of Zone A and D. The reason for these differences are not solely because of the averaging of seasons in the standardisation procedure (see Figure 3). Further investigation is needed to determine whether the process used by Plagányi and Edwards (2007) to eliminate data from the GLM analysis has been exactly replicated in this paper.

Acknowledgements

Thanks to Angus Mackenzie and Genevieve Maharaj of MCM for assistance with queries about the data upon which this paper is based.

Reference

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 GLM analysis to standardise the commercial abalone CPUE series are shown in part a). Note the small number of entries in Zone C during the 1999 and 2000 fishing seasons. Zone CNP was closed during the 2001 fishing season and Zone CP during both the 2001, 2002 and 2003 fishing seasons as indicated by the absence of data for these years. Model years are defined as the period October to September. Note values shown in brackets indicate sample sizes which were considered too small to be included in the analysis. Part b) of the table shows the legal abalone catch per zone (in MT).

a) Number of records

	Zone/subarea				
Model year	**A**	**B**	**CNP**	**CP**	**D**
1980	257	555	73	754	535
1981	192	578	147	622	383
1982	311	610	109	594	608
1983	327	691	144	466	302
1984	334	701	303	366	373
1985	360	620	158	366	583
1986	340	765	222	446	205
1987	446	586	106	494	144
1988	457	434	96	498	147
1989	448	414	91	504	184
1990	527	410	139	458	140
1991	447	404	161	544	167
1992	349	302	98	398	142
1993	299	239	110	336	75
1994	347	290	160	288	162
1995	441	238	138	334	171
1996	514	324	413	460	206
1997	771	249	258	117	197
1998	634	509	225	75	331
1999	725	429	57	8	306
2000	456	335	24	(2)	315
2001	400	291	(1)		135
2002	290	229	103	(2)	96
2003	415	128	54		26
2004	101	576	158	(6)	69
2005	63	599	170	(4)	56
2006	42	673	160	(6)	48
2007		483	(1)		
2008		290			

b) Legal catch (in MT)

Model year	Zone/subarea				
	A	**B**	**CNP**	**CP**	**D**
1980	144.2	173.6	17.1	162.3	183.8
1981	111.0	173.8	38.5	140.7	127.3
1982	144.8	186.6	26.7	131.1	191.1
1983	158.5	200.5	37.0	105.4	78.4
1984	165.2	205.1	82.6	95.4	101.0
1985	136.8	176.4	41.6	98.7	153.8
1986	132.0	229.4	57.3	120.6	50.3
1987	171.8	166.1	29.7	126.4	45.2
1988	194.9	138.7	26.2	139.5	49.2
1989	190.9	137.3	28.6	134.8	51.0
1990	199.2	142.4	40.3	116.6	46.3
1991	182.9	138.0	41.0	119.0	49.6
1992	184.1	147.5	30.3	113.9	56.3
1993	179.2	152.1	31.9	105.2	53.7
1994	174.0	150.0	44.1	91.2	92.5
1995	210.8	152.3	39.1	84.7	90.2
1996	204.1	146.8	64.6	61.3	89.9
1997	196.8	145.9	36.7	16.1	92.6
1998	162.4	148.4	24.3	7.6	108.5
1999	191.5	155.5	11.3	1.0	103.8
2000	179.6	139.7	3.6	0.2	100.1
2001	156.5	113.4	0.2	0.0	34.1
2002	112.6	84.6	30.0	0.5	19.0
2003	119.5	36.5	5.1	0.0	1.4
2004	31.9	149.8	8.6	0.3	10.9
2005	10.1	140.3	8.6	0.2	8.5
2006	7.6	138.2	7.4	0.3	7.2
2007	0.0	72.1	0.0	0.0	0.0
2008	0.0	24.0	0.0	0.0	0.0

Table 2. Nominal and GLM-standardised commercial CPUE series for abalone for model years (October to September) 1980 to 2008 and Zones A, B, C (shown separately for subareas CNP and CP) and D. Both the nominal and the standardised values have been divided by the mean value of the respective series.

a) Nominal CPUE

	Zone/subarea				
Model year	**A**	**B**	**CNP**	**CP**	**D**
1980	1.026	0.766	0.875	0.841	0.908
1981	1.000	0.768	0.900	0.834	0.834
1982	0.878	0.777	0.884	0.834	0.804
1983	0.866	0.758	0.942	0.875	0.724
1984	0.940	0.809	0.961	0.891	0.797
1985	0.880	0.818	0.922	0.965	0.810
1986	0.977	0.910	1.026	1.104	0.771
1987	1.006	0.875	1.137	1.057	0.869
1988	1.078	0.961	1.189	1.149	1.035
1989	0.985	0.977	1.158	1.116	0.895
1990	1.130	1.191	1.422	1.215	1.247
1991	1.129	1.232	1.225	1.099	1.233
1992	1.246	1.296	1.267	1.231	1.165
1993	1.352	1.530	1.093	1.290	1.911
1994	1.283	1.331	1.211	1.321	1.711
1995	1.194	1.435	1.256	1.131	1.469
1996	1.174	1.356	0.976	0.914	1.441
1997	1.117	1.442	0.902	0.722	1.499
1998	1.167	1.324	0.997	0.739	1.550
1999	0.966	1.133	0.985	0.673	1.020
2000	1.006	1.134	1.111		0.952
2001	0.994	1.059			0.842
2002	1.006	1.086	1.225		0.761
2003	0.816	1.034	0.805		0.484
2004	0.763	0.757	0.559		0.444
2005	0.496	0.681	0.512		0.393
2006	0.523	0.575	0.456		0.434
2007		0.496			
2008		0.490			

b) Standardised CPUE

Model year	Zone/subarea				
	A	**B**	**CNP**	**CP**	**D**
1980	1.080	0.868	0.981	0.969	0.982
1981	1.023	0.829	1.015	0.924	0.875
1982	0.894	0.851	1.010	0.911	0.851
1983	0.867	0.803	0.981	0.928	0.750
1984	0.935	0.833	0.973	0.941	0.815
1985	0.882	0.841	0.940	0.966	0.828
1986	0.911	0.920	1.052	1.135	0.895
1987	0.948	0.855	1.103	1.005	0.945
1988	0.983	0.931	1.110	1.067	1.058
1989	0.998	0.993	1.152	1.119	0.990
1990	1.036	1.091	1.266	1.126	1.204
1991	1.050	1.037	1.156	1.060	1.039
1992	1.195	1.252	1.262	1.217	1.192
1993	1.110	1.358	1.075	1.294	1.646
1994	1.076	1.038	1.188	1.203	1.293
1995	1.113	1.201	1.183	1.119	1.202
1996	1.169	1.290	1.048	0.983	1.310
1997	1.118	1.442	0.843	0.722	1.412
1998	1.094	1.310	0.958	0.695	1.456
1999	1.019	1.266	1.064	0.619	1.078
2000	1.010	1.139	1.064		0.956
2001	1.069	1.136			0.879
2002	1.129	1.120	1.206		0.818
2003	0.896	1.046	0.755		0.554
2004	0.898	0.893	0.591		0.700
2005	0.701	0.786	0.544		0.579
2006	0.796	0.718	0.481		0.692
2007		0.652			
2008		0.502			

Table 3. Parameters estimates and standard errors for the single factors *year*, *season* and *zone* included in the GLM model fitted to obtain standardised indices of abundance for abalone.

Parameter	Estimate	Standard error
Year		
1980	0.000	—
1981	0.031	0.024
1982	0.040	0.024
1983	0.055	0.024
1984	0.064	0.024
1985	0.102	0.025
1986	0.080	0.025
1987	0.075	0.027
1988	0.161	0.029
1989	0.167	0.028
1990	0.399	0.029
1991	0.319	0.032
1992	0.498	0.038
1993	0.526	0.047
1994	0.383	0.040
1995	0.329	0.041
1996	0.539	0.029
1997	0.533	0.043
1998	0.585	0.026
1999	0.488	0.028
2000	0.458	0.031
2001	0.430	0.032
2002	0.397	0.033
2003	0.324	0.044
2004	0.331	0.043
2005	0.145	0.029
2006	-0.037	0.029
2007	-0.151	0.045
2008	-0.457	0.034
Season		
Jan-Mar	0.000	—
Apr-Jun	-0.022	0.020
Jul-Sep	0.117	0.020
Oct-Nov	0.249	0.061
Zone		
A	0.385	0.028
B	0.000	—
CNP	-0.058	0.047
CP	-0.061	0.021
D	0.145	0.023

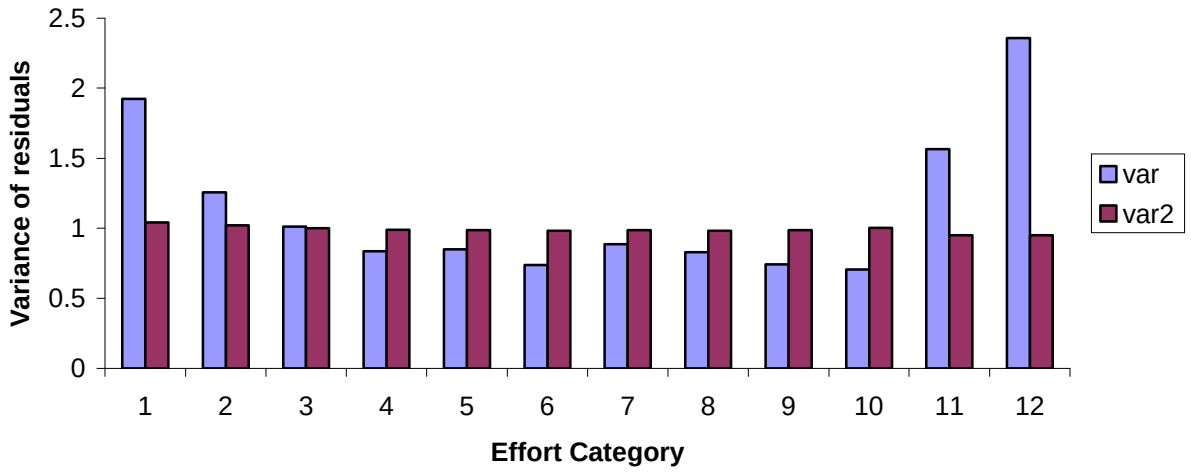

Figure 1. Plot of the average variance of the residuals (corresponding to an initial GLM fit of the data) versus effort category (50 minutes interval). To take account of this heteroscedasticity, a weighted GLM was developed as described in the text, with the final (homoscedastic) pattern of residuals shown alongside (series var 2).

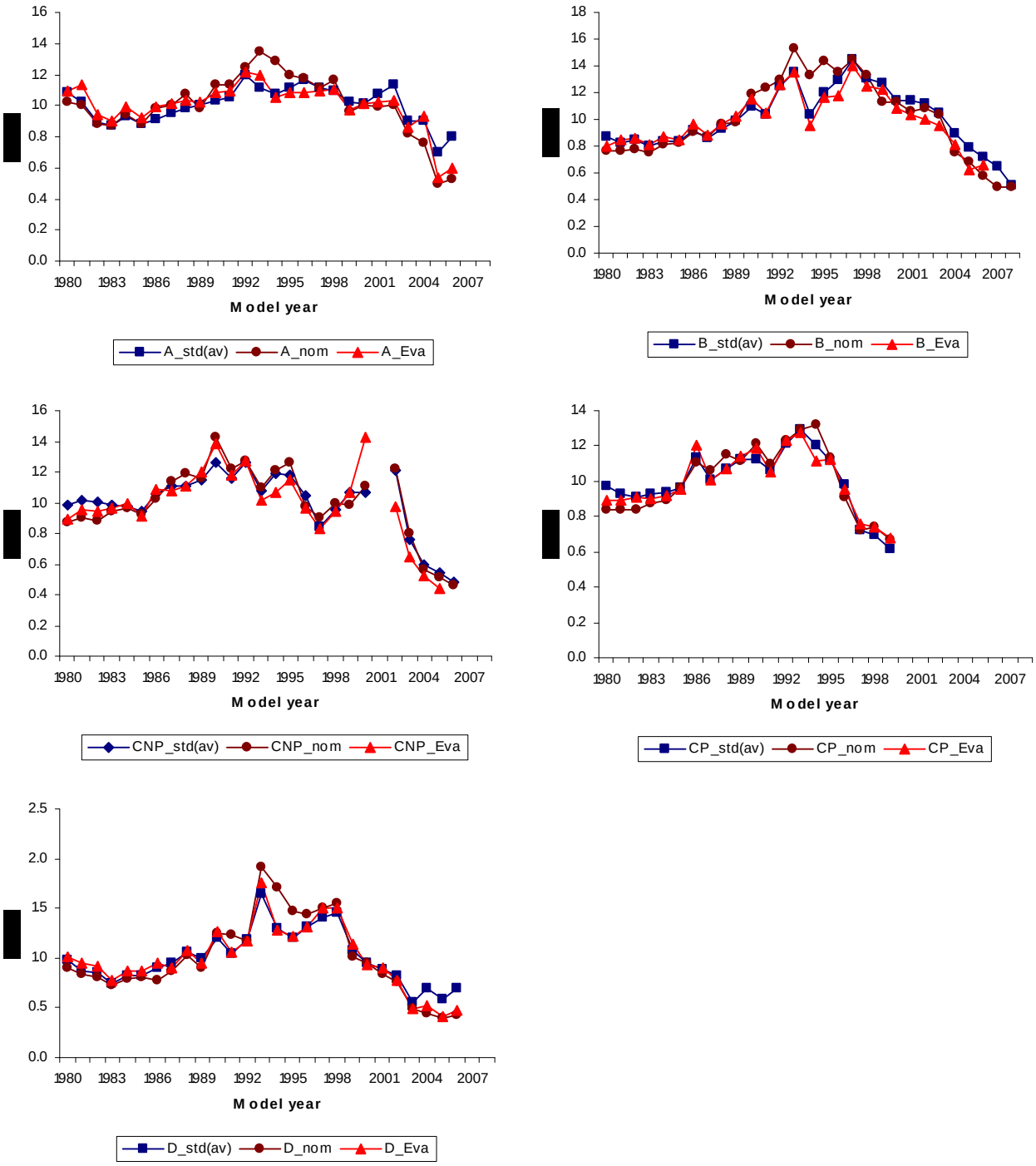

Figure 2. Comparisons between the nominal and GLM-standardised catch-per-unit-effort (CPUE) trends for zones A, B, CNP, CP and D. For ease of viewing, both the nominal and the standardised values have been divided by the mean value of the respective series. Values are derived from an iterative effort weighted GLM (see text). For comparison, the standardised CPUE series obtained by Plagányi and Edwards (2007) ("Eva") are also shown.

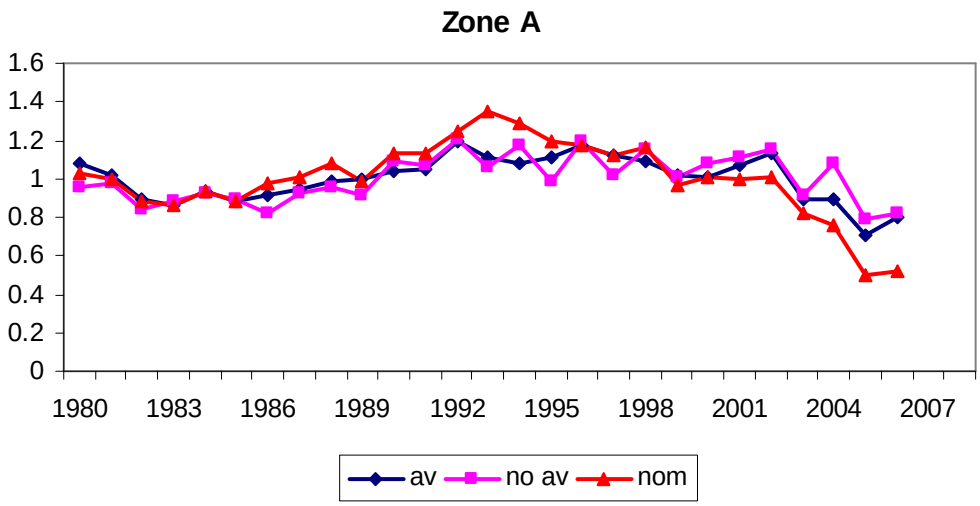

Figure 3. Comparisons between the nominal and GLM-standardised CPUE series obtained using equation (2) (av) and when no averaging over season is performed (no av) for Zone A.