The application of a Generalized Linear Mixed Model to the Area 8 bakkie data

J.P. Glazer and D. S. Butterworth

Introduction

A Generalized Linear Mixed Model (GLMM) has been applied to the Area 8 bakkie CPUE data. The intention was to include data from Areas 10, 11 and East of Hangklip (Areas 12-14) in the analyses and to include sub-area as a factor in the model, but preliminary results indicated that extending to these other Areas did not produce satisfactory results within a random-effects framework. This paper therefore reports standardized indices of abundance for the bakkie fishery from Area 8 only, including sub-area as an effect in the model.

A procedure has been adopted to adjust the standardized index to allow for the movement of lobster into the East of Hangklip area over a period of time (1987-1995). The index is then extended back to 1986 by scaling the pre-1992 indices from the model applied in the past to standardize the Area 8 CPUE data, so that they can be incorporated with the GLMM-based index.

The data

Catch and effort data are available since 1986 for Area 8, with information at a sub-area level being available since 1992 only. The GLMM analyses reported here are therefore restricted to data since 1992.

Certain general data exclusions have been applied prior to the application of the GLMM. These are as follows:

- Month=October (historically very little fishing took place in this month)
- Catch=0

The sample sizes per year and month are shown in Table 1. Given these, it would seem reasonable to include data from the months January – June in the analyses. Table 2 therefore shows the sample sizes per year and sub-area for that period. Sub-areas 4 – 6 have been omitted from the analyses due to patchiness of data over time in those particular sub-areas.

The GLMM and associated results

A model of the form shown in equation (1) was applied to the Area 8 bakkie data from 1992 onwards.

$$\ell n(\text{CPUE})=\alpha+\beta_{year}+\gamma_{\text{month}}+\eta_{sub\text{-}area}+(\text{year} \times \text{month})+(\text{year x subarea})+ \varepsilon \qquad (1)$$

α is the intercept,
year is a factor with 18 levels (1992-2009) associated with the year effect,
month is a factor with 6 levels (January-June) associated with the month effect, and
sub-area is a factor with 3 levels (subareas 1,2, and 3) associated with the sub-area effect.

Both the month and sub-area interactions with year are treated as random effects.

In order to derive an index of abundance the model is run twice; the second run excluding records where the residuals from the first run exceed ± 2SD. This methodology was adopted in order to adjust for outliers (leading to non-normality of the residuals) evident in the initial model run.

The exponent of the year factors, adjusted for movement of lobster into the East of Hangklip area, is taken to be the standardized CPUE index, i.e. $CPUE_{year} = e^{year} \times (\frac{A_{8,year}}{A_8})$. The proportion $(\frac{A_{8,year}}{A_8})$ is applied to adjust the Area 8 area size (2621 km^2) to include East of Hangklip (comprising a total area size of 161.96km^2). $A_{8,year}$ is year-specific (the Area 8 size is expanded in a linear fashion over the period 1987-1995) and A_8 is the area size of Area 8. The resultant year-specific proportions applied to the exponent of the year factors are as follows:

Year	proportion
≤1986	1
1987	1.007
1988	1.014
1989	1.021
1990	1.027
1991	1.034
1992	1.041
1993	1.048
1994	1.055
≥1995	1.062

The standardized index, together with the nominal trend, is shown in Figure 1.

The GLM-standardized index used in the past ("Revised Area 8" in Figure 4 of Glazer and Butterworth, 2011) incorporates data from 1986. A method of combining the GLMM index with that of the GLM index was considered desirable in order to extend the series as far back in time as possible. This was achieved by multiplying the pre-1992 GLM values by the ratio $\frac{\overline{Std\ CPUE_{GLMM,1992-1996}}}{\overline{Std\ CPUE_{\ GLM,1992-1996}}}$ in order to scale them to the GLMM index and then combine them with the GLMM index. The resulting combined index is reported in Table 3 and shown in Figure 2.

The GLMM fitted assumes that the random effects are homoscedastic and uncorrelated. Figures 3 and 4 show the random effects by month and by sub-area respectively. There is no obvious indication of substantial non-randomness.

The assumption of normality of the error term was investigated by examining the unstandardized residuals obtained from the GLMM fit after the exclusion of outliers. The mean, median and mode are 0, 0.04 and -0.6 respectively. The skewness and kurtosis statistics (which for a normal distribution should equal 0) are -0.4 and 0.04 respectively. Given that the median (0.04) is much less than the standard deviation of the residuals (0.48), the non-normality of the residual distribution is probably not too much of a cause for concern. The residual distribution is shown in Figure 5.

Reference

Glazer, J.P and D. S. Butterworth. 2011. Updated GLM analyses of Area 8+. Unpublished Working Group Document: Fisheries/2011/MAR/SWG-WCRL 04. 10pp.

Table 1: Sample sizes per year and month for Areas 8. Data from the shaded cells will be included in the GLMM analyses.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Nov	Dec	Total
1992	38	141	172	73	77	86				53	111	751
1993	106	158	160	163	115	65	8			46	95	916
1994	199	129	115	12	114	119	5			64	136	893
1995	66	120	125	96	14	13		18		85	56	593
1996	130	36	87	102	15		91	29		66	69	625
1997	37	69	85	41	77	55	61	35	25		48	533
1998	27	20	102	38	83	56	74	71	51		33	555
1999	54	66	58	122	104						59	463
2000	101	44	53	63	82	52	3	5			44	447
2001	26	29	87	124	258	405						929
2002	63	76	162	329	403	558	42		1	1	7	1642
2003	92	56	123	323	448	644				5	17	1708
2004	42	86	219	292	310	539	1		2	1	1	1493
2005		10	133	119	220	224						706
2006	45	96	188	138	332	291	1			8	44	1143
2007	133	161	161	227	32	143					13	870
2008	112	181	114	85	66	130				19	23	730
2009	46	132	198	85	107	49				2	36	655
Total	1317	1610	2342	2432	2857	3429	286	158	79	350	792	15652

Table 2: Sample sizes per year and sub-area for the January to June period. Data from the shaded cells will be included in the GLMM analyses.

	SA 1	SA 2	SA 3	SA 4	SA 5	SA 6	Total
1992	147	328	112				587
1993	115	422	230				767
1994	384	127	118	59			688
1995	207	186		41			434
1996	173	137	60				370
1997	148	166	44	4		2	364
1998	55	131	140				326
1999	29	6	369				404
2000	54	19	300	20		2	395
2001	625	6	283	8	1	6	929
2002	942	518	41	65		25	1591
2003	698	614	20	289	2	63	1686
2004	411	743	7	261	3	63	1488
2005	206	390	17	69	1	23	706
2006	262	523	47	206		52	1090
2007	223	228	51	304		51	857
2008	149	98	29	356		56	688
2009	97	164	62	268		26	617
Total	4925	4806	1930	1950	7	369	13987

Table 3: Bakkie standardized CPUE indices for Area 8.

Year	CPUE
1986	0.346
1987	0.525
1988	0.538
1989	
1990	0.719
1991	0.405
1992	0.705
1993	0.885
1994	1.285
1995	1.475
1996	1.235
1997	1.447
1998	1.611
1999	1.175
2000	1.369
2001	1.145
2002	1.192
2003	1.091
2004	1.038
2005	1.074
2006	0.964
2007	0.839
2008	0.893
2009	1.062

Figure 1: Area 8 standardized CPUE index. The nominal CPUE trend is also shown. Both indices have been normalized to their respective means.

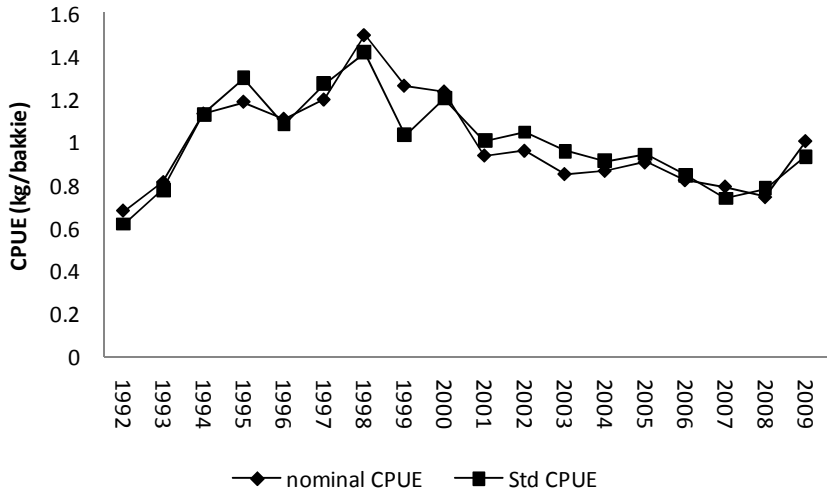

Figure 2: Area 8 standardized CPUE index extended back to 1985.

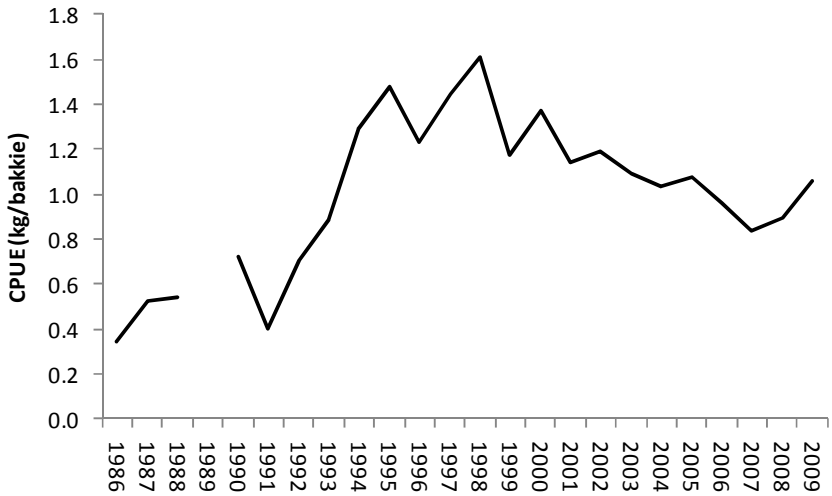

Figure 3: Random effect estimates by month obtained from the GLMM.

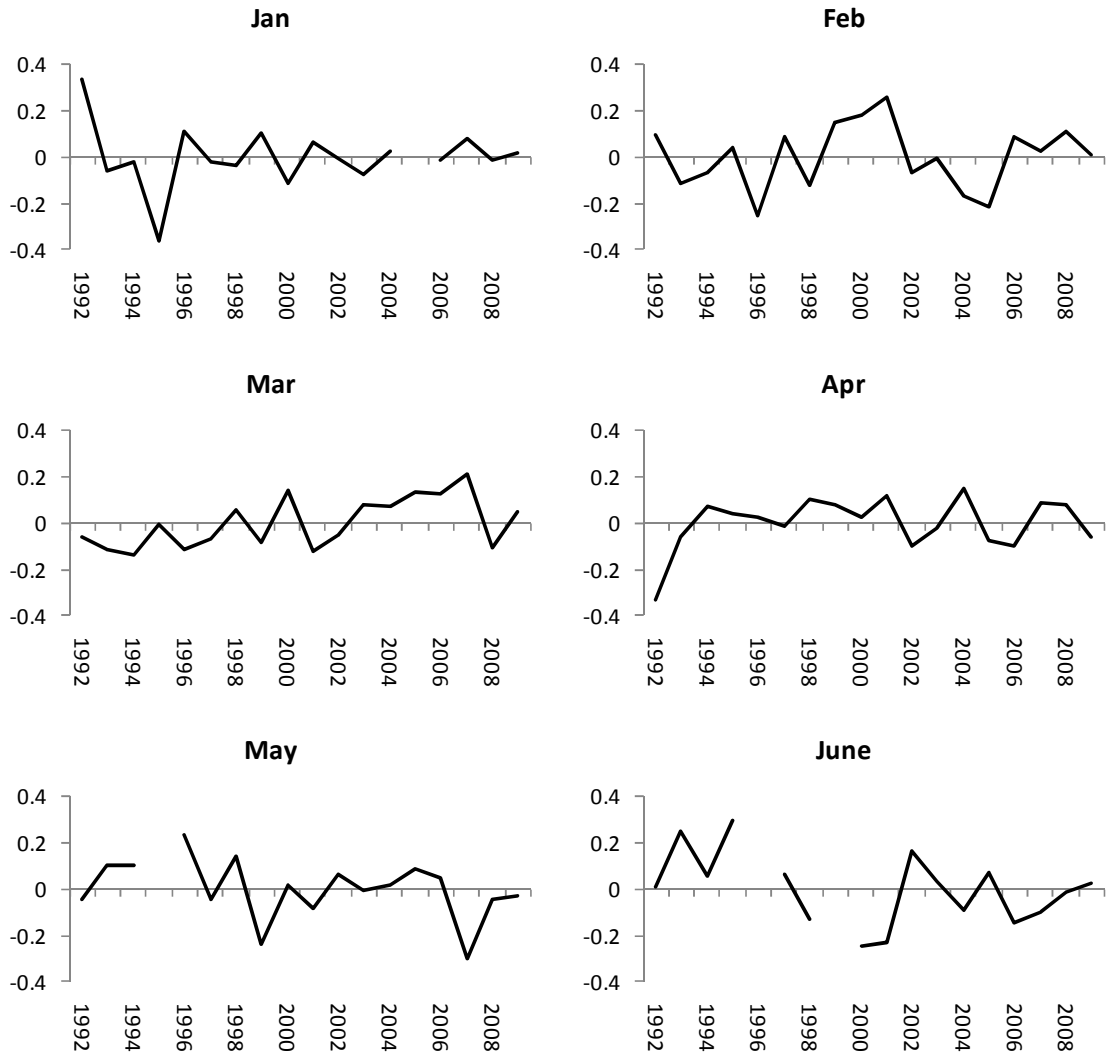

Figure 4: Random effect estimates by sub-area obtained from the GLMM.

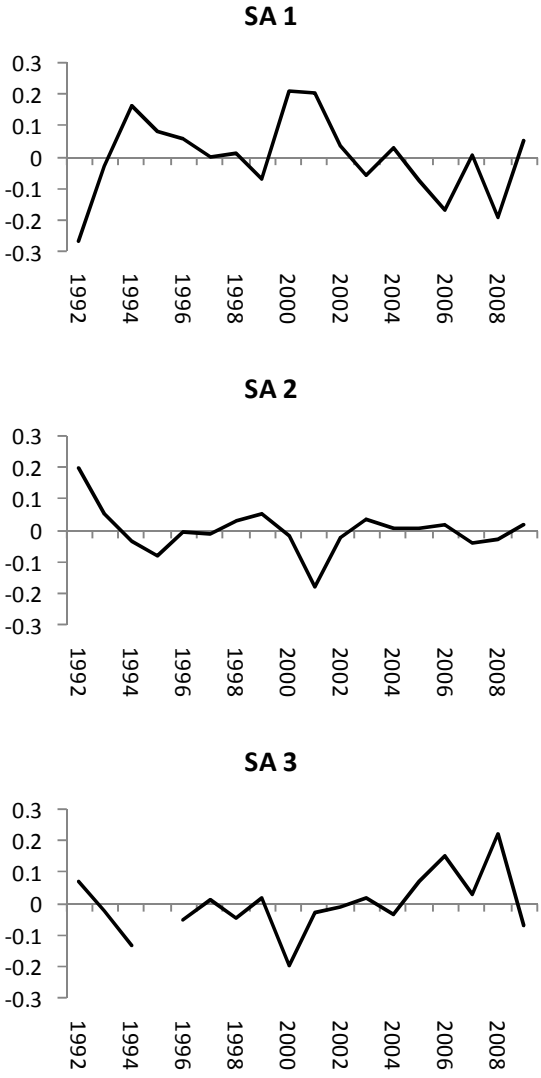

Figure 5: Distribution of unstandardized residuals obtained from the GLMM.