The inclusion of sub-areas in the standardisation of the Area 8 trapboat CPUE data through the application of a Generalized Linear Mixed Model

J.P. Glazer and D. S. Butterworth

Introduction

Data for the West Coast rock lobster Areas are available at a sub-area level from 1992. The data are fairly patchy at a sub-area level for most areas other than Area 8 for trapboats and Area 1 for bakkies. This paper presents preliminary results from an analysis of the trapboat Area 8 CPUE data where the sub-areas are included as factors in the model.

The data

Table 1 lists the sub-areas within each Area.

Tables 2-4 show the sample sizes per fishing method, Area and sub-area. From these it is considered appropriate to first attempt an analysis of the trapboat Area 8 data since there is a fair representation of data across all sub-areas within this particular Area.

Certain data exclusions are prior to analyses being conducted. These are as follows:

Data source	Exclusions
Trap-boats	• Month=October (historically very little fishing took place in this month) • Hout Bay Fishing vessels over the period 1997-2000 (catch data incorrect) • Effort (traps)=0 • Catch=0
Bakkies	• Month=October (historically very little fishing took place in this month) • Catch=0
Deck-boats	• Month=October (historically very little fishing took place in this month) • Effort (nets)=0 • Catch=0

The Models

The existing Area 8 GLM for trapboats (GLM_{BC}), which does not account for differences in CPUE trends in the different sub-areas, is of the form:

$$\ell n(\text{CPUE})=\alpha+\beta_{year}+\gamma_{month}+(\text{year}\times\text{month})+\varepsilon \qquad (1)$$

The data included in this model cover the period 1985 – 2009. In order to derive the 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 calculation of the standardized index is then limited to a subset of core months (January-June), and the indices are integrated over area to allow for an expansion of the resource into the area East of Hangklip over the period 1987-1995.

It should be noted that for all other models considered in this paper the outlier exclusion process and adjustments of the area size integrated over to include the area East of Hangklip are applied.

A variant of GLM_{BC} was run where the data included in the analysis were restricted to data from the period January to June (GLM_{BC1}). The resulting index is shown in Figure 1 together with that of the existing model. It is clear that there is very little difference between the two trends other than in 2005 where the difference can be attributed to interpolation which is required for two of the year/month interaction cells (no data for January and February of that year). Further analyses considered in this paper are therefore restricted to data from the January to June period only.

GLM_{BC2} (equivalent to GLM_{BC1}, but including sub-area as a main effect) was run for years 1992 – 2009. The standardized CPUE is determined by integrating over the sizes of the sub-areas, averaging over months and adjusting for the movement of lobster into the East of Hangklip area:

$$CPUE_{year} = [(\sum_{month=Jan}^{June} (\sum_{subarea=1}^{6} e^{(\alpha+\beta_{year}+\gamma_{month}+\eta_{subarea}+year\times month)}) \times A_{subarea})) / \sum_{month=Jan}^{June} 1\ ] \times \frac{A_{8,y}}{A_8} \qquad (2)$$

The sub-area sizes were determined from van Zyl *et al.* (2009) and are shown in Table 5. The sum of the individual subarea sizes do not add up to the total size for Area 8 (Brouwer, 2006), and these were therefore scaled so that their sum does equal that of Area 8. The proportion $\frac{A_{y,8}}{A_8}$ is applied to adjust the area size to include East of Hangklip where $A_{y,8}$ is a year-specific size (the size of Area 8 is expanded in a linear fashion over the period 1987-1995) and A8 is the actual size of Area 8 alone (2621km^2). The resulting trend, compared with that of GLM_{BC1}, is shown in Figure 2 and indicates that the inclusion of a sub-area effect in the model results in a lower CPUE in recent years than is the case if this effect is excluded from the model.

A General Linear Mixed Model (GLMM) model of the form shown in equation (3) was applied to the January – June data from 1992 onwards to allow for possible year-subarea interactions:

$$\ell n(\text{CPUE}) = \alpha + \beta_{year} + \gamma_{month} + \eta_{sub\text{-}area} + (\text{year} \times \text{month}) + (\text{year x subarea}) + \varepsilon \qquad (3)$$

Both the month and sub-area interactions with year are treated as random effects. 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. Figure 3 plots the GLMM trend together with that of the GLM_{BC2} for comparative purposes. Each index has been normalized to its mean. The resulting trends are similar and it is therefore suggested that the GLMM be adopted as the methodology applied for standardizing the Area 8 trapboat CPUE data as it avoids the need in the GLM for *ad hoc* treatments for cells without data amongst the interaction terms.

The GLMM fitted assumes that the random effects are homoscedastic and uncorrelated. Figures 4a-b and Figures 5a-b show the random effects by month and by sub-area. 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 exclusion of outliers as described above (Figure 6). The mean, median and mode are 0, 0.03 and -1.31 respectively. The skewness and kurtosis statistics (which for a normal distribution should equal 0) are -0.24 and -0.28 respectively. Given that the median (0.03) is much less than

the standard deviation of the residuals (0.50), the non-normality of the residual distribution is probably not too much of a cause for concern.

For ease of reference, Table 6 summarizes the models considered in this paper.

References

Brouwer, S. 2006. Area calculations for the South African West Coast rock lobster resource. Unpublished Fisheries Working Group Document, WG/03/06/WCRL18. 4pp.

Van Zyl, D. Auerswald, L. and D. Merkle. 2009. FIMS Area calculations, Station Numbers, Category, Repeats and Position. Unpublished Fisheries Working Group Document, MCM/2009/JUL/SWG/WCRL/04. 23pp.

Table 1: Sub-areas associated with each Area.

Area	Sub-area
1	1,2,3,4,5,6,7,8,9
2	1,2,3
3	1,2,3,4,5
4	1,2,3
5	1,2,3,4,5,10
6	1,2
7	1
8	1,2,3,4,5,6

Tables 2a and b: Sample size per year, Area and sub-area in the trap-boat fishery. No trap-boat fishing takes place in the most northerly regions (Areas 1 & 2). The shaded sub-areas likely reflect punching errors since they are not valid subareas within the Areas that they are shown to occur (as per the listings in Table 1).

a)

Year	Area 3						Area 4							Area 5						
	Sub-area						Sub-area							Sub-area						
	1	2	3	4	5	Total	1	2	3	4	5	7	Total	1	2	3	4	5	10	Total
1992	37	20	3	60	339	459	151	378	81	1			611	231	495	362	282	25	6	1401
1993	3	1			1	5	557	419	88	1		1	1066	20	100	284	57	39	1	501
1994	10	5	12	149	66	242	443	923	1272	1			2639			12	39	248		299
1995	27		14	15	3	59	197	134	370	2			703			2	41	51		94
1996	74	1		5		80	13	146	352				511			2	1	85		88
1997				4	1	5	49	20	91				160	1			95			96
1998	1	5		28		34	1	60	26	2			89				7			7
1999	6	21	57	111	5	200	1	8	1				10			1	2			3
2000	1	2	41	333		377		2					2							
2001		3	8	172	1	184	1	3		4			8							
2002				1		1		28					28							
2003		5	16	27	3	51		1					1							
2004	19	2	5	36		62		1					1							
2005								17	6				23							
2006				1		1														
2007	28	14	2	33		77	3	18					21							
2008	17	26	88	145		276														
2009	25	15	88	76		204		67					67							
Total	248	120	334	1196	419	2317	1416	2225	2287	7	4	1	5940	252	595	663	524	448	7	2489

b)

Year	Area 6							Area 7								Area 8							
	Sub-area							Sub-area								Sub-area							
	1	2	3	4	5	6	Total	1	2	4	5	6	7	10	Total	1	2	3	4	5	6	7	Total
1992	554	630	3	2			1189	923						1	924	252	594		285	41	78		1250
1993	347	618					965	807						1	808	388	413	18	327	118	88		1352
1994	118	393					511	1276	2						1278	555	549	13	220	87	39	17	1480
1995	143	140					283	756				1			757	655	360		110	46	14		1185
1996	229	56					285	636				4			640	606	447	38	360	40	90		1581
1997	137	117				1	255	1197							1197	534	641	22	157	151	77		1582
1998	75	16				2	93	986							986	261	754	83	318	46	236		1698
1999	21	23			8		52	1020							1020	351	580	47	458	25	73		1534
2000	11	1					12	1053		1					1054	572	218	63	236	60	181		1330
2001	10	14					24	1080							1080	609	628	17	251	245	530		2280
2002								1796			1		1		1798	527	339		528	173	474		2041
2003	1	2					3	1941					1		1942	776	480		273	448	477		2454
2004	2	2					4	2404							2404	705	336		275	393	1218		2927
2005								2239	6				1		2246	124	425		705	104	841		2199
2006								2521							2521	211	321		1231	51	2004		3818
2007								1463							1463	268	575	3	659	525	737		2767
2008								1427							1427	155	445	1	718	355	613		2287
2009								987							987	229	629	1	412	7	85		1363
Total	1648	2012	3	2	8	3	3676	24512	8	1	1	5	3	2	24532	7778	8734	306	7523	2915	7855	17	35128

Tables 3a and b: Sample size per year, Area and sub-area in the bakkie fishery. The shaded sub-areas likely reflect punching errors since they are not valid subareas within the Areas that they are shown to occur (as per the listings in Table 1).

a)

Year	Area 1										Area 2			
	Sub-area										Sub-area			
	1	2	3	4	5	6	7	8	9	Total	1	2	3	Total
1992	11	312	60	333	47	417	9	38	3	1230				
1993	8	354	32	216	9	696	40	95	12	1462				
1994	13	155	59	418	5	423	52	84	1	1210				
1995	50	99	21	129	8	396	55	55	10	823				
1996		18	10	170	11	495	51	76		831				
1997	14	93	45	286		541	53	76	1	1109				
1998		14	5	36		189	18	26	4	292				
1999	1	8	5	50		365	43	42	26	540				
2000		17	12	59		562	59	108	5	822				
2001		4	3	55	1	381	30	52	18	544	66	1		67
2002	2	8		164	2	615	31	42	13	877	2			2
2003	12	49	12	137	14	690	59	85	52	1110				
2004	10	36	34	195	11	584	87	101	48	1106				
2005	9	41	2	51	2	437	60	75	13	690		1		1
2006	1	34	1	133	3	501	41	68	1	783	13	1		14
2007	6	56	8	34	1	418	50	64	50	687	24	8		32
2008	1	30	2	38	1	483	56	75	12	698	23			23
2009	1	6	3	55		279	37	41	5	427	23			23
Total	139	1334	314	2559	115	8472	831	1203	274	15241	151	9	2	162

b)

Year	Area 3						Area 4						Area 5						
	Sub-area						Sub-area						Sub-area						
	1	2	3	4	5	Total	1	2	3	4	5	Total	1	2	3	4	5	10	Total
1992	2	40	306	3	6	357	68	1894	160			2122	5	1	535	276	88		905
1993	2	43	22	1	7	75	207	2511	106	2		2826	23	3	875	111	349		1361
1994	4	147	330	1		482	49	4268	82	4		4403	25		728	12			765
1995		9	43			52	25	1579	83	1		1688	1		409	2			412
1996		5	1			6		2083	44			2127			523	15	75		613
1997				1		1	108	1537	32			1677		1	515	145			661
1998							13	971	6			990	41		195	33			269
1999		1		1		2		1783				1783			443	110	6		559
2000				33		33		1195	2			1197		2	333	11	2		348
2001	2	78	11	75		166	1	638	15			654	21		8		4		33
2002		506	83	461	38	1088	328	918	31	3	3	1283	3		237	91			331
2003	16	312	4	467		799	86	554	76	1		717			263	65			328
2004	125	345	45	100		615	22	1000	226	1		1249	1		306	56			363
2005	96	397	51	4		548		1024	2			1026		17	14			2	33
2006	15	144	160	13	62	394	6	578				584			48				48
2007	254	379	72	4	11	720	3	581	8			592	4	1	111			230	346
2008	35	27	42	27		131	7	744	41			792	14	1	59	6	11	115	206
2009	11	20	10	11		52	2	869	29			900	212		55		2		269
Total	562	2453	1180	1202	124	5521	925	24727	943	10	5	26610	350	26	5657	933	537	347	7850

Table 3c: Sample size per year, Area and sub-area in the bakkie fishery. The shaded sub-areas likely reflect punching errors since they are not valid subareas within the Areas that they are shown to occur (as per the listings in Table 1).

Year	Area 6							Area 7		Area 8							
	Sub-area							Sub-area		Sub-area							
	1	2	3	4	5	6	Total	1	Tota	1	2	3	4	5	6	8	Total
1992	672	210		4	2		888	2	2	164	475	112					751
1993	203	83			2		288			128	536	252					916
1994	31	2			1		34			462	227	145	59				893
1995	8	4					12	2	2	254	298		41				593
1996	48	1					49			309	138	172	5		1		625
1997										156	205	166	4		2		533
1998										63	171	321					555
1999		4					4	3	3	31	6	426					463
2000								1	1	58	19	344	20		6		447
2001	264	51				1	316			625	6	283	8	1	6	6	935
2002	844	94					938			961	546	41	68		26		1642
2003	592	62			11		665			713	619	20	291	2	63	1	1709
2004	442	21				1	464			412	744	7	261	3	66		1493
2005	250	25					275			206	390	17	69	1	23		706
2006	250	165					415			313	525	47	206		52		1143
2007	152	2	3				157			227	233	51	306		53		870
2008	160	2					162			182	99	30	363		56		730
2009	77						77			112	175	64	276		28		655
Total	3993	726	3	4	16	2	4744	8	8	5376	5412	2498	1977	7	382	7	15659

Tables 4a and b: Sample size per year, Area and sub-area in the deck-boat fishery. The shaded sub-areas likely reflect punching errors since they are not valid subareas within the Areas that they are shown to occur (as per the listings in Table 1).

a)

Year	Area 1								Area 2				
	sub-area								sub-area				
	1	2	3	4	6	9	10	Total	1	2	3	4	Total
1992	3	1			26			30	50		4		54
1993	8	1	1	1	39	6		56	17		1	1	19
1994	16			8	55	4		83	3		5		8
1995	3	15			34		1	53					
1996	6	1			27			34					
1997	3	4		2	20	3		32			2		2
1998										2	10		12
1999											1		1
Total	39	22	1	11	201	13	1	288	70	2	23	1	96

b)

Year	Area 3							Area 4						Area 5					Area 6				Area 7		Area 8			
	sub-area							sub-area						sub-area					sub-area				sub-area		sub-area			
	1	2	3	4	5	7	Total	1	2	3	4	8	Total	1	2	4	5	Total	1	2	5	Total	1	Total	1	2	4	Total
1992	112	17	1	2	1	1	134		30	23			53	6		34	7	47	145	67	1	213						
1993	6						6	2	42	63		1	108	6	1	1	16	24	99	69	1	169						
1994	76	12			1		89		121	16	1		138	31	1	1	22	55	124	34	1	159	74	74		6		6
1995	17						17	7	25	44			76										9	9	7			7
1996	18	4					22		91	4			95										3	3				
1997								1	98	9			108															
1998	3	9	4	16	2		34	6	62	18	4		90										67	67		4		4
1999	6	21		5	3		35	4	8				12										17	17				
2000		2	1	3			6		4	8			12										33	33				
2001		1					1	1	14				15															
Total	238	66	6	26	7	1	344	21	495	185	5	1	707	43	2	36	45	126	368	170	3	541	203	203	7	6	4	17

Table 5: Sizes of the sub-areas in Area 8 based on block number sizes as detailed in van Zyl *et al.* (2009). The sub-area sizes have been scaled so that their sum equals the official size of Area 8 (2621km^2) as detailed in Brouwer (2006).

block numbers (van Zyl *et al*., 2009)	Sub-area	Size (km^2) 200 m contour	Scaled size (km^2)
8.1-8.10	1	190.85	192.32
8.11-8.21	2	186.93	188.37
8.22-8.31	3	159.70	160.93
8.32-8.47	4	364.18	366.98
8.48-8.82	5	1495.63	1507.12
8.83-8.86	6	203.63	205.19
	Total	2601	2621

Table 6: Summary of the Models considered in this paper.

Name	Type	Model specification and standardization equation	Period analysed	Notes
GLM_{BC}	GLM	$\ell n(\text{CPUE})=\alpha+\beta_{year}+\gamma_{month}+(\text{year}\times\text{month})+\varepsilon$ $CPUE_{year} = (\sum_{month=Jan}^{June} e^{(\alpha+\beta_{year}+\gamma_{month}+year\times month)})\times A_8 / \sum_{month=Jan}^{June} 1$	1985-2009	All data utilized in analysis. Standardization restricted to core months (January-June). Outliers excluded. Area size adjusted to include EOH.
GLM_{BC1}	GLM	$\ell n(\text{CPUE})=\alpha+\beta_{year}+\gamma_{month}+(\text{year}\times\text{month})+\varepsilon$ $CPUE_{year} = (\sum_{month=Jan}^{June} e^{(\alpha+\beta_{year}+\gamma_{month}+year\times month)})\times A_8 / \sum_{month=Jan}^{June} 1$	1985-2009	Analysis restricted to data for the January-June period. Outliers excluded. Area size adjusted to include EOH.
GLM_{BC2}	GLM	$\ell n(\text{CPUE})=\alpha+\beta_{year}+\gamma_{month}+\eta_{subarea}+(\text{year}\times\text{month})+\varepsilon$ $CPUE_{year} = [(\sum_{month=Jan}^{June} (\sum_{subarea=1}^{6} e^{(\alpha+\beta_{year}+\gamma_{month}+\eta_{subarea}+year\times month)})\times A_{subarea})) / \sum_{month=Jan}^{June} 1\ ]\times \frac{A_{y,8}}{A_8}$	1992-2009	Analysis restricted to data for the January–June period. Outliers excluded. Area size adjusted to include EOH. 1985-1991 data excluded Subarea factor added
GLMM	GLMM	$\ell n(\text{CPUE})=\alpha+\beta_{year}+\gamma_{month}+\eta_{subarea}+(\text{year}\times\text{month})+(\text{year}\times\text{subarea})+\varepsilon$ $CPUE_{year} = e^{\text{year}}\times\frac{A_{y,8}}{A_8}$	1992-2009	Analysis Restricted to data for the January–June period. Outliers excluded. Area size adjusted to include EOH. 1985-1991 data excluded Subarea factor and interactions added

Figure 1: Area 8 standardized CPUE indices as derived from GLMs including data from i) all months (GLM_{BC}) and ii) January – June only (GLM_{BC1}). Each index has been normalized to its mean.

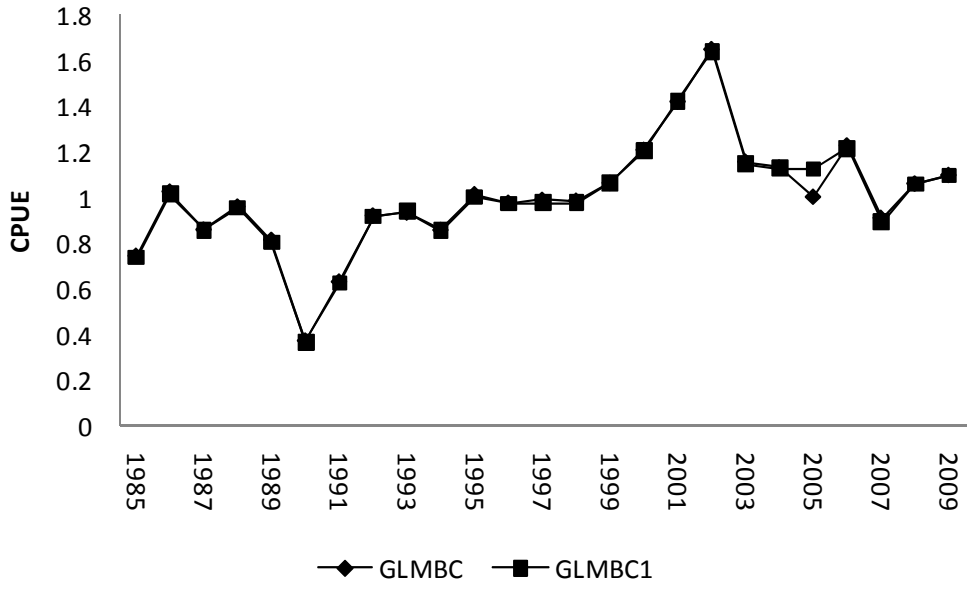


Figure 2: GLM_{BC1} (year+month+year×month) *vs* GLM_{BC2} (year+month+subarea+year×month). Each index has been normalized to its mean over the common period (1992-2009).

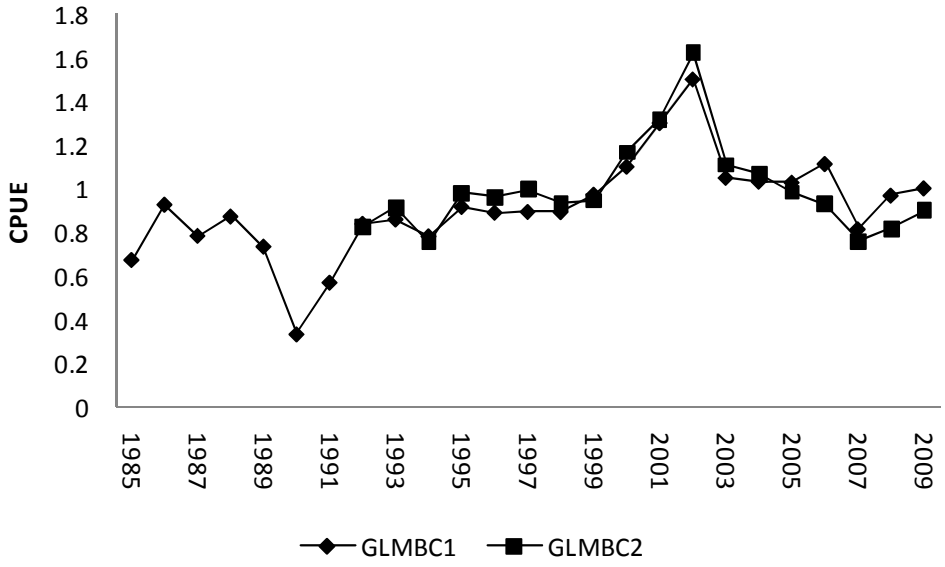


Figure 3: GLM_{BC2} *vs* GLMM standardized indices of abundance. Each index has been normalized to its mean over the common period (1992-2009).

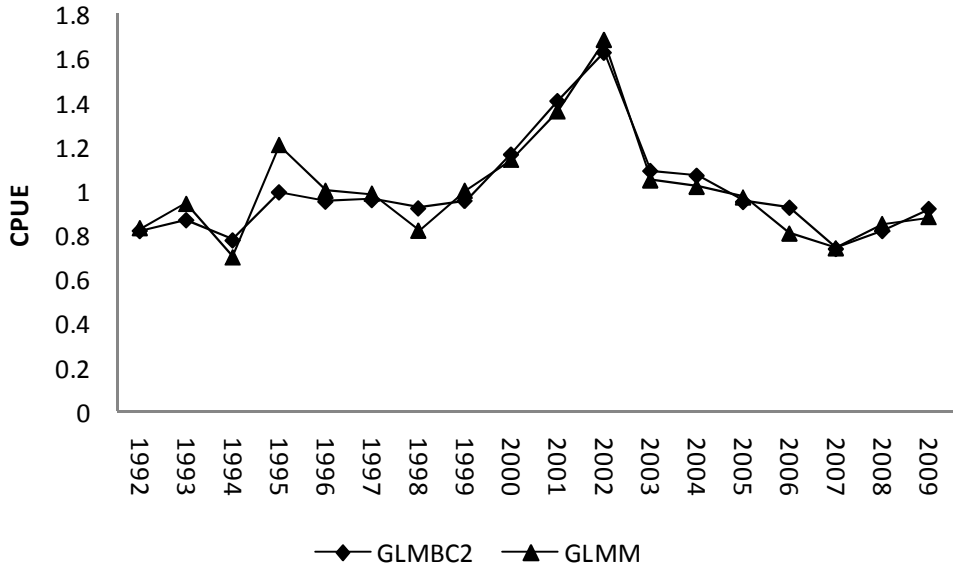
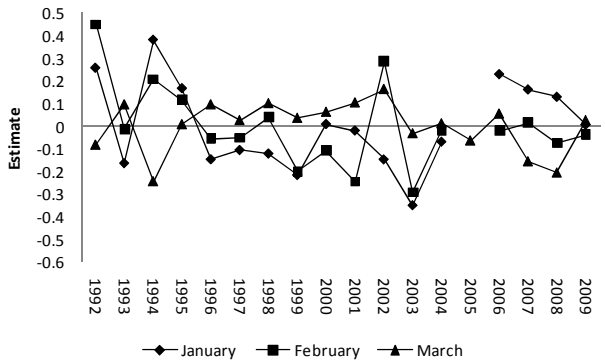


Figure4 a-b: Random effect estimates by month obtained from the GLMM.



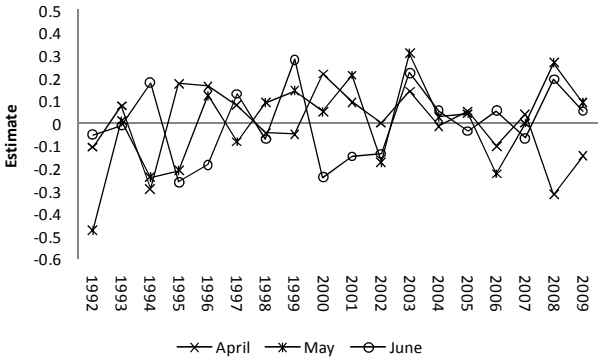


Figure 5a-b: Random effect estimates by sub-area obtained from the GLMM.

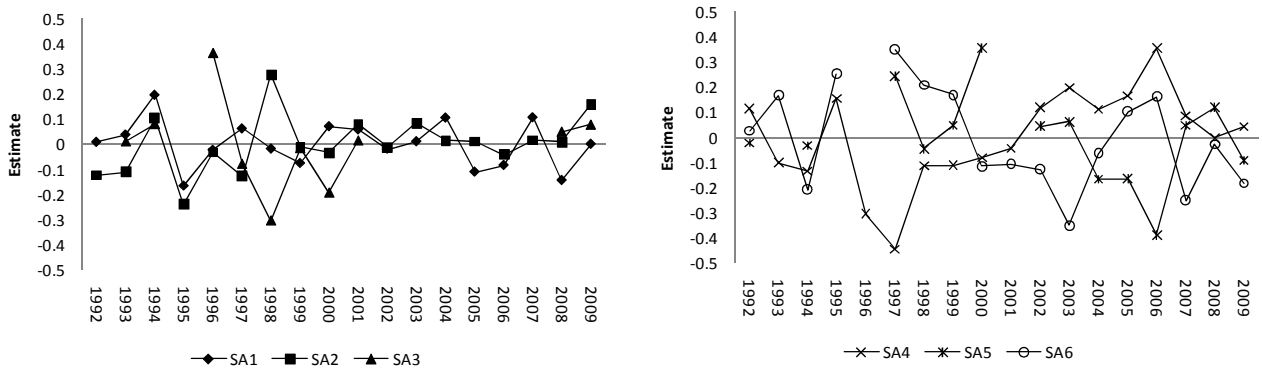


Figure 6: Distribution of unstandardized residuals obtained from the GLMM.