Standardised longline lobster CPUE from Inaccessible Island for the 1996-2005 period.

S.J. Johnston, A. Brandão, C.T.T. Edwards and D.S. Butterworth.

MARAM
Department of Mathematics and Applied Mathematics
University of Cape Town
Rondebosch, 7701

May 2008

Introduction

The commercial CPUE series of a resource is often used as an index of population density and consequently population abundance when modelling the dynamics of the underlying population. It is known, however, that a number of other factors besides density may influence the recorded values of CPUE. Where sufficient data exist, General Linear Mixed Model (GLMM) standardisation is able to take some of these further effects into account, thereby producing a more reliable index of abundance. This document reports the application of a GLMM standardisation to *Jasus tristiani* lobster catch and effort data from around Inaccessible Island for the period 1996-2005.

Methodology

<u>Data</u>

Raw Logsheet data
The logsheet data for Inaccessible Island have been entered electronically into an EXCEL spreadsheet. Logsheet data from the fishery are available for the Season-Years between 1996 and 2005, where a Season-Year is taken to run from May until April the following year, i.e. Season-year 2005 refers to the period from May 2005 to April 2006. Unfortunately logsheet data for 2006 have been misplaced (James Glass, pers. comm.). Logsheet data are also incomplete for Season-Years 1996 and 1999 (Edwards and Glass, 2007).

Summary sheet data
Data summary sheets recorded by the Agriculture and Natural Resources Department on Tristan da Cunha are available from Season-Year 1996 to 2006. These contain summary data from both the logsheets (total catch and total effort) and factory reports (Edwards, 2007). It should be noted that the Summary sheet data available for 1996 do not overlap with the logsheet data for that year (i.e. the data from the two sources refer to different trips).

<u>Accounting for inaccurate records</u>

Although logsheet data are valuable data as they record details of the catches, e.g. location and soak-time which are needed for standardisation, the logsheet entries are known to be inaccurate (Edwards, 2007). In particular, longline catch and powerboat effort are unreliable. Furthermore there is currently insufficient information concerning the different catch rates for longline monster and powerboat traps, thereby precluding the standardisation of the catch rate across different types of fishing. All powerboat data were therefore excluded from the analyses presented here.

Because of inaccurate longline catch records, the total logsheet catch for each year differ from the actual catch taken. A more accurate (best) estimate of the total longline catch in year *y* (C_y) is provided by subtracting the total powerboat catch from the total packed weight (both recorded on the Summary sheets), where the packed weight is scaled upwards to account for weight lost during processing (Edwards, 2007). This catch estimate can then be used to adjust the longline catch records so that the total catches from both sources are equal. Unfortunately there are logsheets missing for some years. An adjustment coefficient k_y was therefore developed using the ratio of total recorded effort for the summary sheets and logsheets, to scale adjustments.

Adjusted logsheet catches were calculated as follows:

$$c_{i,y} \rightarrow c^*_{i,y} = c_{i,y} k_y = c_{i,y} \frac{C_y}{C_y^{LS} \dfrac{E_y^{SS}}{E_y^{LS}}} \qquad (1)$$

where

$c_{i,y}$ is the *i*'th logsheet longline catch record for Season-Year *y*,

C_y^{LS} is the total logsheet longline catch for Season-Year *y*,

C_y is the best estimate of the total longline catch for Season-Year *y* (based on summary sheets),

E_y^{LS} is the total logsheet longline effort for Season-Year *y*, and

E_y^{SS} is the total Summary sheet longline effort for Season-year *y*.

Adjusted catches were then used to calculate Adjusted CPUE values (I_y^*) for each Season-year:

$$I_y = \frac{1}{n_y} \sum_i \frac{c_{i,y}}{e_{i,y}} \rightarrow \frac{1}{n_y} \sum_i \frac{c^*_{i,y}}{e_{i,y}} = I_y^* \qquad (2)$$

where

I_y is the nominal CPUE for Season-Year *y*,

$e_{i,y}$ is the *i*'th logsheet longline effort record for Season-year *y*, and

n_y is the number of logsheet records for Season-year *y*.

For the 1996 Season-Year, the only logsheets available are from a period of experimental fishing. There are no factory records that correspond to these logsheet data and therefore it is impossible to estimate k_{1996} as described above. The value of k_{1996} is thus set equal to the average value of the *k* values for 1997-2001, which results in k_{1996}=0.90.

Other data manipulations and filters

The raw data were filtered as follows:

- Remove all records with a "NA" or "0" in a critical field e.g. zero effort records and "NA" catch or area records.
- Remove any very high nominal CPUE values (>20 kg/trap)

Appendix 1 provides more detailed information on the final input data to be used for the GLMM analysis for each category of data.

The General Linear Mixed Model

A GLMM which includes both fixed and random effects is used to standardise the lobster CPUE data, where catches are the adjusted logsheet catches of Equation (1) and effort is logsheet effort. (Note that this approach assumes that the logsheet data represent an unbiased sample of all the fishery in each Season-year.). This model allows for possible annual differences in the areal distribution of the lobsters (which is considered to be a fixed effect) and for annual differences in each month (considered as a random effect). This model is given by:

$$\ln(CPUE + \delta) = \mathbf{X}\alpha + \mathbf{Z}\beta + \varepsilon \qquad (3)$$

where:

α	is the unknown vector of fixed effects parameters (in this case this consists of the factors given by equation (4) below),
$\mathbf{X}$	is the design matrix for the fixed effects,
β	is the unknown vector of random effects parameters (which in this application consists of a year-month interaction for reasons explained below),
$\mathbf{Z}$	is the design matrix for the random effects,
δ	is a small constant added to the rock lobster CPUE to allow for the occurrence of zero CPUE values (0.1 kg/trap in this case), and
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 E(ln($CPUE+\delta$)) = $\mathbf{X}\alpha$. 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 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:

$$\mathrm{Cov}(Incr) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^{\mathrm{T}} + \mathbf{R},$$

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

The sum of the factors that are considered as fixed effects (i.e. $\mathbf{X}\alpha$ in equation (1)) in the GLMM is given by:

$$\mu + \phi_{year} + \zeta_{month} + \gamma_{area} + \eta_{trap\ type} + \lambda_{soak\ time} + \theta_{depth} + \tau_{year\times area} \tag{4}$$

where:

μ	is the intercept,
year	is a factor with 10 levels associated with the years (i.e. the Season-Years: 1996-2005),
month	is a factor with 7 levels associated with the fishing month (1, 2, 8, 9, 10, 11 or 12),
area	is a factor with 4 levels associated with groupings of fishing areas (i.e. level 1: area 1, level 2: areas 2 and 9, level 3: area 6, and level 4: remaining areas from 1-9),
trap type	is a factor with 3 levels associated with the trap type (beehive, monster and plastic pots),
soak time	is a factor with 3 levels associated with the soak time period ("0.25–0.49" days, "0.5–1.9" days and "2+" for 2 or more days),
depth	is a factor with 4 levels associated with fishing depth ranges ("10–-" for depths < 10m, "10–39.9"m, "40–89.9"m, and "90+" for depths ≥ 90 m), and
year×area	is the interaction between year and area.

The categories used for area, soak time and depth were determined by final analysing of the data using a finer categorisation, and then continuing categories for which the estimate proved very similar.

For this model, because of the fixed effect interaction of area with year (which implies changing spatio-temporal distribution patterns), an index of overall abundance needs to integrate the different trends in density in each area (see Figure 1) over the size of these areas. Accordingly the standardised CPUE series is obtained from:

$$CPUE_{year} = \left[\sum_{area}\left(\left(\exp\left(\mu + \phi_{year} + \gamma_{area} + \tau_{year\times area}\right) - \delta\right) * A_{area}\right)\right] \Big/ A_{total} \tag{5}$$

where:

A_{area} is the surface size of the area concerned,

A_{total} is the total size of the fishing ground considered (the division by A_{total} is to keep the units and size of the standardised CPUE index comparable with those of the nominal CPUE).

In this application the CPUE has been standardised on the month of *September*, trap type *Monster*, soak time "0.5-1.9" days, and depth of "40-80"m.

Often models with interaction terms have missing cells for certain combinations of levels of factors. To be able to compute equation (5) for standardising the CPUE, the missing cells were replaced by the average of the estimable factors for the pre- and post-years of the missing year and the same area. If a missing cell occurs when there is no pre- or post-year, the average of all estimable factors for that area is used. Year-month interaction terms are significant, but their inclusion as fixed effects would have resulted in too many missing cells; this is why they were treated as a random effect.

The sizes of the areas for each of the nine fishing areas are given in Table 1.

Results

A total of 3920 records were included in the analysis. Table 2 provides standardised CPUE values derived from the GLMM considered, with more detailed results reported in Appendix 2. For comparison, the Adjusted Nominal CPUE values are also shown. Figure 2 shows the indices of abundance provided by the random effects model. These are compared to the Adjusted Nominal CPUE values. Both indices show upward trends, though those obtained from the GLMM standardisation reflect a greater increase over the whole period (Table 3 and Figure 3).

Rate of Increase

The average rate of increase over the period considered (1996 to 2005) can be calculated by the log linear regression of the values in Table 2. These results are shown both in Table 3 and Figure 3. Results show that the average rate of increase is about 16% using the Adjusted Nominal CPUE series, and 23% using the GLMM CPUE series. The reason for the difference is that the rate of increase in CPUE in area 1 is greater than in the other areas (see Figure 1), and the size of this area is much greater that the other areas (Table 1).

Conclusion

From the analyses of this paper, the catch and GLMM standardised CPUE series shown in Table 4 are put forward as the best upon which to base assessment of the resource.

References

Edwards, C.T.T. 2007. Sources of data from the lobster fisheries on Inaccessible, Nightingale, Gough and Tristan da Cunha. Technical Report MARAM/Tristan/07/Dec/05, Ovenstone Fisheries.

Edwards, C.T.T. and Glass, J.P. 2007. Reconciliation of data from the lobster fisheries on Inaccessible, Nightingale, Gough and Tristan da Cunha. Technical Report MARAM/Tristan/07/Dec/06, Ovenstone Fisheries.

Table 1: The size (km^2) of each fishing area (see Figure A1.3).

Area	Name	Size
1	Bank	53.58
2	North point	5.88
3	Salt beach	1.10
4	East Point	10.14
5	Toms beach and Black spot	3.60
6	South Hill	3.60
7	Pyramid rock and Blinder	5.23
8	West point	5.04
9	Blendon Hall	4.32

Table 2: Standardised longline CPUE series for Inaccessible Island using the GLMM model detailed in the text. The number of data points for each year (*n*) is shown alongside the scaling coefficient (*k*) used to estimate the Adjusted CPUE (see Equations 1 and 2) from the Nominal CPUE.

Season-Year	*N*	*k*	CPUE		
			Nominal	**Adjusted**	**Standardised**
1996	115	0.90	1.82	1.64	1.35
1997	227	0.91	2.75	2.51	0.92
1998	726	0.98	2.39	2.36	1.64
1999	360	0.76	3.58	2.71	2.30
2000	406	0.87	2.95	2.57	2.19
2001	545	0.98	3.30	3.24	2.69
2002	419	1.07	4.29	4.58	4.74
2003	243	0.85	6.34	5.41	5.05
2004	415	0.91	7.10	6.44	8.22
2005	464	1.01	6.82	6.92	6.60

Table 3: The average annual rate of increase as estimated over the data period (1996-2005) for the Nominal, Adjusted Nominal and GLMM standardised CPUE series.

CPUE series	**Mean annual rate of increase**	**95% CI**
Nominal	14.6%	[10.5% - 18.6%]
Adjusted Nominal	15.5 %	[12.2% - 18.7%]
GLMM standardised	22.8%	[17.2% - 28.4%]

Table 4: The catch and GLMM standardised CPUE series to be used for assessment purposes.

	Total Catch (kg)	GLMM standardised CPUE
1970	80000	
1971	147000	
1972	116000	
1973	214000	
1974	282000	
1975	133000	
1976	224000	
1977	138000	
1978	123000	
1979	141000	
1980	74000	
1981	115000	
1982	92000	
1983	72000	
1984	77000	
1985	90000	
1986	62000	
1987	81000	
1988	72000	
1989	67000	
1990	78781	
1991	56552	
1992	71625	
1993	59886	
1994	61586	
1995	61465	
1996	73306	1.35
1997	62521	0.92
1998	61492	1.64
1999	64176	2.3
2000	66637	2.19
2001	70512	2.69
2002	70775	4.74
2003	77283	5.05
2004	84484	8.22
2005	92945	6.6
2006	103281	
2007	100000	

Figure 1: CPUE trends in different fishing areas (obtained from the GLMM which includes *year x area* as a fixed effect)

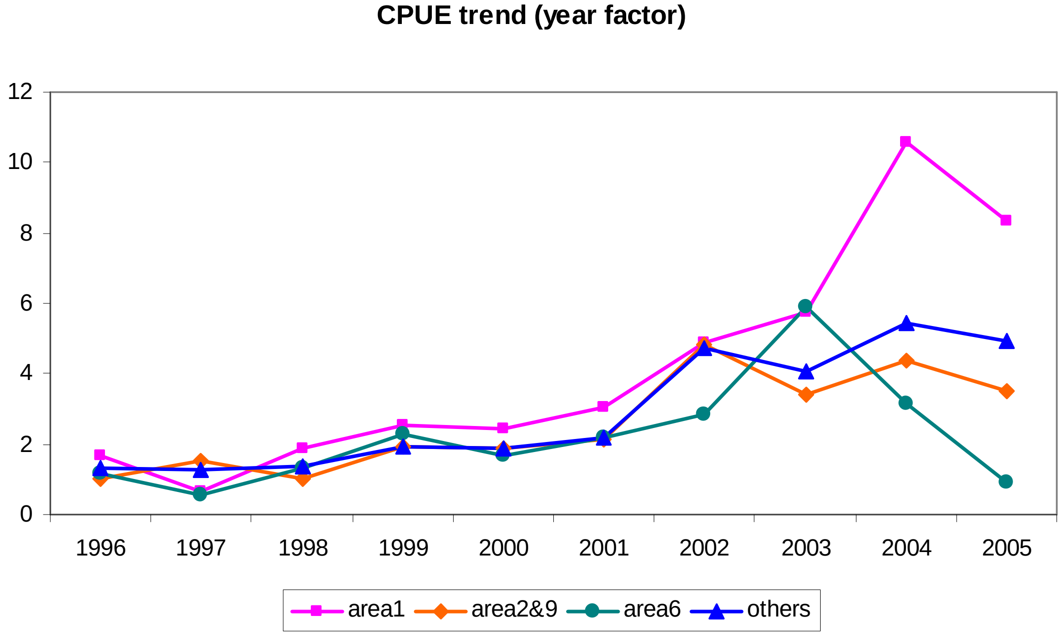


Figure 2: The GLMM standardised CPUE index for Inaccessible Island, compared to the Adjusted Nominal CPUE series. The trend shown as a dashed line is a log-linear regression fitted to annual estimates for the standardised GLMM series.

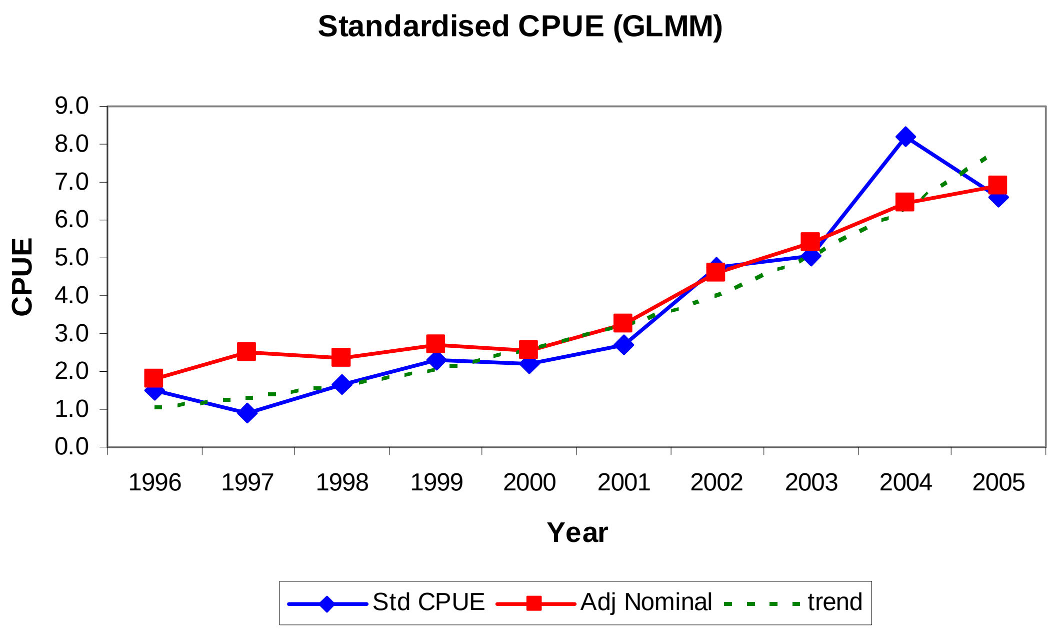
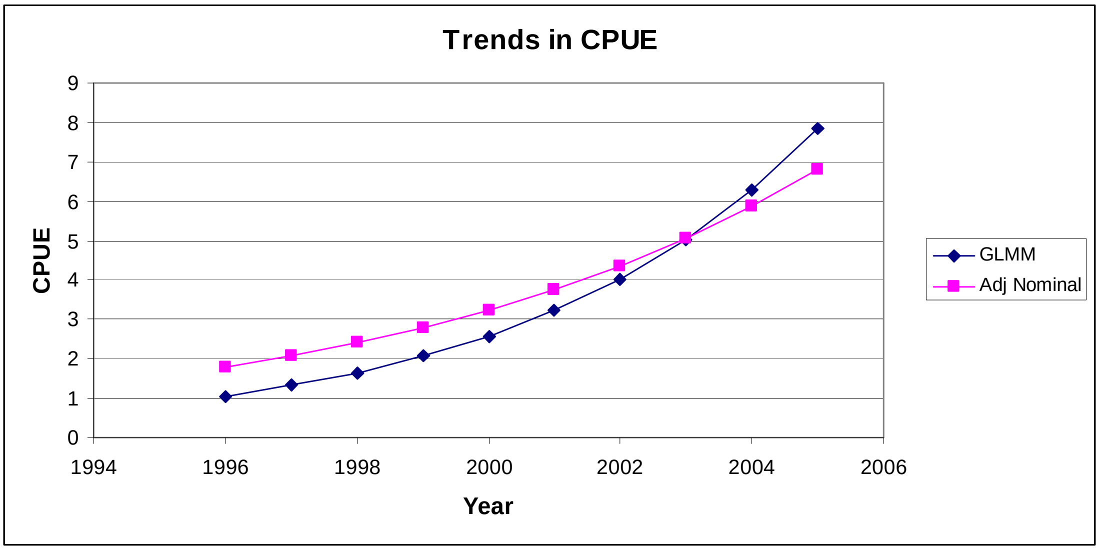


Figure 3: Comparison of the annual log-linear regressions fitted to both the Adjusted Nominal CPUE series (showing a 15.5% annual rate of increase) as well as the GLMM standardised CPUE series (showing a 22.8% annual rate of increase).



Appendix 1

The total number of data records to be used in the GLMM, after applying the various eliminating filters listed in the main text is 3920.

Figure A1.1: % data records for each Season-Year.

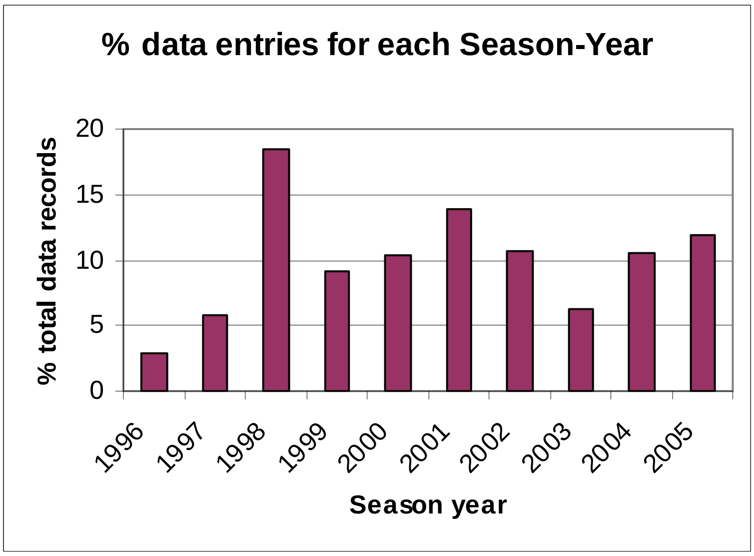


Figure A1.2: % data records for each area.

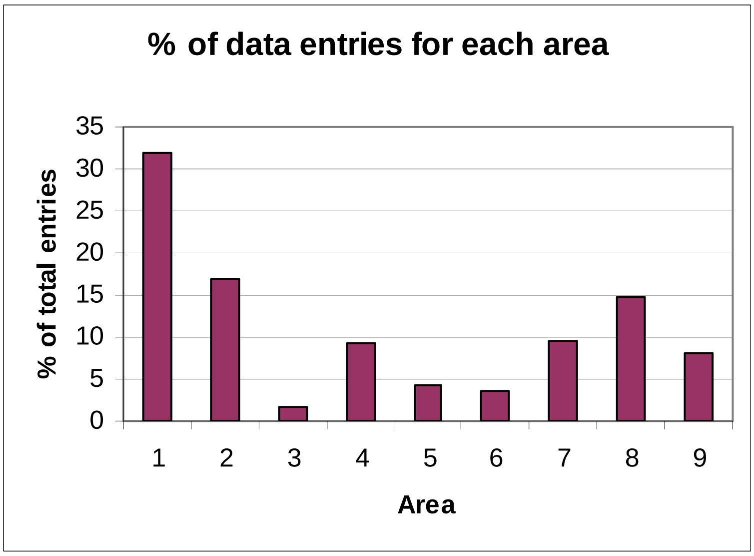
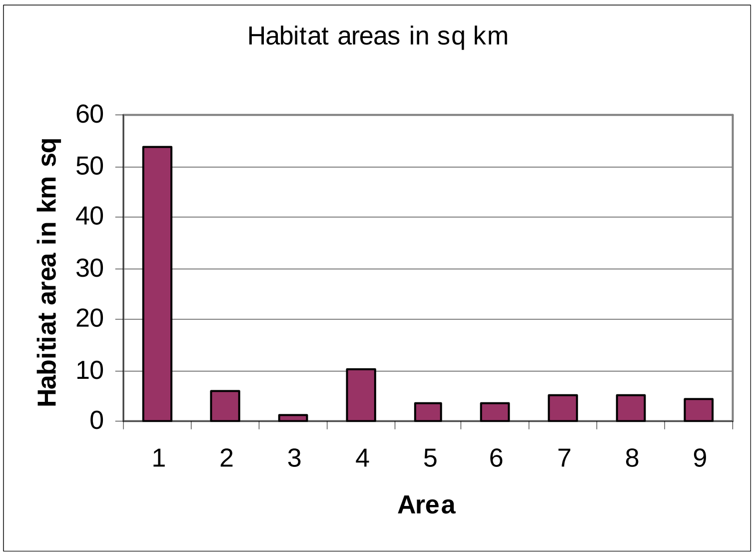


Figure A1.3: Habitat area in km^2 for each area.



Area	Name
1	Bank
2	North point
3	Salt beach
4	East Point
5	Toms beach and Black spot
6	South Hill
7	Pyramid rock and Blinder
8	West point
9	Blendon Hall

Figure A1.4: % data records for each fishing month.

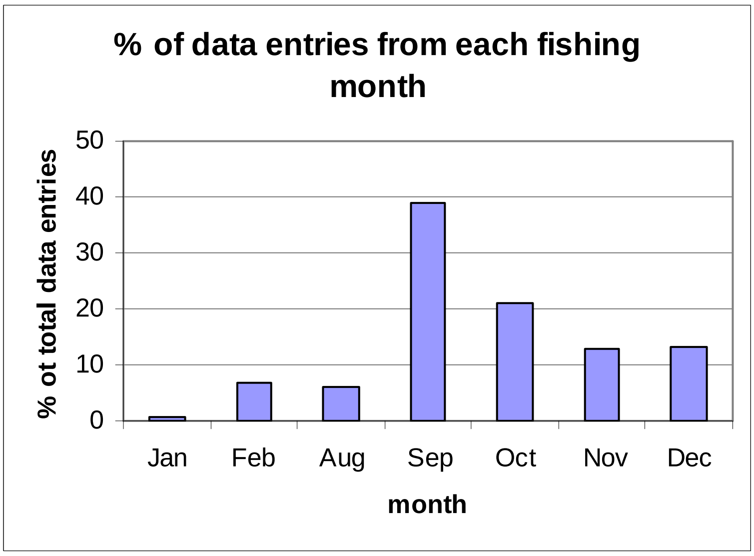


Figure A1.5: % data records for each trap type.

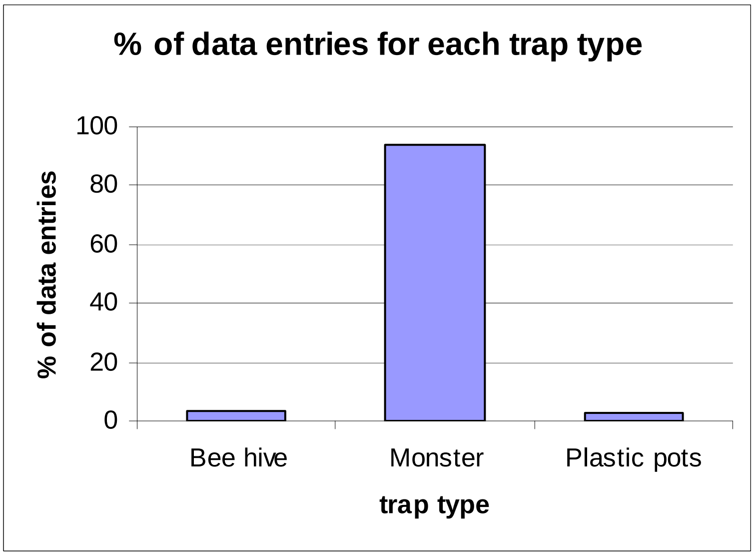


Figure A1.6: % data records for each depth bin (0 reflects 0-9m).

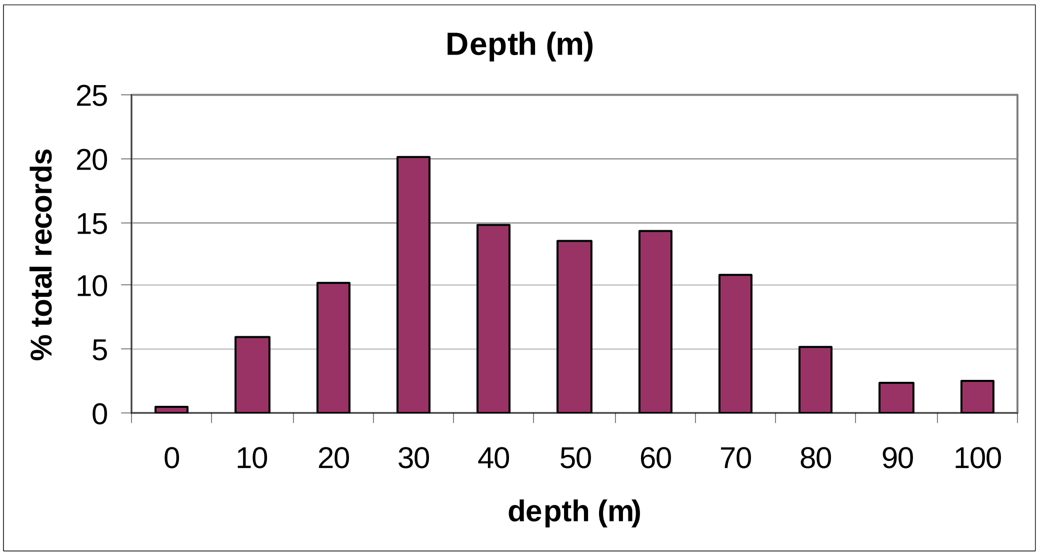


Figure A1.7: Average nominal CPUE for each depth bin (0 reflects 0-9m).

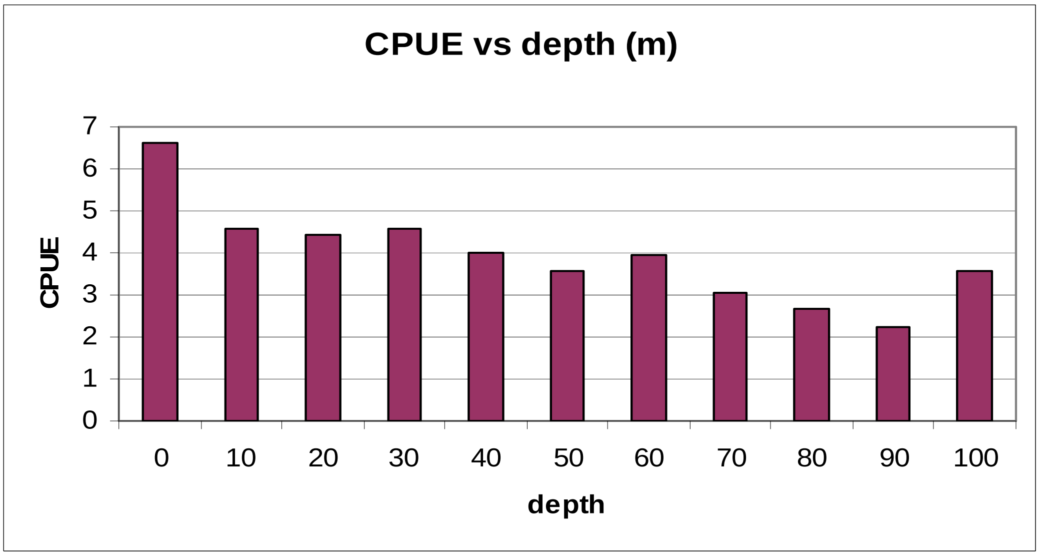


Figure A1.8: % data records for each soak time bin (0 reflects 0 to 0.2 days, i.e. less than 6 hours).

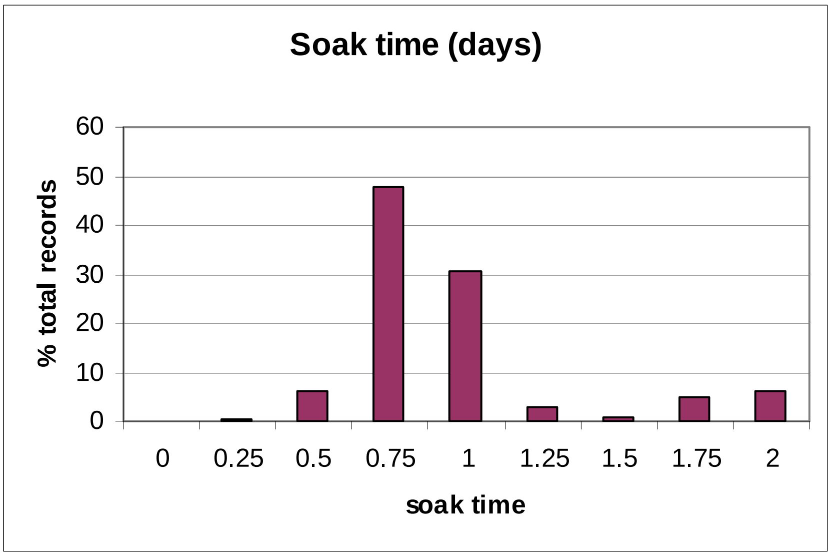


Figure A1.9: Average CPUE at each soak time bin (0 reflects 0 to 0.2 days, i.e. less than 6 hours).

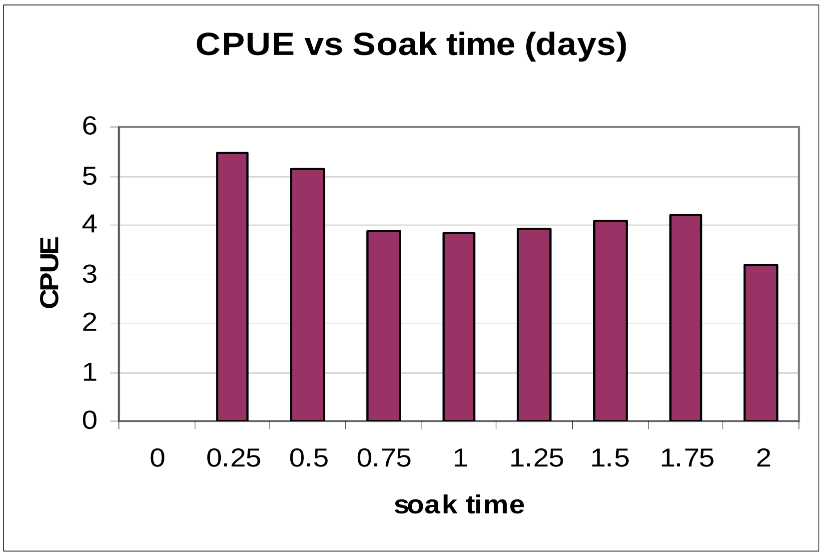


Figure A1.10: Map showing the fishing areas around Inaccessible Island.

Appendix 2

Further results from the GLMM standardisation of the lobster CPUE data for Inaccessible Island.

Table A2.1: The F statistic and its associated p-value for each fixed effect term included in the GLMM. Note that the F tests reported here are for the sequential fit of each term to the model (i.e. each row gives the effect of adding that term to a model that contains all the terms in the preceding rows).

Model term	F statistic	p-value
ϕ_{year}	6.70	0.002
ζ_{month}	0.20	0.969
γ_{area}	30.87	< 0.001
θ_{depth}	26.77	< 0.001
$\lambda_{soak\ time}$	9.48	< 0.001
$\eta_{trap\ type}$	22.7	< 0.001
$\tau_{year \times area}$	5.04	< 0.001

Figure A2.1: Main effects of the fixed effect factors in the GLMM. The error bars represent the ± one standard error