A summary of the General Linear Modelling approach applied to standardize the CPUE data for the offshore trawl fishery for *Merluccius capensis* and *M. paradoxus* off the coast of South Africa.

J.P. Glazer and D.S. Butterworth

Introduction

The models applied to standardize the CPUE data of *Merluccius capensis* and *M. paradoxus* caught offshore off the coast of South Africa are summarised here. This is not straightforward because CPUE indices are required at the species level, but the offshore trawl commercial catch data are recorded only for both species combined. Consequently algorithms developed by Gaylard and Bergh (2004), which make use of species proportions by size at depth, as estimated from research surveys, have been applied to split the hake catches by species at a coast level (west and south) before combining the data from both coasts to perform coast-combined species-specific analyses. Note that this approach can be used from 1978 onwards only, as prior to that the depth of drags was not recorded.

The data used in the analyses are obtained from the Marine and Coastal Management (MCM) demersal database. A fair amount of pre-processing of the data is necessary to ready them for GLM analysis purposes. This includes accumulating the drag-by-drag data on a daily basis per vessel and applying rules to identify and exclude potential "outliers" from the analyses. The daily accumulation is necessary because in certain cases operational constraints prevent the recording of catches per trawl. In such cases effort is recorded per trawl, whereas the catch for the day is logged against the effort of only one of the trawls (usually the last) of the day. Alternatively, the skipper may average the daily catch across the drags of the day.

Separating the species

The algorithms from Gaylard and Bergh (2004) that were used to split the catches by species are summarized below. These splits are made for each trawl.

The proportion of *M. capensis* in size category *s* in each trawl is given by:

$$\bar{p}_s = \frac{1}{1+e^{B_s}} \tag{1}$$

For the west coast:

$$B_s = \kappa_s[d - (d_s^* + \alpha_y + \beta_L + 0.5\gamma_{summer})] \tag{2}$$

For the south coast:

$$B_s = \kappa_s[d - (d_s^* + \beta_L)] \tag{3}$$

where: κ_s is the coast-specific slope parameter for size category *s*,

d is the trawl depth in metres,
d_s^* is the coast-specific shift parameter for size category s,
α_y is the coast-specific year parameter for year y,
β_L is the coast-specific long-shore parameter for long-shore category L, and
$0.5\gamma_{summer}$ is the average of the west coast summer and winter season factors estimated in the fit to the survey data.

Note that the α, β and γ parameters are estimated taking them to be independent of size category. Season and year factors are omitted for the south coast, as they were not significant in the Gaylard and Bergh (2004) GLM analyses of the survey data. The parameter values estimated are shown in Table 1 and Figure 1 plots, as an example, the proportion of *M. capensis* per depth and size category in the 23°-24° E area on the south coast.

The General Linear Models

The following two models (equations 4 and 5) were applied to the *M. capensis* and *M. paradoxus* CPUE data respectively:

$$\begin{aligned}\ell n(\mathrm{CPUE}_{\mathrm{capensis}}+\delta) = {} & \alpha+\beta_{year}+\gamma_{depth}+\eta_{area}+\kappa_{seas}+\lambda_{vessel}+\nu(\mathrm{snoek\,CPUE}) \\ & +\nu'(\mathrm{snoek\,CPUE})^2+\varpi(\mathrm{hmack\,CPUE})+\varpi'(\mathrm{hmack\,CPUE})^2 \\ & +interactions+\varepsilon\end{aligned} \qquad (4)$$

$$\begin{aligned}\ell n(\mathrm{CPUE}_{\mathrm{paradoxus}}+\delta) = {} & \alpha+\beta_{year}+\gamma_{depth}+\eta_{area}+\kappa_{seas}+\lambda_{vessel}+\nu(\mathrm{snoek\,CPUE}) \\ & +\nu'(\mathrm{snoek\,CPUE})^2+\varpi(\mathrm{hmack\,CPUE})+\varpi'(\mathrm{hmack\,CPUE})^2 \\ & +interactions+\varepsilon\end{aligned} \qquad (5)$$

(Note: to avoid clutter, the subscripts "*capensis*" and "*paradoxus*" for the parameters of equations 4 and 5 have been omitted)

where: $\mathrm{CPUE}_{\mathrm{capensis}}$ is the catch of *M. capensis* per unit of (hake-directed – the recorded data specifies the target species for each trawl) effort,

$\mathrm{CPUE}_{\mathrm{paradoxus}}$ is the catch of *M. paradoxus* per unit of (hake-directed) effort,

α is the intercept,

year is a factor with 26 levels (1978-2003) associated with the year effect,

depth is a factor with 8 levels in both the *M. capensis* and *M. paradoxus* models:

$d1_{wc}$: 0 - 100m
$d2_{wc}$: 101 - 200m
$d3_{wc}$: 201 – 300m
$d4_{wc}$: 301 – 400m
$d5_{wc}$: > 400m
$d6_{sc}$: 0 - 100m
$d7_{sc}$:101 - 200m
$d8_{sc}$: > 200m

area is a factor with 6 levels in both the *M. capensis* and *M. paradoxus* models*:*
$a1_{wc}$: ≤ 31°00S
$a2_{wc}$: 31°00S - 33°00S
$a3_{wc}$: 33°00S - 34°20S
$a4_{wc}$: > 34°20S
$a5_{sc}$: < 22°00E
$a6_{sc}$: ≥ 22°00E,

seas is a factor with 4 levels in both the *M. capensis* and *M. paradoxus* models:
Summer: December - February
Autumn: March - May
Winter: June - August
Spring: September - November,

vessel is a factor associated with each individual vessel in the dataset being analyzed (note that for the same vessel, different values of this factor may be estimated for *M. capensis* and *M. paradoxus*),

snoek CPUE and hmack CPUE refer to the CPUE of the bycatch species snoek and horse-mackerel respectively (unlike other major by-catch species, these two species tend **not** to co-occur with hake, so that trawls with proportionally larger catches of these two are reflective of some redirection of fishing effort away from hake, of which account needs to be taken in the GLM),

interactions refer to *year*×*depth*, *year*×*area* and *depth*×*area* interactions which allow for spatial density patterns which have changed over time, and

ε is the error term, assumed to follow a normal distribution.

δ is a (usually small) constant added to the CPUE of the species being modelled to allow for the occurrence of zero CPUE values - here δ is taken to be 10% of the average CPUE of the species being modelled in the respective datasets.

Standardizing the CPUE

The introduction of interactions with year requires that the standardized CPUE (assumed to provide an index of local density) be integrated over area to determine an index of abundance. The boundary separating the west and south Coasts is shown in Figure 2 as being from Cape Agulhas to the tip of the Agulhas Bank so that the whole of the major fishing area of Brown's Bank is included in the west coast.

The formula applied to standardize the CPUE for *M. capensis* and *M. paradoxus* is therefore:

$$CPUE_y = \sum_{strata} [e^{\{\alpha + \beta_{year} + \gamma_{depth} + \eta_{area} + \text{autumn} + \text{median vessel estimate} + \nu(\text{snoek } \overline{CPUE}) + \nu'(\text{snoek } CPUE^2) + \varpi(\text{hmack } \overline{CPUE}) + \varpi'(\text{hmack } CPUE^2) + \text{interactions}\}} - \delta] * \frac{A_{stratum}}{A_{total}} \qquad (6)$$

where $A_{stratum}$ is the size of the area of the stratum in nm^2 (e.g. depth 200-300m and latitude 31 - 33°), and
A_{total} is the total size of the area considered (it is not strictly necessary to divide by A_{total}, but this keeps the units and size of the standardised CPUE index comparable with those of the basic CPUE data).

For the west coast the standardised CPUE is calculated for depths > 200m since very little fishing takes place at depths below 200m. The majority of hauls within the 0 - 200m depth range occur very close to the 200m depth contour, and accordingly are of questionable representativeness of densities within the whole depth-latitude stratum to which the above equation would take them to refer. Similarly, the standardized CPUE for the south coast is calculated for depths > 100m only.

Results

276955 records (vessel-days) were included in the *M. capensis* and *M. paradoxus* CPUE analyses. The amount of variation explained by the *M. capensis* model was 65.4%, while that for the *M. paradoxus* model was 55.6%. The standardized CPUE series are presented in Figures 3 and 4, and indicate that (to the extent which these indices provide unbiased indices of abundance) the abundance of *M. capensis* has increased at an average annual rate of 1.3% per annum, while that of *M. paradoxus* has increased at an average annual rate of 0.01% per annum, over the time period considered (1978-2003).

A Spearman's rank correlation coefficient of 0.4 ($p<0.0001$) was obtained between the *M. capensis* and *M. paradoxus* residuals in the GLM analysis, indicating (unsurprisingly) that catch rates for the two species in the same haul are not independent. Residual distribution plots for the two species indicate that the residuals are not normally distributed; however these deviations from normality do not seem visually too severe (Figures 5 and 6).

Table 1: Parameter values for substitution into equations (2) and (3): the coast-and size-specific algorithms used to split the hake catches by species (Gaylard and Bergh, 2004).

West coast		South coast	
Size category values (κ_S)		**Size category values (κ_S)**	
κ_{small}	0.04722	κ_{small}	0.09074
κ_{medium}	0.03325	κ_{medium}	0.03786
κ_{large}	0.02784	κ_{large}	0.02085
Depth parameter values (d_s^*)		**Depth parameter values (d_s^*)**	
d_{small}^*	177.46		181.62
d_{medium}^*	282.76		257.29
d_{large}^*	325.60		386.85
Year parameter α_y			
< 1985	14.04		
1985	21.95		
1986	13.52		
1987	8.02		
1988	0.50		
1989	11.34		
1990	32.73		
1991	11.45		
1992	21.14		
1993	16.31		
1994	4.84		
1995	26.70		
1996	-6.6		
1997	7.22		
1998	5.25		
1999	4.07		
2000	5.25		
2001	5.25		
2002	21.51		
2003	0.00		
Longshore (latitude) factors (β_L)		**Longshore (longitude) factors (β_L)**	
North of 29°S	0.00	West of 21°E	0.00
29-30°S	-4.02	21-22°E	18.92
30-31°S	4.81	22-23°E	-20.74
31-32°S	1.99	23-24°E	-33.63
32-33°S	5.75	24-25°E	-34.00
33-34S	14.93	25-26°E	-11.64
34-35°S	34.81	East of 26°E	44.51
South of 35°S	36.27		
Season factor			
γ_{summer}	-17.02		

Figure 1: Proportion of *M. capensis* per depth and size class on the south coast in the 23º-24ºE area.

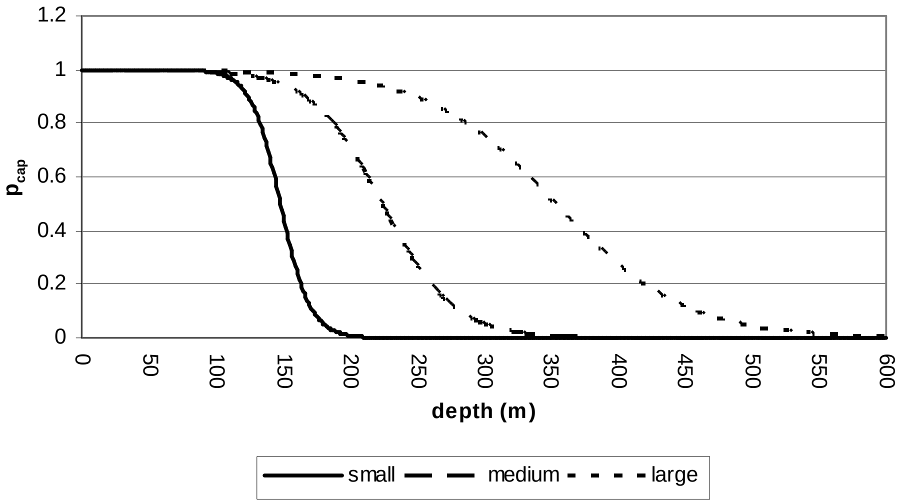


Figure 2: Demarcation of boundaries separating the west and south coasts in the hake fishery. The "Old boundary" was set by ICSEAF and was used to separate coasts until 2004 after which it was agreed by the MCM Demersal Working Group to adopt the "New boundary" for future analyses so that the boundary did not split Brown's Bank. The depth contours shown are the 200m and 1000m contours respectively.

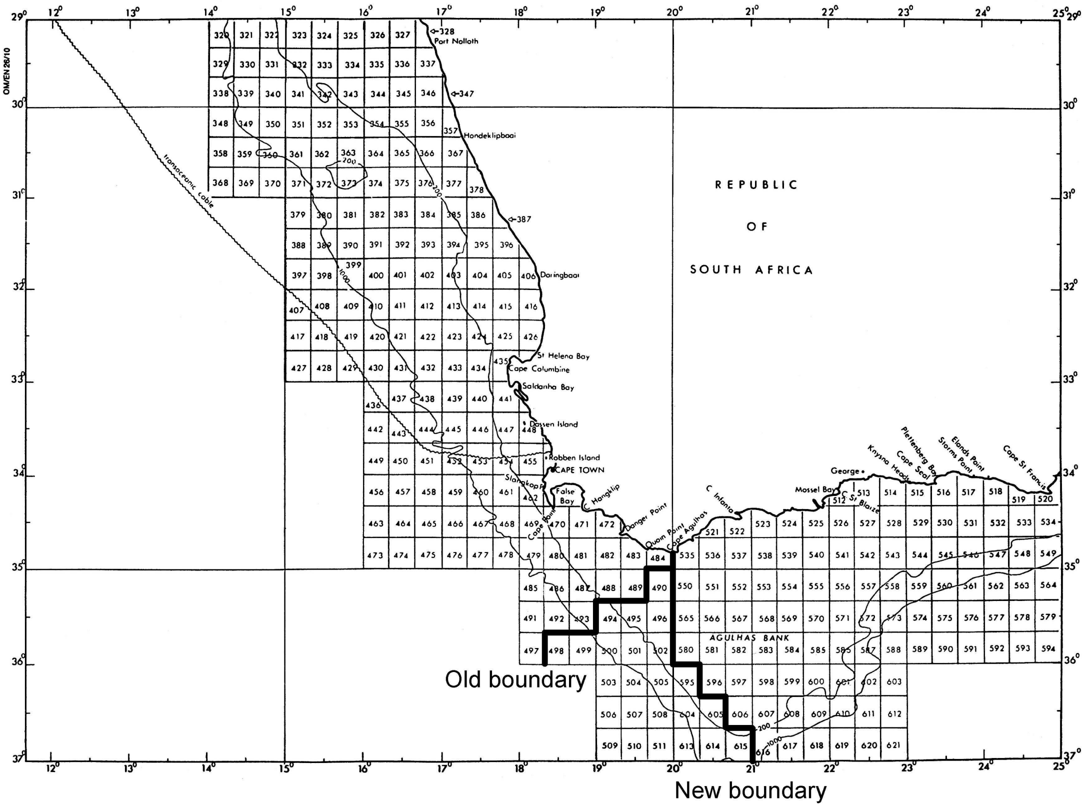


Figure 3: Coast-combined *M. capensis* standardized CPUE.

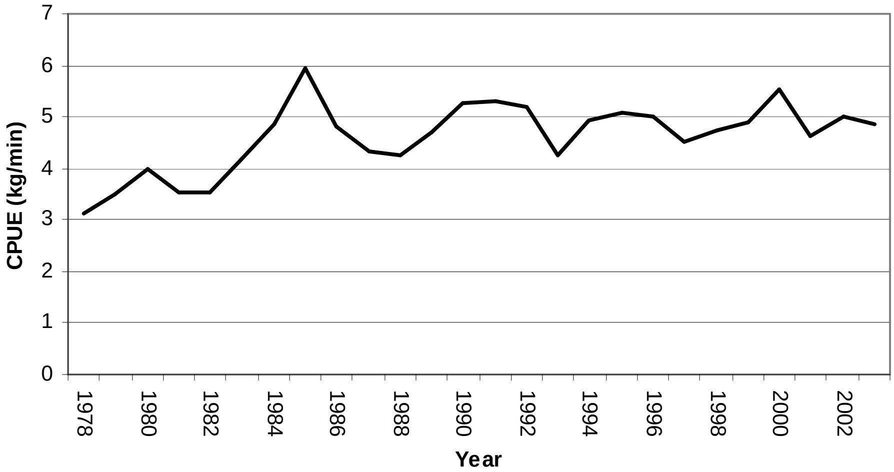


Figure 4: Coast-combined *M. paradoxus* standardized CPUE.

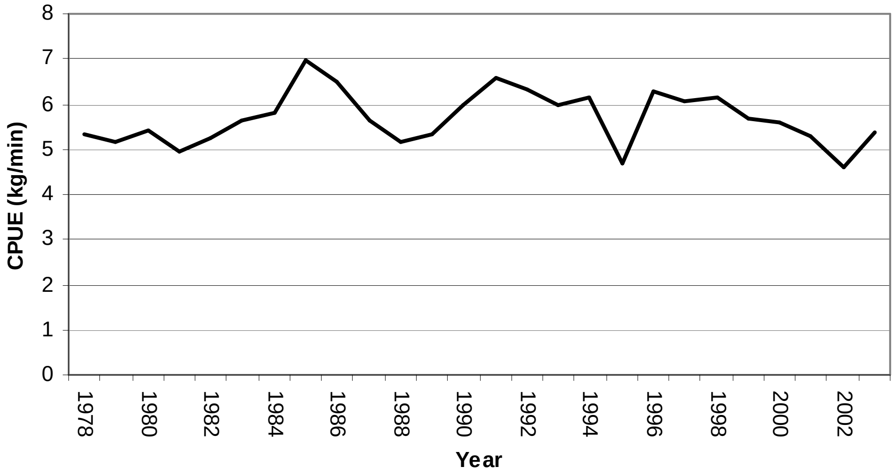


Figure 5: Histogram of unstandardized residuals from the *M. capensis* model.

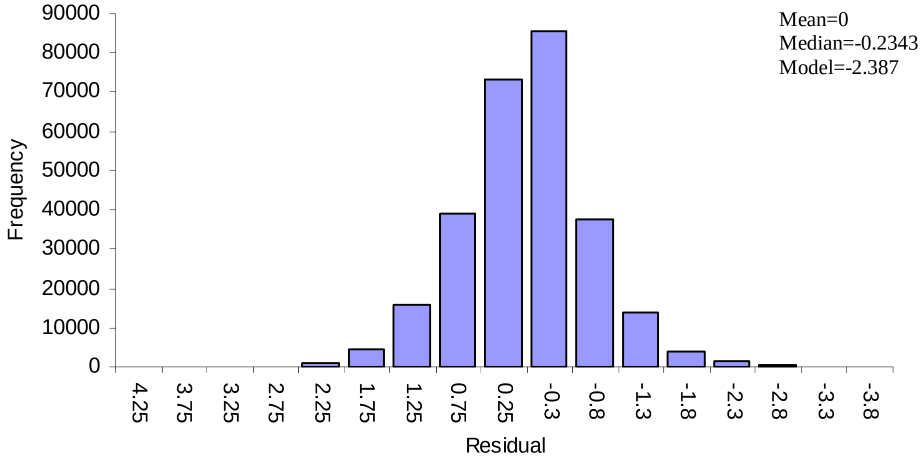
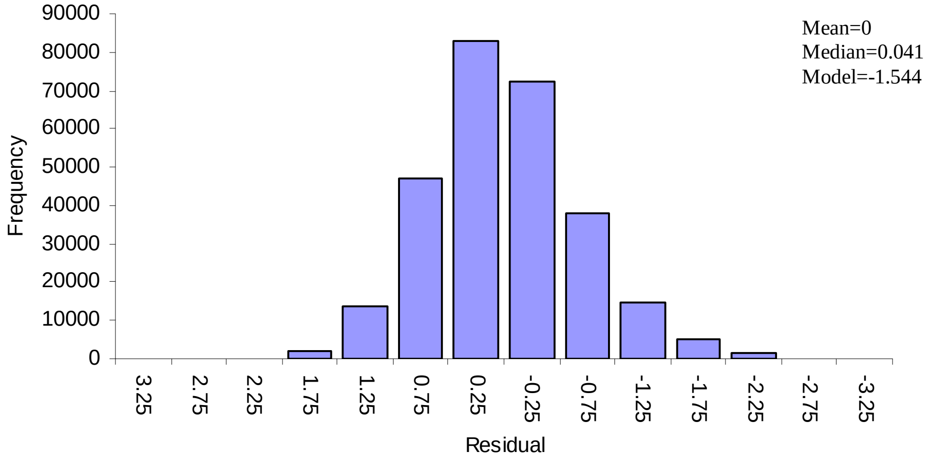


Figure 6: Histogram of unstandardized residuals from the *M. paradoxus* model.



Reference

Gaylard J. D. and M.O. Bergh. 2004. **A species splitting mechanism for application to the commercial hake catch data 1978 to 2003**. Unpublished MCM Demersal Working Group Document, WG/09/04 D:H:21.