# GLM-BASED STANDARDIZATION OF THE CATCH PER UNIT EFFORT SERIES FOR SOUTH AFRICAN WEST COAST HAKE, FOCUSING ON ADJUSTMENTS FOR TARGETING OTHER SPECIES

## J. P. GLAZER* and D. S. BUTTERWORTH†

Catch per unit effort (*cpue*) data for the South African west coast Cape hakes *Merluccius capensis* and *M. paradoxus* were standardized by applying the Generalized Linear Modelling (GLM) technique. The resulting standardized series indicated that resource abundance declined at a rate of 0.4% per annum over the period 1978–1994. Further investigation revealed that the model applied did not adequately adjust for effort being directed away from hake towards other species. This was because, for low bycatch *cpue*, there was a positive rather than the expected negative correlation with hake catch rates. A method to correct for this inadequate adjustment, taking account of an assumed underlying reason of positively correlated fluctuations in catchability of hake and by-catch species, was developed and tested by simulation. This resulted in a more optimistic view of resource status, indicating a 0.6% per year increase in abundance over the period considered. This result remained contrary to general perceptions in the industry that there was a substantial improvement in resource abundance over the 17-year period considered, as suggested by a 4% increase in abundance per year in nominal *cpue*. The reasons that the GLM indicated a lesser increase in resource abundance were that the effective average power of the fleet increased, the fleet moved over time to deeper water, where catch rates were higher, and the increase in fish density in deeper water was more than offset by a simultaneous decrease in density in shallower waters. The inadequate bycatch adjustment arose from the analyses inappropriately including bycatch species that co-occur with hake, and hence are taken incidentally in hake-directed fishing, instead of only species whose capture requires a different targeting of effort.

Key words: bycatch corrections, Cape hakes, catch per unit effort, Generalized Linear Modelling

The two species of Cape hake, *Merluccius capensis* (shallow-water hake) and *M. paradoxus* (deep-water hake), form the mainstay of the bottom trawl industry off South Africa and constitute the country's most valuable fishery, with an annual wholesale value of just over R1 billion (Stuttaford 2000). It is therefore important that the status of the Cape hakes be assessed regularly to ensure that exploitation does not exceed sustainable levels. Because both species of hake are morphologically similar, it has not been possible to distinguish them in commercial catch statistics. Consequently, they have been treated as a single species for assessment purposes.

The standard assumption for past assessments has been that there are two stocks of Cape hake, one off the west coast and another off the south coast of South Africa (Fig. 1). For these stocks, assessments assume that catch per unit effort (*cpue*) is proportional to abundance. The nominal *cpue* (hake catch divided by actual time trawled) for the West Coast stock has grown steadily over the period 1978–1994, increasing at an average rate of 3.8% per year. However, this compu-

tation makes no allowance for differences in effective fishing power between vessels in the fleet, or that fishing patterns (both spatial and temporal) changed over this period. In other words, these *cpue* data have not been collected on the basis of some balanced design that would have provided unbiased estimates automatically. Therefore, before inferences about trends in resource abundance can be drawn from the data, they need to be standardized to adjust for the potentially biasing effects mentioned above. Previously, *cpue* data had been standardized by applying vessel power factors that were estimated crudely in the early 1970s (Andrew 1986). The appropriateness of applying these factors to more recent fishing had been questioned (Andrew 1986, Andrew and Butterworth 1987), so the data were re-standardized by applying a Generalized Linear Modelling (GLM) technique, which is widely used for such a purpose. The modelling was carried out using the SAS (1985) GLM package, because it takes account of unequal numbers of observations for different combinations of categorical variables specified in the model.

* Marine & Coastal Management, Department of Environmental Affairs and Tourism, Private Bag X2, Rogge Bay, 8012, South Africa. E-mail: jglazer@mcm.wcape.gov.za
† MARAM (Marine Resource Assessment and Management Group), Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch 7001, South Africa
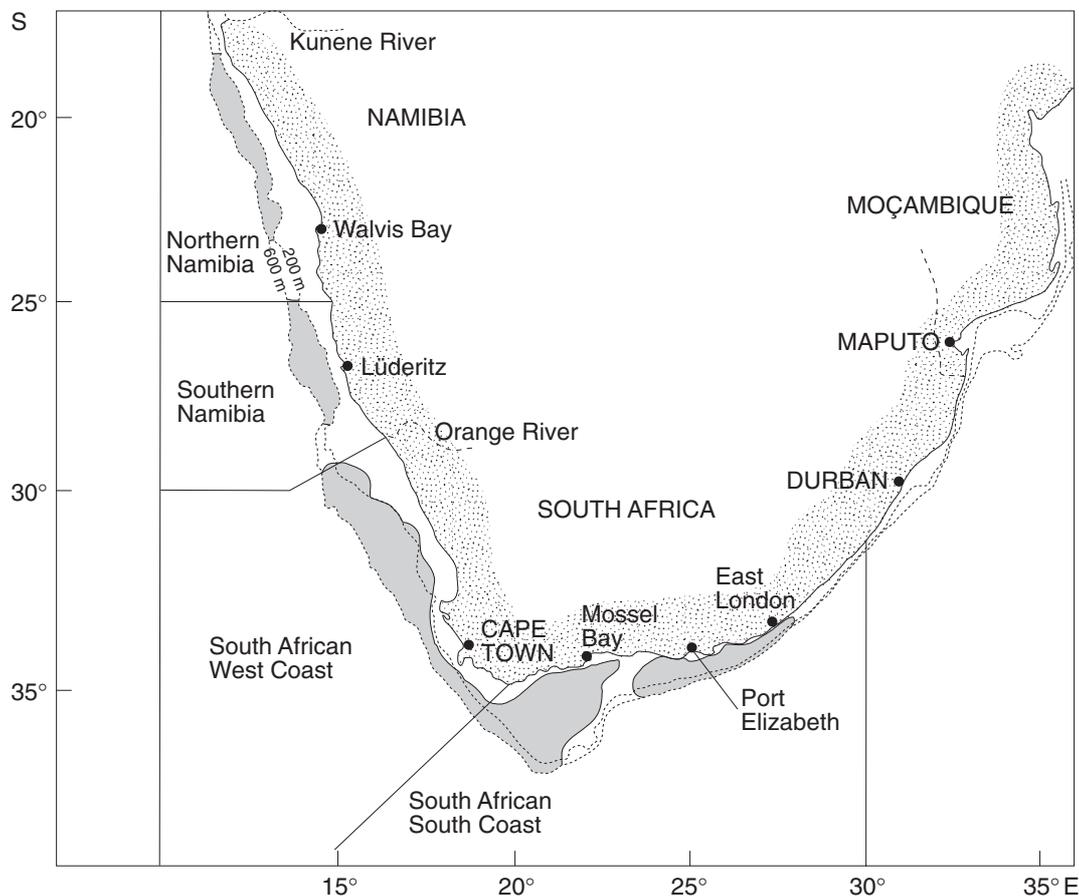
Fig. 1:  Map of South Africa showing the demarcation of West and South coasts in terms of separating the hake stocks off South Africa. The grounds on which Cape hake are caught are illustrated (after Punt 1992)

The data and model used to standardize the *cpue* of West Coast hake are described and the results are presented and discussed. In addition, a method to address the fact that the GLM in its basic form was unable to account adequately for targeting other species is developed.

## MATERIAL AND METHODS

### The data and model

The data used were extracted from the demersal database of Marine & Coastal Management (MCM) and cover the period 1978–1994 (data prior to 1978 are available only at a much less detailed level). The information extracted is listed in Table I.

A GLM of the form shown in Equation 1 was used to model the *cpue* of the West Coast hake stock:

$$\ell n(cpue + \delta) = \alpha + \beta_{year} + \omega_{season} + \eta_{depth} + \tau_{latitude} \\ + \lambda_{vessel} + \gamma(bycatch\ cpue) \\ + \gamma'(bycatch\ cpue)^2 + interactions + \varepsilon \quad ,$$

(1)

where *cpue* is the hake catch (kg) per minute trawled, $\delta$ a (usually small) constant added to *cpue* to allow for zero values, $\alpha$ the intercept, $\beta_{year}$ a year effect, where *year* is a factor with 17 levels (corresponding to 1978 to 1994), and $\omega_{season}$ is a season effect, where *season* is a factor with four levels:

Table I: The information contained in the dataset used to standardize the West Coast hake *cpue* data. Data were extracted from the demersal database of Marine & Coastal Management and used for the analysis after exclusion of a small number of records as a result of validation checks

| Parameter |
| --- |
| Company code (a unique code assigned to each fishing company for identification purposes) |
| Vessel code (a unique code assigned to each fishing vessel for identification purposes) |
| Power factor (as crudely calculated in the early 1970s) |
| Vessel class (vessels were separated into broad categories according to their gross registered tonnage) |
| Landing date (date on which the catch was landed at port) |
| Drag date (date on which a drag took place) |
| Start time (time to the nearest minute at which drag started) |
| Effort (the amount of time net was dragged; recorded in minutes) |
| ICSEAF Division (identifying the Division in which the catch took place: 1.6 = West Coast; 2.1 or 2.2 = South Coast) |
| Grid block in which catch was taken (the fishing grounds are divided into 20 minute squares so that catch positions can be reported accurately) |
| Depth at which catch was taken |
| Mesh size used (75, 85 or 110 mm) |
| Species targeted[1] |
| Total hake catch (kg)[2] |
| Total horse mackerel *Trachurus trachurus capensis* catch (kg)[2] |
| Total monk *Lophius upsicephalus* catch (kg)[2] |
| Total kingklip *Genypterus capensis* catch (kg)[2] |
| Total Agulhas sole *Austroglossus pectoralis* catch (kg)[2] |
| Total West Coast sole *Austroglossus microlepis* catch (kg)[2] |
| Total snoek *Thyrsites atun* catch (kg)[2] |
| Total mackerel *Scomber japonicus* catch (kg)[2] |
| Total chokka squid *Loligo vulgaris reynaudii* catch (kg)[2] |
| Total red squid *Todapopsis eblanae* and *Todarodes angolensis* catch (kg)[2] |
| Total catch (kg) of other species (e.g. ribbon fish *Lepidopus caudatus* and panga *Pterogymnus laniarius*)[3] |
| Mass of hake in the large size category (kg) |
| Mass of hake in the medium size category (kg) |
| Mass of hake in the small size category (kg) |
| Mass of hake in the fillets category (kg) |
| Latitude position at which catch was taken (minutes were converted to decimalized minutes) |
| Longitude position at which catch was taken (minutes were converted to decimalized minutes) |

[1] Analyses were restricted to drags/days indicated as hake-directed. However, this field was not completed consistently, so that many indications of "hake direction" in fact reflected effort directed at other species

[2] Space is provided in the logbooks to declare the quantity of each of these species caught. Apart from hake, the other species are referred to as "declared" bycatch. These bycatch species combined constitute the bycatch *cpue* explanatory variable in the GLM

[3] Space is not provided in the logbooks for declaring the catch of these species, and they are referred to as "undeclared" bycatch. The catch of these species is determined only at the landing site, and is apportioned across the drags of the trip in the same ratio as the catch of the targeted species across drags. These species are not included as explanatory variables in the GLM because the target species is generally recorded as hake, and there is therefore a direct correlation between the catches of hake and these "undeclared" species because of the system used for apportionment

*Summer* = December–February
*Autumn* = March–May
*Winter* = June–August
*Spring* = September–November,

$l1 \leq 31°00'S$
$31°00'S < l2 \leq 33°00'S$
$33°00'S < l3 \leq 34°20'S$
$l4 > 34°20'S,$

$\eta_{depth}$ is a depth effect, where *depth* is a factor with five levels:

$d1 \leq 100$ m
$100$ m $< d2 \leq 200$ m
$200$ m $< d3 \leq 300$ m
$300$ m $< d4 \leq 400$ m
$d5 > 400$ m,

$\tau_{latitute}$ is a latitudinal effect, where *latitude* is a factor with four levels:

$\lambda_{vessel}$ is a vessel effect, where *vessel* is a factor with 110 levels (each vessel being uniquely identified by the vessel code recorded in the database), *bycatch cpue* is a continuous variable relating to the catch rate of bycatch species (the sum of all "reported" bycatch species divided by total effort per vessel day), upon which the logarithm of hake *cpue* is assumed to depend quadratically, *interactions* refer to *year × depth*, *year × latitude* and *depth × latitude* interactions, and $\varepsilon$ is the error term, which is assumed to follow a normal distribution $N(0,\sigma^2)$.

```
                                                              #
  +2.25 ┤ *                                                   31
         │ *                                                  467
         │ ******                                           4 955
         │ ************************                        23 261
         │ ***************************************************  45 247
         │ ********************************************    36 143
         │ ******************                             16 949
         │ *******                                         6 398
         │ ***                                             2 467
         │ *                                                 730
         │ *                                                  53
  −3.25 ┤ *                                                   1
         └─────────────────────────────────────────────
                        FREQUENCY
```
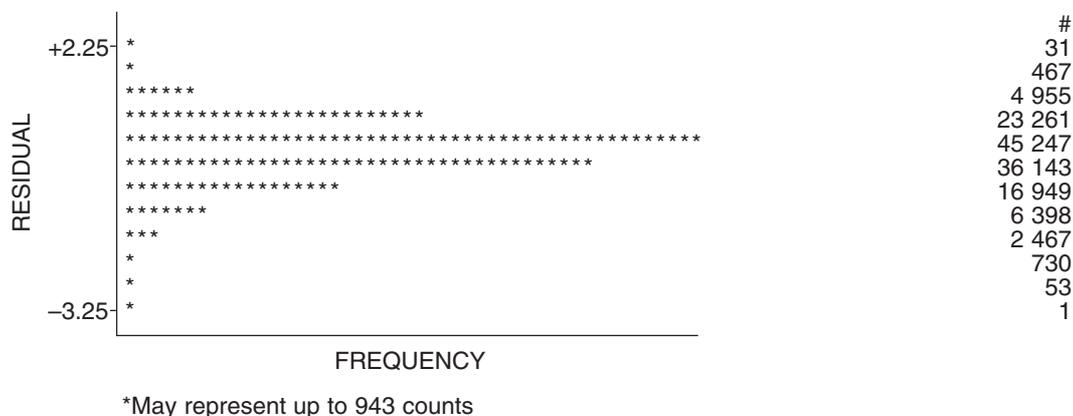
RESIDUAL

*May represent up to 943 counts

Fig. 2: Frequency plot of the residuals obtained from the GLM of Equation 1 applied to the West Coast hake *cpue* data

Latitude was included in the model to account for the possibility that there had been a longshore as well as offshore (reflected by the depth variable) shift in fish density or fishing patterns over time. Table I defines what constitutes a "reported" bycatch species, and explains why catches of other "unreported" species could not be taken into account in this analysis.

Although mesh size is an important factor that influences catches, it would not be appropriate to include it as an explanatory variable in the model, because the mesh size recorded in the database is generally the minimum mesh size legally required and not the actual mesh size used. Furthermore, the trawling industry has admitted that substantial illegal lining of nets took place during the period under analysis – see Results and Discussion section.

The factors included in the model were selected on the basis of a forward selection procedure, where the change in $r^2$ (the proportion of variation explained by the model) was the criterion upon which the inclusion/exclusion of explanatory variables was based. Although in GLM analyses the inclusion/exclusion of such variables is normally based on tests of statistical significance, this criterion was not applied here because, for large datasets ($n = 136\ 702$ in the present case), factors can be statistically significant but have a minimal quantitative impact on predictions. It would be technically more correct to base variable inclusion/exclusion decisions on the adjusted $r^2$ statistic, where $r^2_{adj} = 1 - (1 - r^2)(n - 1)/(n - m)$, and $m$ is the number of parameters estimated, to account for loss of degrees of freedom as more parameters are estimated. However, in the cases considered here, $m \ll n$, so $r^2$ and $r^2_{adj}$ hardly differ. Furthermore, inferences of significance are based on the assumption of independence of data,

which is unlikely to hold here.

It would be desirable to apply a pre-specified quantitative criterion in deciding the inclusion/exclusion of explanatory variables (Doonan 1991). However, in the context of this study, decisions with respect to the inclusion of factors in the model (based on $r^2$) were made by the Demersal Working Group of MCM, the body responsible for reviewing hake assessment analyses to provide scientific recommendations for management. Viewed in hindsight, these decisions did not always reflect consistent application of identical criteria. It should be appreciated that such decisions often reflected compromises between opposing views in the context of a wider debate at the time concerned.

**Selection of δ**

The analysis is based upon the assumption that the data are log-normally distributed. This was made partly for convenience, because it allowed for the use of statistical packages that are restricted to this assumption (and were the only type available during the present study that could handle so large a dataset). However, one problem that can arise when log-normality is assumed is the occurrence of zero *cpue* values. This is usually dealt with by adding a small constant (δ) to the *cpue* before taking logs (Campbell *et al*. 1995, ICCAT 1995).

As the scale for *cpue* values is essentially arbitrary, the choice of δ must be linked to the scale adopted. This is usually achieved by defining δ to be some percentage of the overall nominal mean (ICCAT 1995). This approach was adopted here, although it does have the disadvantage that, in absolute terms, the value of δ changes slightly each year as additional data become available

and are included in the analysis, because this in turn leads to a change in the average nominal *cpue* value.

Selection of δ was initially intended to be based on the requirement of (near-) normality of the residual distribution obtained from the model fit, as assumed in estimating the parameters of Equation 1 in the GLM (residuals are the differences between the actual data and the values predicted by the model). However, for reasons associated with reducing bias linked to attempts to correct for bycatch (which are discussed later), the δ value was selected to be 10% of the average hake *cpue*. A frequency plot of the residuals for the model described in Equation 1 is shown in Figure 2. From this it is evident that the distribution is skewed to some extent. Although this non-normality precludes accurate statements about confidence intervals for the resultant estimates, the primary concern of this work was to minimize possible bias in the trend of the estimates of standardized *cpue* over time.

## Standardizing the *cpue*

The method of calculating the annual standardized *cpue* is more complicated than simply assuming that annual abundance is proportional to the exponentiated year-factor (Kimura 1981). This complication arises as a result of the inclusion of δ, which introduces non-linearity in the log-transformed *cpue* as a function of its explanatory variables. Therefore, δ needs to be factored out when calculating the standardized annual *cpue* (because it is assumed that fish density is proportional to *cpue*, and not *cpue* + δ). This was achieved by applying Equation 2 below and assuming conditions pertinent to "average" fishing; hence the choices made were that average season = *autumn*, average vessel is that corresponding to the median of the vessel factor estimates, and the bycatch *cpue* is the average of that recorded over the period. The *cpue* (assumed to provide an index of local density) was then integrated over area to determine an index of abundance (Honma 1974):

$$cpue_y = \sum_{strata} \left[ e^{\left( \begin{array}{c} \alpha + \beta_{year} + \omega \overline{_{season}} + \eta_{depth} \\ + \tau_{latitude} + \lambda_{median\ vessel} \\ + \gamma \left( \overline{bycatchcpue} \right) \\ + \gamma' \left( \overline{bycatchcpue} \right)^2 \\ + interactions \end{array} \right)} - \delta \right] \times \frac{A_{stratum}}{A_{total}}$$

(2)

where $cpue_y$ is the standardized index of abundance for year $y$, $A_{stratum}$ the size of the stratum (e.g. depth 200–300 m, latitude 31–33°S), and $A_{total}$ is the total size of the area considered (it is not strictly necessary
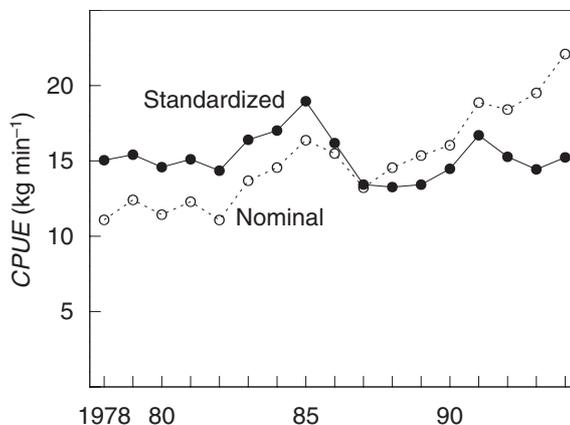


Fig. 3: West Coast hake nominal and GLM-standardized *cpue* (using Equations 1 and 2)

to divide by $A_{total}$, but this keeps the units and size of the standardized *cpue* index comparable with those of the basic *cpue* data, which assists interpretation).

There are some empty cells for the model of Equation 1 because no fishing took place in the 0–100 m depth zone in certain years. The abundance index of Equation 2 was consequently restricted to depths >200 m. The rationale for this is that, although the area from 0–200 m makes up a substantial portion (54%) of the hake habitat up to 500 m, very little fishing (some 2% of the hauls) takes place shallower than 200 m. Furthermore, most hauls within the 0–200 m depth range are taken close to the 200-m depth contour, and accordingly are of questionable representativeness of density within that whole depth-latitude stratum.

The standardized *cpue* obtained from the model indicated that the abundance of the West Coast hake stock had declined at a rate of 0.4% per year over the period 1978–1994, i.e. a different situation from that indicated by the nominal trend of +3.8% per year (Fig. 3).

Strictly speaking, the exponent in Equation 2 should include a correction factor for bias ($\frac{\sigma^2}{2}$), because if log($x$) ~ $N(\eta, \sigma^2)$, then $E(x) = e^{(\eta + \frac{\sigma}{2})}$ (Kimura 1981). Here, the inclusion of this bias correction factor had minimal impact on the standardized index of abundance (a change of <0.01% per year in the estimate of trend) and was therefore not included in the standardization calculation.

## Improved method of adjusting for bycatch

Inclusion of the bycatch *cpue* as an explanatory variable in the model was to account for the impact that targeting other species would have on the hake *cpue*. Anon.
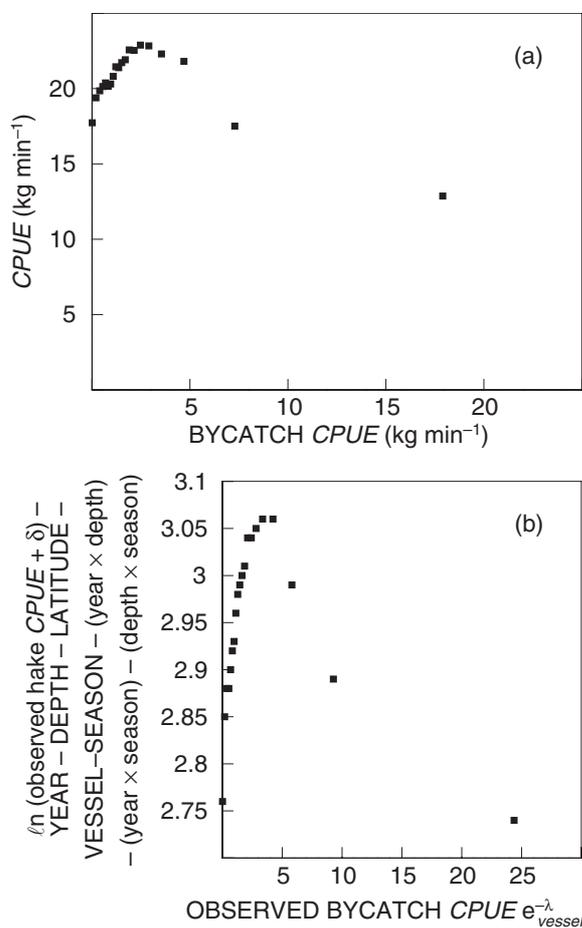
Fig. 4: Relationship between (a) West Coast hake nominal *cpue* and bycatch *cpue* and (b) West Coast hake standardized *cpue* and bycatch *cpue.* Each point represents the average hake and "reported" bycatch *cpue* in five percentile intervals

(1996) derived a relationship between the *cpue* of the species being modelled and bycatch, which could be used in GLMs to adjust for targeting bycatch species when only the total effort expended (both on the species under consideration and on bycatch) is known. The relationship suggests an approximately linear relationship between the log of observed hake *cpue* and bycatch *cpue* in the limit of a relatively small proportion of effort targeted at bycatch. A quadratic term was included in Equation 1 to allow for deviations from linearity for greater degrees of targeting.

The model suggests that, as bycatch *cpue* increases (i.e. bycatch species are increasingly targeted), so hake *cpue* decreases, because a greater proportion of the total effort used to compute this *cpue* is being directed elsewhere. The predicted decrease, however, is not evident in this case. Figure 4a shows that, at low levels of bycatch *cpue*, there is a positive correlation between hake and bycatch *cpue*, whereas the expected negative correlation is evident only at higher levels of such bycatch *cpue* (Bergh and Barkai 1996). Given the strength of this positive correlation in the dataset being analysed, use of bycatch *cpue* as an explanatory variable in the GLM would not allow proper adjustment for targeting bycatch. This was highlighted by the fact that the bycatch *cpue* parameter ($\gamma$) estimated in the GLM was a very small negative number, indicating, counter-intuitively, that very little hake *cpue* had to be forfeited in order to increase bycatch *cpue* (Bergh and Barkai 1996).

It could conceivably be that the upward portion of the plot in Figure 4a is explained by the temporal and spatial effects already included in the model of Equation 1. To check for this, Figure 4b was developed to indicate the relationship between hake and bycatch *cpue*, after correcting for temporal and spatial effects. Specifically $\ell n(hake\ cpue + \delta) - \beta_{year} - \omega_{season} - \eta_{depth} - \tau_{latitude} - \lambda_{vessel} - (year \times depth) - (year \times season) - (depth \times season)$ is plotted against bycatch *cpue* (corrected for vessel effects). It is quite clear that the positive correlation is still present, and hence that some procedure to remove this effect needs to be applied.

This positive correlation between hake and bycatch *cpue* possibly arises from the fact that there are periods of high and low catchability that are related to environmental conditions, which would affect both hake and bycatch *cpue* in the same direction. Adjusting for bycatch *cpue* in the GLM as in Equation 1 in such circumstances would lead to an incorrect downward adjustment of the hake abundance index. However, adjustments are still required to remove the bias caused by the lack of specifying effort targeted away from hake in favour of bycatch (causing the negative correlation at larger bycatch *cpue* values). The problem is that both effects were present for West Coast hake, and it was therefore necessary to disentangle them. This was achieved (to some extent) by adopting the procedure described below.

The residuals associated with the bycatch *cpue* (as an index of abundance of bycatch species) were assumed to be correlated with those of the hake *cpue* (as an index of abundance of hake), with correlation coefficient $\rho$. This correlation is assumed to be a manifestation of related fluctuations in catchability. The bycatch *cpue* was therefore adjusted as follows:

$$cpue_i^{o*} = cpue_i^o \times e^{-\lambda_{vessel}} \times e^{-\rho\varepsilon_i} \qquad (3)$$

Table II:  Values of $r^2$ (the proportion of the variance explained) from applying the iterative correction procedure of Equation 3 to 17 years of simulated data (100 data points in each year). Results for δ = 0.1 × mean hake *cpue*, and δ = 0.5 × mean hake *cpue* are shown for cases where the true annual rate of biomass increase β is 0.01 or 0.02

| Parameter | | True ρ | | | | |
|---|---|---|---|---|---|---|
| | | 0.0 | 0.25 | 0.50 | 0.75 | 1.0 |
| β = 0.01, δ = 0.10 | 0.0 | 60.9% | 56.0% | 51.6% | 47.6% | 44.5% |
| | 0.25 | 69.5% | 66.0% | 62.2% | 59.5% | 57.4% |
| Assumed ρ | 0.50 | 74.7% | 71.9% | 69.1% | Oscillates between 66.4 and 66.7% | Oscillates between 64.4 and 64.7% |
| | 0.75 | No convergence | No convergence | No convergence | No convergence | No convergence |
| | 1.0 | No convergence | No convergence | No convergence | No convergence | No convergence |
| β = 0.02, δ = 0.10 | 0.0 | 60.9% | 56.1% | 51.6% | 47.6% | 44.5% |
| | 0.25 | 69.6% | 66.0% | 62.7% | 59.5% | 57.5% |
| Assumed ρ | 0.50 | 74.7% | 72.0% | 69.2% | Oscillates between 66.5 and 66.8% | Oscillates between 64.5 and 64.7% |
| | 0.75 | No convergence | No convergence | No convergence | No convergence | No convergence |
| | 1.0 | No convergence | No convergence | No convergence | No convergence | No convergence |
| β = 0.01, δ = 0.50 | 0.0 | 53.2% | 48.3% | 43.8% | 39.9% | 36.8% |
| | 0.25 | 57.9% | 53.8% | 50.0% | 46.6% | 43.8% |
| Assumed ρ | 0.50 | 61.6% | 58.1% | 54.8% | 51.7% | 49.3% |
| | 0.75 | 64.9% | 61.5% | 58.5% | 55.7% | 53.7% |
| | 1.0 | 67.2% | 64.3% | 61.6% | 59.0% | Oscillates between 56.8 and 56.9% |
| β = 0.02, δ = 0.50 | 0.0 | 53.4% | 48.5% | 44.1% | 40.1% | 37.0% |
| | 0.25 | 58.1% | 54.1% | 50.3% | 46.8% | 44.1% |
| Assumed ρ | 0.50 | 61.9% | 58.4% | 55.1% | 52.0% | 49.6% |
| | 0.75 | 64.9% | 61.8% | 58.8% | 56.0% | 53.0% |
| | 1.0 | 67.5% | Oscillates between 64.5 and 64.6% | 61.9% | 59.3% | Oscillates between 57.1 and 57.2% |

where $cpue_i^{o*}$ is the corrected observed bycatch *cpue* for vessel-day $i$, $cpue_i^o$ the observed bycatch *cpue* for vessel-day $i$, $e^{-\Lambda_{vessel}}$ a correction term for the vessel effect, $\varepsilon_i$ the residual of the GLM fit to $\ell$n(hake *cpue* + δ) for vessel-day $i$, and ρ is the correlation assumed.

The reason for adjusting by the vessel factor is to standardize effort similarly for both hake and bycatch *cpue*. This is in the spirit of the model underlying the expected linearity of the log(hake *cpue*) v. bycatch *cpue* relationship.

In order to implement this bycatch adjustment, an iterative procedure was applied as follows:

(1)  Run the model of Equation 1 and retain the residuals.
(2)  Adjust the bycatch *cpue* by applying Equation 3, using an externally specified value for ρ.
(3)  Rerun the model, replacing the bycatch *cpue* with that calculated in (2), and retain the residuals.
(4)  Repeat (2) and (3) until convergence is obtained.

Stability of the slope statistic (obtained from regressing $\ell$n [standardized *cpue*] against time) was used as a criterion to indicate when adequate convergence had been achieved.

An assumption underlying this procedure is that there is no temporal trend in the abundance of bycatch species. In this case, little is known about trends in abundance for the bycatch species considered, except for those of chokka squid *Loligo vulgaris reynaudii* and kingklip *Genypterus capensis*, both of which are considered to have decreased in abundance off South Africa (Punt and Japp 1994, Roel *et al.* 2000). Neglect of these trends introduces some negative bias into trends of standardized indices of hake abundance.

In order to evaluate possible biases inherent in this method, which is proposed to adjust for the positive correlation evident between hake and bycatch *cpue*, at low levels of the latter, a simulation testing exercise was undertaken (see Appendices 1, 2).

Table III: Estimates of the slope β (annual trend in abundance) from applying the iterative correction procedure of Equation 3 to 17 years of simulated data (100 data points in each year). Results for δ = 0.1 × mean hake *cpue*, and δ = 0.5 × mean hake *cpue* are shown for cases where the true annual rate of biomass increase β is 0.01 or 0.02

| Parameter | | | True ρ | | | | |
|---|---|---|---|---|---|---|---|
| | | | 0.0 | 0.25 | 0.50 | 0.75 | 1.0 |
| β = 0.01, δ = 0.10 | | 0.0 | 0.010 | 0.009 | 0.009 | 0.008 | 0.008 |
| | | 0.25 | 0.011 | 0.010 | 0.010 | 0.010 | 0.008 |
| Assumed ρ | | 0.50 | 0.011 | 0.011 | 0.011 | 0.010 | 0.010 |
| | | 0.75 | No convergence | No convergence | No convergence | No convergence | No convergence |
| | | 1.0 | No convergence | No convergence | No convergence | No convergence | No convergence |
| β = 0.02, δ = 0.10 | | 0.0 | 0.018 | 0.018 | 0.017 | 0.017 | 0.017 |
| | | 0.25 | 0.019 | 0.019 | 0.018 | 0.018 | 0.017 |
| Assumed ρ | | 0.50 | 0.020 | 0.020 | 0.019 | 0.019 | 0.018/0.019 |
| | | 0.75 | No convergence | No convergence | No convergence | No convergence | No convergence |
| | | 1.0 | No convergence | No convergence | No convergence | No convergence | No convergence |
| β = 0.01, δ = 0.50 | | 0.0 | 0.006 | 0.006 | 0.005 | 0.005 | 0.005 |
| | | 0.25 | 0.006 | 0.006 | 0.006 | 0.006 | 0.005 |
| Assumed ρ | | 0.50 | 0.007 | 0.006 | 0.006 | 0.006 | 0.006 |
| | | 0.75 | 0.007 | 0.007 | 0.007 | 0.006 | 0.006 |
| | | 1.0 | 0.007 | 0.007 | 0.007 | 0.007 | 0.006 |
| β = 0.02, δ = 0.50 | | 0.0 | 0.012 | 0.012 | 0.011 | 0.011 | 0.011 |
| | | 0.25 | 0.012 | 0.012 | 0.012 | 0.012 | 0.011 |
| Assumed ρ | | 0.50 | 0.013 | 0.013 | 0.012 | 0.012 | 0.012 |
| | | 0.75 | 0.013 | 0.013 | 0.013 | 0.013 | 0.012 |
| | | 1.0 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 |

## RESULTS AND DISCUSSION

The results from the simulation exercise for the variety of choices for δ indicated that the two smaller values considered (0.05 and 0.1 × mean hake *cpue*) yielded virtually unbiased estimates of biomass rates of increase β (under both 1 and 2% per year simulated biomass increases), whereas the larger values led to negatively biased estimates. The small values of δ also yielded near-identical estimates of β, irrespective of the true and assumed values of ρ. The variation explained by the model ($r^2$) increased with assumed ρ, irrespective of the true value of ρ; therefore, the true value of ρ cannot be estimated on the basis of the choice that yields the highest $r^2$. As examples, those for two of the options for δ (0.1 and 0.5 × mean hake *cpue*) are presented in Tables II and III, the former reporting the variation explained by the model and the latter the estimates of β.

Given the bias indicated by the simulation results for larger values of δ, 10% of mean hake *cpue* was used in the analyses of the actual data. The iterative method discussed above (Equation 3 and following) was then applied to the actual data for various values

of ρ (0, 0.25, 0.5, 0.75 and 1) until convergence was achieved in terms of the slope statistics. Convergence was not obtained for ρ = 0.75 and 1.0, so a ρ value of 0.5 was selected as the basis for results to be input to resource assessments. This choice was likely to achieve as optimistic a trend in estimated abundance as possible with the methods applied. Although there is some concern that the correction method failed for higher values of ρ, the assumption of ρ = 0.5 reflects a fairly high degree of correlation, and it could be argued that any value greater than 0.5 is unrealistic.

This bycatch correction procedure was applied in practice over a subsequent period of three years. Each year, as additional data became available, the procedure was applied to update and to extend the standardized *cpue* index input to the algorithms used to provide management advice.

The results from applying the iterative procedure for correcting for the positive correlation observed between hake and bycatch *cpue* (assuming δ = 10% of the average hake *cpue* and ρ = 0.5) are shown in Table IV. The method converges fairly quickly in terms of the slope statistics, moving from an initial downward trend in resource abundance of some 0.4% per year (ignoring the correlation effect) to an increasing trend in resource
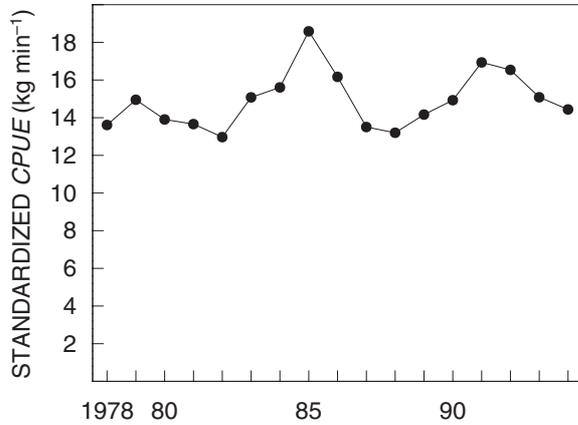
Fig. 5: West Coast hake standardized *cpue* corresponding to the final iteration in Table IV to adjust for the bycatch *cpue* correlation effect

abundance of about 0.6% per year.

The standardized *cpue* series obtained from the final iteration in Table IV is shown in Figure 5. The relatively small upward trend was inconsistent with a general perception by industry that there had been a substantial improvement in the resource over the 17-year period considered in this study, as suggested, for example, by the nominal *cpue* increase of approximately 4% per year. The difference between the extent of re-

Table IV: Results from the iterative procedure applied to the actual data to correct for the positive correlation observed between hake and bycatch *cpue*. The value $r^2$ indicates the proportion of the variance explained by the model and β is the slope statistic obtained from regressing ℓn (standardized *cpue*) against year. The model applied to standardize the hake *cpue* data is given by Equation 1. The case of ρ = 0 is synonymous with the basic approach, in which no correction was applied to the bycatch *cpue*

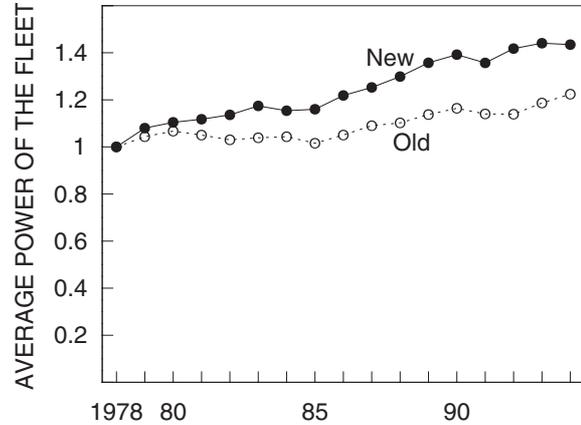| ρ | Iteration | $r^2$ (%) | Slope (%) | Bycatch *cpue* parameter estimate, γ (kg min⁻¹)⁻¹ | (Bycatch *cpue*)² parameter estimate, γ' (kg min⁻¹)⁻² |
|------|-----------|-----------|-----------|--------------------------|-------------------------|
| 0.0  |           | 29.5      | -0.37     | 0.000234                 | -0.000518               |
| 0.50 | 1         | 35.5      | 0.43      | -0.05466                 | 0.000444                |
|      | 2         | 35.2      | 0.62      | -0.06307                 | 0.000426                |
|      | 3         | 34.9      | 0.58      | -0.05037                 | 0.000056                |
|      | 4         | 35.3      | 0.61      | -0.06071                 | 0.000340                |
|      | 5         | 34.9      | 0.60      | -0.05083                 | 0.000055                |
|      | 6         | 35.4      | 0.58      | -0.06801                 | 0.000639                |
|      | 7         | 34.7      | 0.58      | -0.04861                 | 0.000049                |



Fig. 6: Average power of the West Coast fishing fleet, normalized to 1978. The "old" series refers to the case where effort is standardized by applying the power factors crudely calculated in the early 1970s, and the "new" series results from standardizing effort by applying the vessel factor estimates obtained from the final GLM (including the bycatch *cpue* correlation correction) presented herein

source recovery indicated by the GLM and the nominal trend arises mainly from three factors:

(i)   an increase in the average fishing power of the vessels in the fleet;
(ii)  vessels moving to deeper water to fish, where catch rates tend to be higher;
(iii) an increase in fish density in deeper water over time.

Only the first of these factors was taken into account (albeit roughly) in previous analyses of hake catch rate, when the power factors that had been crudely calculated in the past were applied to standardize the effort. These decreased the estimated annual rate of increase from 3.8 to 1.7%, but the plot in Figure 6 indicates that the extent of this correction was still too small. In the Figure, the effective average power of the vessels (normalized to 1978) is plotted over time by applying the equation:

$$P_y = \frac{\sum_v \left( f_v E_{v,y} \right)}{\sum_v E_{v,y}} \quad , \qquad (4)$$

where $P_y$ is the average power of the vessels in the fleet in year $y$, $f_v = e^{\lambda_{vessel}}$, where $\lambda_{vessel}$ is the vessel effect estimate obtained from fitting the model in Equation 1
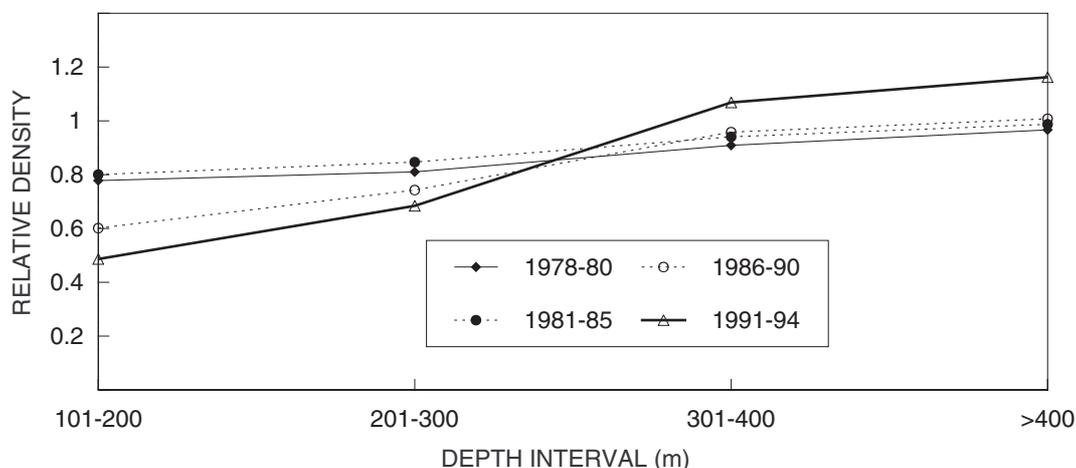
Fig. 7: Relative density of hake (as reflected by GLM-standardized *cpue*) on the West Coast at various depth intervals over a number of sub-periods

(denoted by the "new" series in Figure 6), or alternatively $f_v$ is the old power factor in the case where no GLM standardization was performed (denoted by the "old" series in Figure 6), and $E_{v,y}$ is the nominal fishing effort expended in year $y$ by vessel $v$.

The influence of the remaining two factors is apparent in Figures 7 and 8. There is a trend of increasing density with depth for all subperiods; however, with time, although hake density in deeper water in-
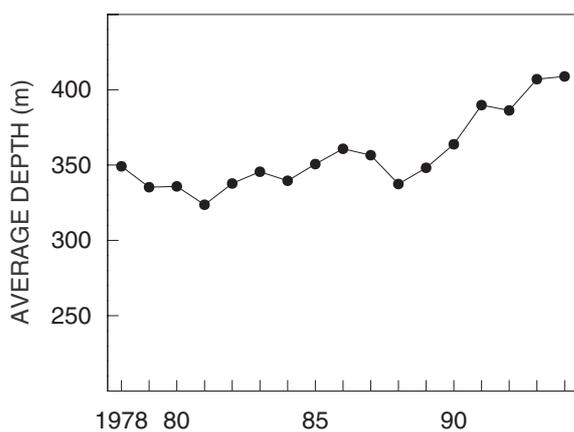


Fig. 8: Average depth at which fishing took place on the West Coast. Only hake-directed drags (refer to Table I) were considered in the construction of this plot
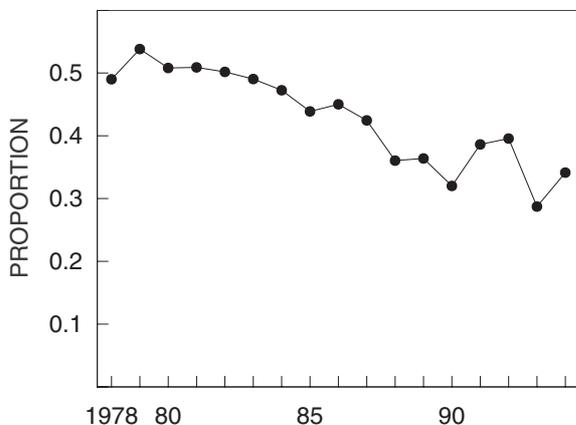
creased, the density in shallow waters (100–300 m) decreased (Fig. 7). The latter effect more than offsets the former when the relative areas of the strata are taken into account. Figure 8 indicates that the average depth of trawls made each year has increased over time. Taken together, the two depth effects contribute to a trend in nominal catch rate that is greater than the trend in resource abundance.

Figure 9 relates to hake abundance as indexed by standardized *cpue* as an index of density multiplied by area. The plot shows the abundance in the 200–300 m depth range expressed as a proportion of that in the 200–500 m range, plotted as a function of time. This proportion decreased from 0.49 in 1978 to 0.34 in 1994, further confirming the shift in distribution of hake towards deeper water over time.

The GLM-standardized *cpue* series also appears to be inconsistent with trends based on research surveys (summer and winter) of hake abundance (Fig. 10). The summer series for the full depth range trawled (0–500 m) reveals an increase (from fitting a log-linear model trend) in hake abundance of $6 \pm 5\%$ (1 *SE*) per year, whereas for depths >200 m there was an increase in abundance of $7 \pm 3\%$ per year. However, GLM-standardized *cpue* results show a smaller increase in abundance of $0.6 \pm 0.5\%$ (although the difference is not significant at the 5% level).

This discrepancy may be explained by examining the commercial catch-at-age data, which suggest that fewer young fish are being caught currently than in the late 1970s. Reasons for this are not clear, but probable

Fig. 9: West Coast hake abundance in the 200–300 m depth range as a proportion of hake abundance in the 200–500 m depth range
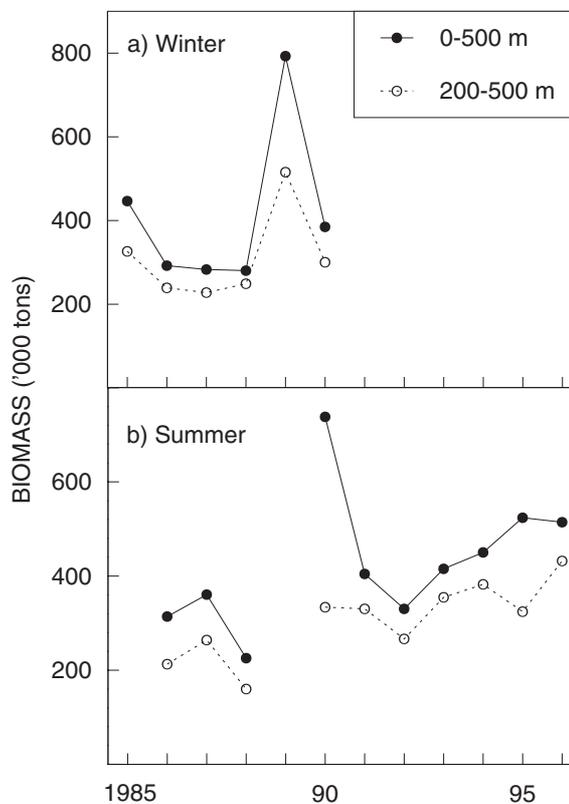


Fig. 10: Survey swept-area biomass estimates for West Coast hake in (a) winter – surveys were discontinued in 1990 and (b) summer – no survey took place in 1989. Biomass estimates are reported for the full depth range (0–500 m) and for depths from 200 to 500 m

contributory factors are that the fleet has moved to deeper water over time (hake size increases with depth – Payne 1995) and that net-liners are no longer being used. It seems likely that there was extensive illegal use of liners in the late 1970s through the 1980s and perhaps in the early 1990s. Whatever the reasons, the overall effect is that the *cpue* at present reflects a smaller component of the total biomass of the resource than was the case some 15–20 years ago. This would explain how the total biomass of the resource, as indicated by the survey results, could have increased faster than the exploitable component of the biomass indexed by the *cpue*. Consequently, assessment models no longer treat the *cpue* series as comparable throughout the full period from 1978 onwards, but rather assume a smooth change in catchability during the period 1984–1993 (Anon. 1997, Geromont and Butterworth 1997).

### FUTURE WORK

Now that statistical packages with appropriate capabilities are available, alternative error structure models should be explored, e.g. Poisson and gamma error models that avoid the problem of specifying $\delta$ and may better reflect the residual distribution. The analyses so far have assumed that the residuals obtained from the GLM have constant variance, i.e. are homoscedastic. If heteroscedasticity is present, an appropriate weighting procedure is needed to achieve minimum variance estimation, and this might be affected by a different choice for the distribution of the residuals.

Disaggregation of the catch data by size is, in principle, an approach to deal with the decreasing catch of younger fish over time as the extent of illegal use of liners has declined. There are, however, associated problems because the size categories (see Table I) used by the industry changed over time, which introduces comparability problems. A better understanding needs to be developed as to why, when the bycatch correlation correction method is applied, the goodness of fit $r^2$ increases with the assumed correlation $\rho$, whatever the true value of $\rho$. Ideally some criterion should be developed that would allow reliable estimation of $\rho$, if this method is to be applied in circumstances

where there is little other basis to guide a choice for ρ.

Although the bycatch correlation correction method went some way towards accounting for features in the data, it remains unsatisfactory, partially because of its failure to converge for higher values of ρ for the actual hake data. This has prompted further research into whether disaggregating bycatch species into those caught incidentally (because of their co-occurrence with hake) and those specifically targeted might better explain the positive correlation observed. This approach, which is detailed in Appendix 3, has achieved some success.

## ACKNOWLEDGEMENTS

## LITERATURE CITED

ANDREW, P. A. 1986 — Dynamic catch-effort models for the southern African hake populations. *Rep. Benguela Ecol. Progm. S. Afr.* **10**: 248 pp.

ANDREW, P. A. and D. S. BUTTERWORTH 1987 — Is $f_{0.1}$ an appropriate harvesting strategy for the Cape hake? In *The Benguela and Comparable Ecosystems.* Payne, A. I. L., Gulland, J. A. and K. H. Brink (Eds). *S. Afr. J. mar. Sci.* **5**: 925–935.

ANON. 1996 — Report of the bluefin tuna methodology session (Madrid, Spain – April 16 to 19, 1996). *Collve Vol. scient. Pap. Int. Commn. Conserv. Atl. Tunas* **46**(1): 187–212.

ANON. 1997 — Recommendation to the Sea Fisheries Advisory Committee for a provisional hake *TAC* for 1998. Unpublished Report, Marine and Coastal Management, South Africa **WG/05/97/D:H:49**: 4 pp.

BERGH, M. [O.] and A. BARKAI. 1996 — Hake abundance trends from catch rate and size structure data. Unpublished Report, Sea Fisheries Research Institute, South Africa **WG/11/96/D:H:39**: 63 pp.

CAMPBELL, R. A., TUCK, G. and T. NISHIDA. 1995 — Indices of abundance for southern bluefin tuna from analysis of fine-scale catch and effort data. Unpublished Report, Southern Bluefin Tuna Working Group Document **SBFWS/95/14**: 72 pp.

DOONAN, I. 1991 — Orange roughy fishery assessment, *CPUE* analysis – linear regression, NE Chatham Rise – 1991. New Zealand Fisheries Assessment Research Document. 48 pp.

GEROMONT, H. F. and D. S. BUTTERWORTH 1997 — Assessments of West Coast hake using an age-structured production model to provide a basis for simulation testing of a revised Operational Management Procedure. Unpublished Report, Sea Fisheries Research Institute, South Africa **WG/08/97/D:H:35**: 24 pp.

GLAZER, J. P. 1999 — The application of General Linear Modelling methods to estimate trends in abundance of the hake and rock lobster stocks off South Africa. M.Sc. thesis, University of Cape Town: 239 pp.

HONMA, M. 1974 — Estimation of overall effective fishing intensity of tuna long-line fishery. *Bull. far Seas Fish. Res. Lab.* **10**: 63–85.

ICCAT 1995 — *Report for Biennial Period, 1994-95, Part I (1994) – 2.* Madrid; International Commission for the Conservation of Atlantic Tunas: 283 pp.

KIMURA, D. K. 1981 — Standardized measures of relative abundance based on modelling log (c.p.u.e.), and their application to Pacific ocean perch (*Sebastes alutus*). *J. Cons. perm. int. Explor. Mer* **39**: 211–218.

PAYNE, A. I. L. 1995 — Cape hakes. In *Oceans of Life off Southern Africa,* 2nd ed. Payne, A. I. L. and R. J. M. Crawford (Eds). Cape Town; Vlaeberg: 136–147.

PUNT, A. E. 1992 — Selecting management methodologies for marine resources, with an illustration for southern African hake. In *Benguela Trophic Functioning.* Payne, A. I. L., Brink, K. H., Mann, K. H. and R. Hilborn (Eds). *S. Afr. J. mar. Sci.* **12**: 943–958.

PUNT, A. E. and D. W. JAPP 1994 — Stock assessment of the kingklip *Genypterus capensis* off South Africa. *S. Afr. J. mar. Sci.* **14**: 133–149.

ROEL, B. A., COCHRANE, K. L. and J. G. FIELD 2000 — Investigation into the declining trend in chokka squid *Loligo vulgaris reynaudii* catches made by South African trawlers. *S. Afr. J. mar. Sci.* **22**: 121–135.

SAS INSTITUTE INC. 1985 — *SAS Users Guide: Statistics, Version 5 Edition.* Cary, North Carolina; SAS Institute Inc.: 1290 pp.

STUTTAFORD, M. (Ed.) 2000 — *Fishing Industry Handbook. South Africa, Namibia and Moçambique*, 28th ed. Cape Town; Exbury Publications: 464 pp.

ZAR, J. H. 1984 — *Biostatistical Analysis*, 2nd ed. Englewood Cliffs, New Jersey; Prentice-Hall: xiv + 718 pp.

## APPENDIX 1

### Simulation testing of bycatch correction procedure of Equation 3

The following assumptions were made.

(i) The time trend in hake biomass is

$$B_{hake}(y) = 1 \times e^{\beta^* y} \quad , \qquad \text{(App. 1.1)}$$

where $B_{hake}(y)$ refers to the hake biomass in year $y$, $\beta^*$ is the true annual rate of biomass increase (values of 1 and 2% were investigated), and $y$ refers to year, where $y = 1\ldots17$.

(ii) Bycatch species biomass = 1 for all years.
(iii) Catchability coefficient for hake ($q_{hake}$) = 1.
(iv) Catchability coefficient for bycatch species ($q_{byc}$) = 0.5.
(v) Of a total effort $E = 1$ applied by each vessel each day, the fraction of effort exerted on bycatch species is

$$E_{byc} = [0.1 + \omega y]x^+ \qquad \text{(App.1.2)}$$

and on hake is

$$E_{hake} = 1 - E_{byc} \quad , \qquad \text{(App.1.3)}$$

where $x^+$ is a value drawn from the probability distribution of bycatch *cpue*s normalized to their mean, which is taken to reflect the variability in the extent of targeting of effort onto bycatch species (see Appendix 2).

Note that the value of $\omega = 0.007$ investigated, which reflects an approximate doubling of effort directed at bycatch over the full period considered, is roughly compatible with the actual data for the West Coast hake fishery. If $E_{byc}$ is >1, it is set to 0.95, and hence $E_{hake}$ is set to 0.05. The catchability coefficients are simply scalar multipliers to give relative abundance indices, so that their values are not critical to the analysis; round figures were therefore chosen for convenience.

(i) Hake catch for vessel-day $i$,
$$C_{hake} = q_{hake}B_{hake}E_{hake}e^{\varepsilon_i} \quad , \qquad \text{(App. 1.4)}$$

where $\varepsilon_i$ is drawn from $N(0, \sigma^2)$.

(ii) Bycatch species catch for vessel-day $i$,
$$C_{byc} = q_{byc}B_{byc}E_{byc}e^{\xi_i}, \quad , \qquad \text{(App.1.5)}$$
where

$$\xi_i = \rho\varepsilon_i + \sqrt{(1-\rho^2)}\phi_i \qquad \text{(App. 1.6)}$$

and $\phi_i$ is drawn from $N(0, \sigma^2)$, i.e. the log-residuals have a correlation coefficient of $\rho$ with those for the hake catch.

(iii) The observed hake *cpue* is given by
$C_{hake}/\text{total effort} = C_{hake}/E = C_{hake}/1 = C_{hake}$
(App. 1.7)

(iv) The observed bycatch *cpue* is given by
$C_{byc}/E = C_{byc}.$ (App. 1.8)

For various true values of $\rho$, 100 vessel-days of data were generated for each of the 17 years. A value of $\sigma = 0.4$ (Equations App. 1.4 and 1.6) was used in the data-generation process because, for a range of $\sigma$ values tested, this resulted in a relationship between the simulated hake and bycatch *cpue* most similar to that shown in Figure 4.

The GLM fitted in the simulation testing exercise was

$$\ell n(hake\ cpue + \delta) = \alpha + \beta y + \gamma(bycatch\ cpue) + \gamma'(bycatch\ cpue)^2 + \varepsilon \ ,$$
(App. 1.9)

where $\delta$ is a constant added to the hake *cpue* to allow for the occurrence of zero *cpue* (options of $\delta$ that were investigated were 5, 10, 20 and 50% of the mean *cpue* respectively), $\alpha$ a constant, $\beta$ the estimated year-trend parameter, $\gamma$ and $\gamma'$ are the bycatch *cpue* parameters.

For different assumed values of $\rho$, the iterative procedure described in the main text is applied to determine how well the true $\beta^*$ is estimated. The results (in the form of the regression correlation $r^2$ and $\beta$) pertain to assumed biomass increases of 1 and 2% per year (see Tables II and III for selected $\delta$ values). Although up to 85 iterations were performed, convergence was not obtained in some cases. Note that in this simulation exercise there was no need to adjust for vessel, spatial or seasonal affects, because these were not included in the data-generation process under the assumption that appropriate adjustments for them would have already been achieved by the GLM. The purpose of the simulation testing was to determine how successfully the effects of increasing hake biomass and increasing targeting on other species could be separated by the analysis.

## APPENDIX 2

### Developing a bycatch *cpue* distribution for simulation testing

For the process of generating a simulated dataset to test the validity of the method proposed for correcting for the positive correlation evident between hake and bycatch *cpue*, a distribution to reflect the actual bycatch *cpue* was required (see Appendix 1). From this distribution, a random number ($x^+$) was drawn to provide a specific "observation" for a vessel-day. This Appendix details the method used to develop this distribution and the selection criteria applied for accepting or rejecting the random number drawn.

The observed bycatch rates (per vessel per day) for each year were re-normalized so that their means were all 1, i.e.:

$$x_{y,i}^+ = \frac{x_{y,i}}{\frac{1}{n_y}\sum_j x_{y,j}} \quad , \qquad \text{(App. 2.1)}$$

where $x_{y,i}$ is the $i^{th}$ bycatch *cpue* observed in year $y$ and $n_y$ is the sample size in year $y$.

The $x_{y,i}^+$ values were rounded at the third decimal place to yield a finite number of possible values of $x^+$ between 0 and $x_{max}^+$. The $x_{y,i}^+$ values are henceforth denoted by $x^+$, because they will no longer be used in a year-dependent context, i.e. it is assumed that the $x^+$ distribution is year-independent.

The method used to generate random $x^+$ values corresponding to the empirical probability distribution is detailed below.

The frequency of occurrence of the $x^+$ values was determined, allowing the generation of a frequency distribution. The frequencies of occurrence were then normalized to a maximum of 1, by dividing each frequency of occurrence by the largest frequency of occurrence, yielding ($P(x^+)$). Random $x^+$ values were then generated from $U[0, x_{max}^+]$ and $\Phi$ from $U[0,1]$. These would then correspond to some $x^+$ and $P(x^+)$ in the normalized frequency distribution. The random value $x^+$ is accepted if $\Phi < P(x^+)$, otherwise new $x^+$ and $\Phi$ values were generated until $\Phi < P(x^+)$.

Linear interpolation was used to generate $P(x^+)$ values for all $x^+$ values between 0 and $x_{max}^+$ not represented by the actual data.

## APPENDIX 3

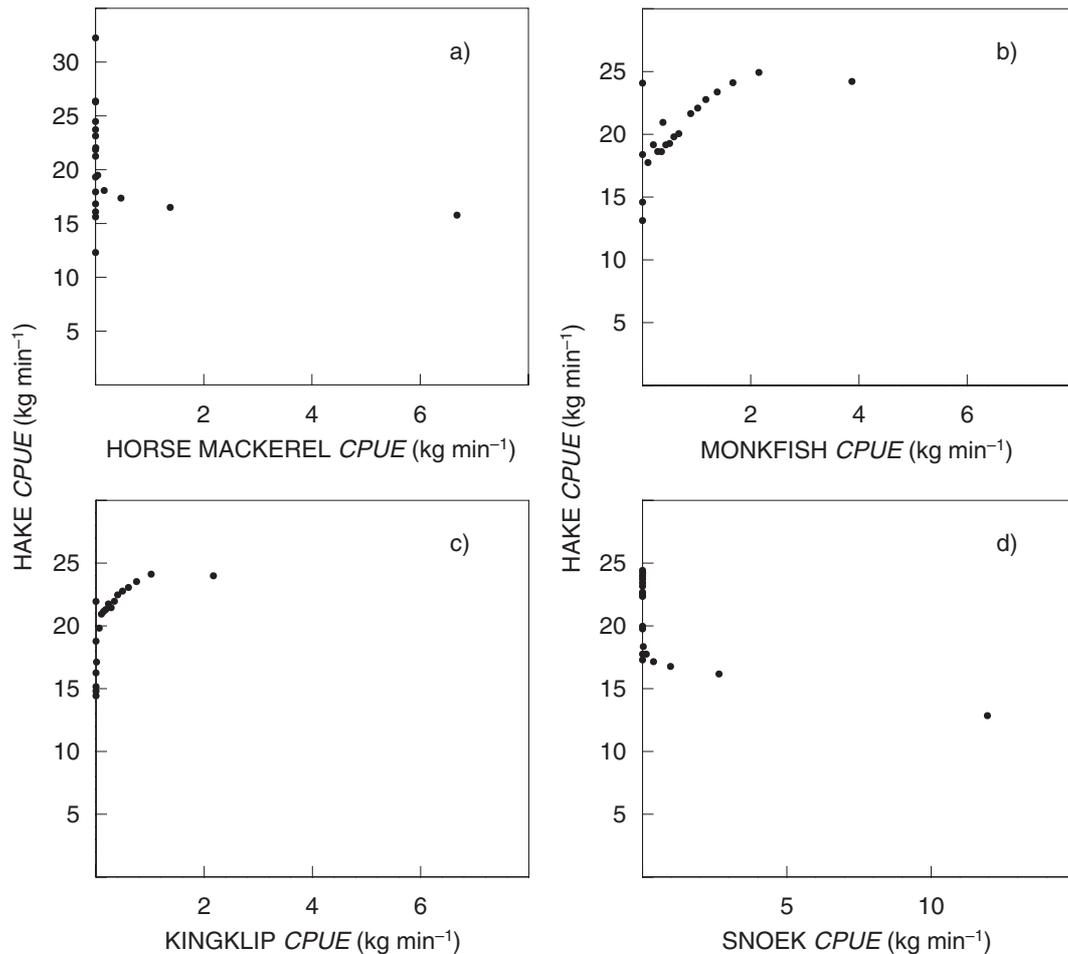### Consequences of differentiating between species in the bycatch



Fig. App. 3.1: Relationship between the *cpue* of hake and the main bycatch species

The bycatch *cpue* in Equation 1 consists of the sum of the catch of the "reported" bycatch species divided by total effort for a specific vessel-day. Given that these species are lumped, the shape of the net relationship shown in Figures 4a and b might arise from a combination of different relationships between hake *cpue* and the *cpue* of the bycatch species, where the associated correlations are positive in some cases and negative in others. Treating each bycatch species *cpue* separately

in the model is therefore investigated in an attempt to ascertain whether this can explain the overall effect.

Note that the analyses pertain to data for the period 1978–1997, whereas the bycatch iterative procedure was developed and tested on data for the period 1978–1994.

The "reported" bycatch species were separated into three categories, according to their importance in the bycatch relative to hake.

(1) Chokka squid, red squid *Toderopsis eblanae*, mackerel *Scomber japonicus* and sole *Austroglossus* spp. are not targeted by the offshore demersal fleet and the catches by such vessels are therefore negligible. They were therefore not considered for inclusion in the GLM.

(2) Monkfish *Lophius vomerinus* and kingklip are found on the hake grounds and are commonly taken as incidental bycatch in hake-targeted fishing. They are sometimes targeted in areas of relatively high abundance, but hake are still more abundant in the catches. Therefore, larger monkfish and kingklip catches do not necessarily reflect a transfer of fishing effort away from hake, and some positive correlation between these catches and those of hake would be expected (as indicated in Fig. App. 3.1), because of the influence of common environmental factors on catchability fluctuations.

(3) Snoek *Thyrsites atun* and horse mackerel *Trachurus trachurus capensis* form tight shoals and their catches are often associated with small catches of hake, because they reflect greater targeting of effort away from hake. It is therefore expected that, as the *cpue* of these bycatch species increases, the *cpue* of hake (in terms of total effort) decreases, i.e. a negative correlation (see Fig. App. 3.1).

The values of Spearman's rank correlation between the catches of hake and each of the bycatch species are shown in Table App. 3.I. They were all significant at the 95% level, and give credence to the categation of the species as indicated above.

It is concluded that it is appropriate to include only the *cpue* of snoek and horse mackerel as explanatory variables in the model, because there is clearly targeting away from hake in favour of those species, but

Table App. 3.I:    Spearman rank correlation coefficients between *cpue* of hake and those of individual "reported" bycatch species

| Bycatch species | Correlation coefficient |
|---|---|
| Horse mackerel | -0.104 |
| Snoek | -0.152 |
| Monkfish | 0.199 |
| Kingklip | 0.267 |

not necessarily any others. Furthermore, each is treated as a separate explanatory variable in the model rather than being lumped.

Equation 1 is therefore revised as follows:

$$\ell n(\text{hake } cpue + \delta) = \alpha + \beta_{year} + \omega_{season} + \eta_{depth} + \tau_{latitude} \\ + \lambda_{vessel} + \gamma(snoek\ cpue) \\ + \gamma(horse\ mackerel\ cpue) + interactions \\ + \varepsilon.$$

(App. 3.1)

An alternative of assuming a quadratic dependence of $\ell n(\text{hake } cpue + \delta)$ on snoek and horse mackerel is also evaluated.

The results of the analyses are shown in Table App. 3.II for the period 1978–1997. Cases where only snoek and horse mackerel are considered when correcting for bycatch lead to some increase in the estimated abundance trend compared to cases where all bycatch species are taken into account. The effect is small and not as large as suggested by the iterative procedure of the main text, with $\rho = 0.5$. However, given that this species split of the bycatch does seem to readily separate positive and negative correlations in the hake v. bycatch *cpue* relationship, it would be the preferable approach to adopt for *cpue* standardization.

Table App. 3.II:  Results obtained from various approaches of disaggregating bycatch species to obtain a standardized *cpue* index for the West Coast hake resource

| Model | $r^2$ (%) | Slope (%) | Parameter estimates | |
|---|---|---|---|---|
| | | | Bycatch | Bycatch[2] |
| Includes combined bycatch *cpue*[1] (quadratic[2]): $\rho = 0$ | 31.35 | -0.19 | 0.000448 | -0.00051 |
| $\rho = 0.5$ (i.e. applying the iterative correction procedure) | 36.66 | 0.29 | -0.055082 | 0.00015 |
| Includes *cpue* individually (linear): $\rho = 0$ | 31.76 | -0.06 | Snoek:   -0.025<br>Hm:       -0.015 | |
| Includes snoek and horse mackerel *cpue* individually (quadratic): $\rho = 0$ | 31.79 | -0.05 | Snoek:   -0.034<br>Hm:       -0.018 | Snoek:  0.00032<br>Hm:      0.00021 |
| Includes each bycatch *cpue* individually (linear): $\rho = 0$ | 32.96 | -0.24 | Snoek:   -0.023<br>Hm:       -0.016<br>Monk:   0.055<br>Kklip:    0.081<br>Mack:   -0.022<br>C squid:  0.158<br>R squid:  0.205<br>Ag sole:  2.335<br>Wc sole:  8.557 | |

Hm = Horse mackerel
Monk = Monkfish
Kklip = Kingklip
Mack = Mackerel
C squid = Chokka squid
R squid = Red squid
Ag sole = Agulhas sole
Wc sole = West Coast sole
[1] These figures differ from those in Table IV because the analyses here include data up to 1997
[2] The assumed dependence of $\ell$n (hake *cpue* + $\delta$) on bycatch *cpue*