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 computation 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 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.

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

* 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 7701, South Africa
Manuscript received May 2000; accepted September 2001
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:

\[
\ln(cpue+\delta) = \alpha + \beta_{\text{year}} + \omega_{\text{season}} + \eta_{\text{depth}} + \tau_{\text{latitude}} + \lambda_{\text{vessel}} + \gamma (\text{bycatch } cpue) + \gamma (\text{bycatch } cpue)^2 + \text{interactions} + \epsilon
\]

(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, \( \beta_{\text{year}} \) a year effect, where \( \text{year} \) is a factor with 17 levels (corresponding to 1978 to 1994), and \( \omega_{\text{season}} \) is a season effect, where \( \text{season} \) is a factor with four levels.
Summer = December–February
Autumn = March–May
Winter = June–August
Spring = September–November,

\[ \eta_{\text{depth}} \text{ is a depth effect, where depth is a factor with five levels:} \]
\[ d1 \leq 100 \text{ m} \]
\[ 100 \text{ m} < d2 \leq 200 \text{ m} \]
\[ 200 \text{ m} < d3 \leq 300 \text{ m} \]
\[ 300 \text{ m} < d4 \leq 400 \text{ m} \]
\[ d5 > 400 \text{ m}, \]

\[ \tau_{\text{latitude}} \text{ is a latitudinal effect, where latitude is a factor with four levels:} \]
\[ l1 \leq 31°00´S \]
\[ 31°00´S < l2 \leq 33°00´S \]
\[ 33°00´S < l3 \leq 34°20´S \]
\[ l4 > 34°20´S, \]

\[ \lambda_{\text{vessel}} \text{ is a vessel effect, where vessel is a factor with 110 levels (each vessel being uniquely identified by the vessel code recorded in the database), bycatch cpue is a continuous variable relating to the catch rate of bycatch species the sum of all “reported” bycatch species divided by total effort per vessel day, upon which the logarithm of hake cpue is assumed to depend quadratically, interactions refer to year × depth, year × latitude and depth × latitude interactions, and } \varepsilon \text{ is the error term, which is assumed to follow a normal distribution } N(0,\sigma^2). \]
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 \( r^2_\text{adj} = 1 - (1 - r^2)(n - 1)/(n - m) \), 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 \( r^2_\text{adj} \) 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 \( \delta \)

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 (\( \delta \)) to the \( cpue \) before taking logs (Campbell et al. 1995, ICCAT 1995).

As the scale for \( cpue \) values is essentially arbitrary, the choice of \( \delta \) must be linked to the scale adopted. This is usually achieved by defining \( \delta \) 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 \( \delta \) 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):

\[
\text{cpue}_y = \sum_{\text{strata}} \left[ \alpha + \beta \text{year} + \omega \text{season} + \eta \text{depth} + \gamma \text{vessel} + \delta \text{bycatch cpue} \right] + \epsilon + \text{interactions}
\]

where \( \text{cpue}_y \) is the standardized index of abundance for year \( y \), \( A_{\text{stratum}} \) the size of the stratum (e.g. depth 200–300 m, latitude 31–33°S), and \( A_{\text{total}} \) is the total size of the area considered (it is not strictly necessary to divide by \( A_{\text{total}} \) 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 \( (\frac{\delta}{2}) \), because if \( \log(x) \sim N(\eta, \sigma^2) \), then \( E(x) = e^{\hat{\gamma}(\hat{y} + \frac{\delta}{2})} \) (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.
(Bergh and Barkai 1996) derived a relationship between the \( \text{cpue} \) 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 \( \text{cpue} \) and bycatch \( \text{cpue} \), whereas the expected negative correlation is evident only at higher levels of such bycatch \( \text{cpue} \) (Bergh and Barkai 1996). Given the strength of this positive correlation in the dataset being analysed, use of bycatch \( \text{cpue} \) as an explanatory variable in the GLM would not allow proper adjustment for targeting bycatch. This was highlighted by the fact that the bycatch \( \text{cpue} \) parameter estimated in the GLM was a very small negative number, indicating, counter-intuitively, that very little hake \( \text{cpue} \) had to be forfeited in order to increase bycatch \( \text{cpue} \) (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 \( \text{cpue} \), after correcting for temporal and spatial effects. Specifically \( \ln(\text{hake \ cpue} + \delta) = \beta_{\text{year}} + \omega_{\text{season}} + \gamma_{\text{latitude}} - \lambda_{\text{vessel}} - (\text{year} \times \text{depth}) - (\text{year} \times \text{season}) - (\text{depth} \times \text{season}) \) is plotted against bycatch \( \text{cpue} \) (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 \( \text{cpue} \) 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 \( \text{cpue} \) in the same direction. Adjusting for bycatch \( \text{cpue} \) 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 \( \text{cpue} \) 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 \( \text{cpue} \) (as an index of abundance of bycatch species) were assumed to be correlated with those of the hake \( \text{cpue} \) (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 \( \text{cpue} \) was therefore adjusted as follows:

\[
\text{cpue}_{\text{hake}} e^{-\lambda_{\text{vessel}}} \times e^{-\rho_{\text{fis}}} = \text{cpue}_{\text{fis}} e^{-\lambda_{\text{vessel}}} \times e^{-\rho_{\text{fis}}}
\]
where $cpue^c_i$ is the corrected observed bycatch $cpue$ for vessel-day $i$, $cpue_o$ the observed bycatch $cpue$ for vessel-day $i$, $e^{-\lambda}$ a correction term for the vessel effect, $\epsilon_i$ the residual of the GLM fit to $\log(hake\ cpue + \delta)$ for vessel-day $i$, and $\rho$ 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 $\log(hake\ cpue) \times bycatch\ cpue$ 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 $\rho$.
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 $\log(standardized\ cpue)$ 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 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 $\delta = 0.1 \times mean\ hake\ cpue$, and $\delta = 0.5 \times mean\ hake\ cpue$ are shown for cases where the true annual rate of biomass increase $\beta$ is 0.01 or 0.02

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\beta = 0.01, \delta = 0.10$</th>
<th></th>
<th>$\beta = 0.02, \delta = 0.10$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho$</td>
<td>0.0</td>
<td>$\rho$</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.25</td>
<td>0.50</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>$\beta = 0.01, \delta = 0.50$</td>
<td>0.0</td>
<td>0.25</td>
<td>0.50</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td></td>
<td>1.0</td>
</tr>
</tbody>
</table>

| $\beta = 0.02, \delta = 0.50$ | 0.0           | 0.25 | 0.50           | 0.75 |
|           |                | 1.0 |                | 1.0 |

Stability of the slope statistic (obtained from regressing $\log(standardized\ cpue)$ 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).
RESULTS AND DISCUSSION

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

Given the bias indicated by the simulation results for larger values of $\delta$, 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 $\rho$ (0, 0.25, 0.5, 0.75 and 1) until convergence was achieved in terms of the slope statistics. Convergence was not obtained for $\rho = 0.75$ and 1.0, so a $\rho$ 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 $\rho$, the assumption of $\rho = 0.5$ 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 $\delta = 10\%$ of the average hake cpue and $\rho = 0.5$) 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 abundance of some 0.5% per year.

### Table III: Estimates of the slope $\beta$ (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 $\delta = 0.1 \times$ mean hake cpue and $\delta = 0.5 \times$ mean hake cpue are shown for cases where the true annual rate of biomass increase $\beta$ is 0.01 or 0.02

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\beta = 0.01, \delta = 0.10$</th>
<th>0.0</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>0.0</td>
<td>0.010</td>
<td>0.009</td>
<td>0.009</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>Assumed $\rho$</td>
<td>0.25</td>
<td>0.011</td>
<td>0.010</td>
<td>0.010</td>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>No convergence</td>
<td>No convergence</td>
<td>No convergence</td>
<td>No convergence</td>
<td>No convergence</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>No convergence</td>
<td>No convergence</td>
<td>No convergence</td>
<td>No convergence</td>
<td>No convergence</td>
</tr>
<tr>
<td>$\beta = 0.02, \delta = 0.10$</td>
<td>0.0</td>
<td>0.018</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
</tr>
<tr>
<td>Assumed $\rho$</td>
<td>0.25</td>
<td>0.019</td>
<td>0.018</td>
<td>0.018</td>
<td>0.018</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.020</td>
<td>0.019</td>
<td>0.019</td>
<td>0.019</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>No convergence</td>
<td>No convergence</td>
<td>No convergence</td>
<td>No convergence</td>
<td>No convergence</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>No convergence</td>
<td>No convergence</td>
<td>No convergence</td>
<td>No convergence</td>
<td>No convergence</td>
</tr>
<tr>
<td>$\beta = 0.01, \delta = 0.50$</td>
<td>0.0</td>
<td>0.006</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>Assumed $\rho$</td>
<td>0.25</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
<td>0.006</td>
</tr>
<tr>
<td>$\beta = 0.02, \delta = 0.50$</td>
<td>0.0</td>
<td>0.012</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
</tr>
<tr>
<td>Assumed $\rho$</td>
<td>0.25</td>
<td>0.012</td>
<td>0.012</td>
<td>0.012</td>
<td>0.012</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.013</td>
<td>0.012</td>
<td>0.012</td>
<td>0.012</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.013</td>
<td>0.013</td>
<td>0.013</td>
<td>0.013</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.013</td>
<td>0.013</td>
<td>0.013</td>
<td>0.013</td>
<td>0.013</td>
</tr>
</tbody>
</table>
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 resource 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:

\[ P_v = \frac{\sum (f_v E_{v,y})}{\sum E_{v,y}} \],

where \( P_v \) 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 parameter estimate obtained from fitting the model in Equation 1.

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 \( \beta \) 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 \( \rho = 0 \) is synonymous with the basic approach, in which no correction was applied to the bycatch cpue.

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>Iteration</th>
<th>( r^2 (%) )</th>
<th>Slope (%)</th>
<th>Bycatch cpue parameter estimate, ( \gamma ) (kg min(^{-1}))(^{-1} )</th>
<th>Bycatch cpue parameter estimate, ( \gamma ) (kg min(^{-1}))(^{-2} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.00</td>
<td>29.5</td>
<td>-0.37</td>
<td>0.000234</td>
<td>-0.000518</td>
</tr>
<tr>
<td>0.50</td>
<td>1</td>
<td>35.5</td>
<td>0.43</td>
<td>-0.05466</td>
<td>0.000444</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>35.2</td>
<td>0.62</td>
<td>-0.06307</td>
<td>0.000426</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>34.9</td>
<td>0.58</td>
<td>-0.05037</td>
<td>0.000056</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>35.3</td>
<td>0.61</td>
<td>-0.06071</td>
<td>0.000340</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>34.9</td>
<td>0.60</td>
<td>-0.05083</td>
<td>0.000055</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>35.4</td>
<td>0.58</td>
<td>-0.06801</td>
<td>0.000639</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>34.7</td>
<td>0.58</td>
<td>-0.04861</td>
<td>0.000049</td>
</tr>
</tbody>
</table>

Fig. 5: West Coast hake standardized cpue corresponding to the final iteration in Table IV to adjust for the bycatch cpue correlation effect.

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.
(denoted by the “new” series in Figure 6), or alternatively \( f \), 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 increased, 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 ± 5% (1 SE) per year, whereas for depths >200 m there was an increase in abundance of 7 ± 3% per year. However, GLM-standardized \( cpue \) results show a smaller increase in abundance of 0.6 ± 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
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 \textit{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 \textit{cpue}. Consequently, assessment models no longer treat the \textit{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, Gerromont 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 $\rho$.
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 $\rho$ 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

The constructive criticisms and suggestions made by Dr V. Restrepo (International Commission for the Conservation of Atlantic Tunas, Madrid, Spain) and an anonymous reviewer on an earlier draft are gratefully acknowledged.

LITERATURE CITED

The following assumptions were made.

(i) The time trend in hake biomass is

\[ B_{\text{hake}}(y) = 1 \times e^{\beta \ast y} \quad \text{[App. 1.1]} \]

where \( B_{\text{hake}}(y) \) refers to the hake biomass in year \( y \), \( \beta \ast \) is the true annual rate of biomass increase (values of 1 and 2% were investigated), and \( y \) refers to year, where \( y = 1 \ldots 17 \).

(ii) Bycatch species biomass = 1 for all years.

(iii) Catchability coefficient for hake \((q_{\text{hake}}) = 1\).

(iv) Catchability coefficient for bycatch species \((q_{\text{byc}}) = 0.5\).

(v) Of a total effort \( E = 1 \) applied by each vessel each day, the fraction of effort exerted on bycatch species is

\[ E_{\text{byc}} = 0.1 + \omega y \]

and on hake is

\[ E_{\text{hake}} = 1 - E_{\text{byc}} \quad \text{[App. 1.2]} \]

where \( x^+ \) is a value drawn from the probability distribution of bycatch \( cpue \) 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 \( E_{\text{byc}} > 1 \), it is set to 0.95, and hence \( E_{\text{hake}} \) 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 \( i \),

\[ C_{\text{hake}} = q_{\text{hake}} B_{\text{hake}} E_{\text{hake}} e^{\varepsilon_i} \quad \text{[App. 1.4]} \]

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

(ii) Bycatch species catch for vessel-day \( i \),

\[ C_{\text{byc}} = q_{\text{byc}} B_{\text{byc}} E_{\text{byc}} e^{\xi_i} \quad \text{[App. 1.5]} \]

where

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

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

\[ C_{\text{hake}} / \text{total effort} = C_{\text{hake}} / E_{\text{hake}} / 1 = C_{\text{hake}} \quad \text{[App. 1.7]} \]

(iv) The observed bycatch \( cpue \) is given by

\[ C_{\text{byc}} / E = C_{\text{byc}} \quad \text{[App. 1.8]} \]

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

\[ \ln(\text{hake } cpue + \delta) = \alpha + \beta y + \gamma (\text{bycatch } cpue) + \gamma' (\text{bycatch } cpue)^2 + \varepsilon \quad \text{[App. 1.9]} \]

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, and \( \gamma \) 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 \( \beta \ast \) 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.
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.:

\[
x^+_{y,i} = \frac{x_{y,i}}{n_y \sum_j x_{y,j}}
\]  

(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 \( x^+_{\text{max}} \). 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 \( U[0, x^+_{\text{max}}] \) 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 \( x^+_{\text{max}} \), not represented by the actual data.
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.
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.

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.

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 categorisation 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 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:

\[
\ln(\text{hake cpue} + \delta) = \alpha + \beta_{\text{year}} + \omega_{\text{region}} + \eta_{\text{depth}} + \tau_{\text{latitude}} + \lambda_{\text{vessel}} + \gamma_{\text{snoek cpue}} + \gamma'_{\text{horse mackerel cpue}} + \text{interactions} + \epsilon.
\]

An alternative of assuming a quadratic dependence of \(\ln(\text{hake cpue} + \delta)\) 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 \(p = 0.5\). 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>
<thead>
<tr>
<th>Bycatch species</th>
<th>Correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horse mackerel</td>
<td>-0.104</td>
</tr>
<tr>
<td>Snoek</td>
<td>-0.152</td>
</tr>
<tr>
<td>Monkfish</td>
<td>0.199</td>
</tr>
<tr>
<td>Kingklip</td>
<td>0.267</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Model</th>
<th>$r^2$ (%)</th>
<th>Slope (%)</th>
<th>Parameter estimates</th>
<th>Parameter estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bycatch</td>
<td>Bycatch$^2$</td>
</tr>
<tr>
<td>Includes combined bycatch cpue$^1$ (quadratic$^2$): $\rho = 0$</td>
<td>31.35</td>
<td>-0.19</td>
<td>0.000448</td>
<td>-0.00051</td>
</tr>
<tr>
<td>Includes combined bycatch cpue$^1$ (quadratic$^2$): $\rho = 0.5$ (i.e. applying the iterative correction procedure)</td>
<td>36.66</td>
<td>0.29</td>
<td>-0.055082</td>
<td>0.00015</td>
</tr>
<tr>
<td>Includes cpue individually (linear): $\rho = 0$</td>
<td>31.76</td>
<td>-0.06</td>
<td>Snoek: -0.025</td>
<td>Hm: -0.015</td>
</tr>
<tr>
<td>Includes snoek and horse mackerel cpue individually (quadratic): $\rho = 0$</td>
<td>31.79</td>
<td>-0.05</td>
<td>Snoek: -0.034</td>
<td>Hm: -0.018</td>
</tr>
<tr>
<td>Includes each bycatch cpue individually (linear): $\rho = 0$</td>
<td>32.96</td>
<td>-0.24</td>
<td>Snoek: -0.023</td>
<td>Hm: -0.016</td>
</tr>
</tbody>
</table>

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 $ln$ (hake cpue + δ) on bycatch cpue