# An initial investigation of the information content of sole catch-at-length distributions regarding recruitment trends

A. Ross-Gillespie and D.S. Butterworth[1]

email: *mlland028@myuct.ac.za*

**Summary**

A simple analysis of sole catch-at-length distributional data is carried out to estimate recruitment trends, assuming that availability and mortality do not change over time. Although the information content of these data is limited, they do provide give some indication of a period of generally below average recruitments during the first decade of the century, suggesting justification for expanding to a statistical-catch-at-length (SCAL) assessment in due course.

## Introduction

At present sole assessments are based on production models which aggregate over the length composition information available for the fishery and use only catches and the annual CPUE values for input. Such assessments could be extended to a full Statistical-Catch-at-Length (SCAL) approach, but this would require a fair amount of person-power resources to implement. Before doing so therefore, this initial and relatively simple analysis has been implemented to attempt to coarsely ascertain the information content of the sole CAL data in an assessment context.

The key feature of this simple approach is to assume that the primary driver of the dynamics is recruitment variability, so that it is attempting to estimate the trend in recruitment strength over time under the assumptions that quantities such as fishing mortality show much less variation over time.

## Data

Commercial catch-at-length data are available for the years 1995 to 2015. A growth curve is taken from Payne *et al.* (1986) for the area west of 24°E:

$$L_\infty = 54.739,\ \kappa = 0.176,\ t_0 = 0.089$$

## Methods

The population dynamics are given by:

$$N_{y+1,a+1} = g_a N_{y,a} \tag{1}$$

where

$N_{y,a}$ is the number of fish in age group $a$ at the start of year $y$ available to the fishery, and

$g_a$ is a function that combines the effects of mortality (both natural and fishing) and availability of age-group $a$ to the fishery, given by:

$$g_a = \begin{cases} g_a^{est} & \text{for } a_{min} \leq a \leq 4 \\ 1 & \text{for a=5} \\ g_5 e^{-\alpha(a-5)} & \text{for } 5 < a \leq a_{max} - 1 \end{cases} \tag{2}$$

[1]Marine Resource Assessment and Management Group, University of Cape Town, Rondebosch, 7701

where $a_{min}$ and $a_{max}$ are the minimum and maximum ages considered in the model, and $g_a^{est}$ and $\alpha$ are estimable parameters. Table 1 lists the values assumed for these ages.

For each year, the number of fish available to the fishery in the lowest age-group (expressed in relative rather than absolute quantities) is given by:

$$N_{y,a_{min}} = e^{\epsilon_y} \tag{3}$$

where $\epsilon_y$ are estimable parameters, i.e. recruitment to the fishery is modelled as fluctuations about a flat trend. If $y_0$ is the first year in the assessment (i.e. 1995), then $N_{y_0-1,a_{min}}$ is treated as an additional estimable parameter. Furthermore equilibrium is assumed in year $y_0 - 1$ so that $N_{y_0-1,a+1} = g_a N_{y_0-1,a}$, thus allowing an initial population structure to be calculated.

The model-predicted catch-proportions-at-age are calculated as:

$$p_{y,a} = \frac{N_{y,a}}{\sum_{a'} N_{y,a'}} \tag{4}$$

The model-predicted catch-proportions-at-length are derived from the catches-at-age by:

$$p_{y,l} = \sum_a p_{y,a} A_{a,l} \tag{5}$$

where $A_{a,l}$ is the proportion of fish of age $a$ to be found in the length group $l$. This matrix is calculated assuming that for each age $a$, the lengths are normally distributed about a mean given by the growth curve, i.e.

$$L_a \sim N\left[L_\infty\left(1 - e^{-\kappa(a-t_0)}\right); \theta_a^2\right] \tag{6}$$

The variance $\theta_a^2$ is given by:

$$\theta_a = \phi l_a \tag{7}$$

where $l_a$ is the expected length at age $a$, and $\phi$ is fixed at 0.05 for the results shown in this document.

If $p_{y,l}^{obs}$ are the proportions-at-length calculated from the observed catch-at-length data, then under the assumption of (possibly overdispersed) multinomial error distributions, which can be rendered approximately homoscedastic normal through a square root transformation, the likelihood contribution is given by:

$$-lnL = \sum_y \sum_l \left[ln\sigma_p + \frac{1}{2\sigma_p^2}\left(\sqrt{p_{y,l}^{obs}} - \sqrt{p_{y,l}}\right)^2\right] \tag{8}$$

where the variance parameter $\sigma_p$ is taken to be its maximum likelihood estimate from Equation 8:

$$\sigma_p = \sqrt{\frac{1}{\sum_y \sum_l 1} \sum_y \sum_l \left(\sqrt{p_{y,l}^{obs}} - \sqrt{p_{y,l}}\right)^2} \tag{9}$$

Furthermore, a penalty is added to the negative log-likelihood to stabilise estimation of the strength of cohorts towards the start and end of the time series which have not been sampled on many occasions; this is done by effectively assuming a prior of a normal distribution for each $\epsilon_y$:

$$pen = \sum_y \epsilon_y^2/(2\sigma_\epsilon^2) \tag{10}$$

where $\sigma_\epsilon = 0.2$ for the results in this document.

The analysis was performed in AD Model Builder (Fournier *et al.* 2012).

# Results and Discussion

Table 2 lists the estimates and the standard errors for the recruitment residuals. Figure 1 shows the fits to the catch-at-length data. This Figure also shows the model-predicted proportions-at-age. Figure 2 shows selected summary outputs from the model, as detailed in the Figure caption.

This analysis was intended as exploratory only in nature, so that sensitivities (e.g. to the values of $\phi$ and $\sigma_\epsilon$) have not yet been investigated. The catch-at-length data clearly contain limited information, as indicated by the small extent of variation over the years in the mean of these length distributions (see Figure 2(B)), though those means do nevertheless indicate some slight general reduction over time.

However the most important output is the relative recruitment estimates shown in Figure 2(E). These do seem to give some indication of a period of generally below average recruitments during the first decade of the century, which possibly links to the drop experienced in CPUE over that period. This signal in these length data would seem to indicate sufficient information content in those data to justify expanding to a SCAL assessment in due course.

# References

Payne, A.I.L. 1986. Biology, stock integrity and trends in the commercial fishery for demersal fish on the south-east coast of South Africa. PhD Thesis, University of Port Elizabeth. v+368 pp.

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielsen, A. and Sibert, J. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, 27(2), 233249.

**Table 1:** List of the model parameters and their descriptions.

| Parameter | Description |
|---|---|
| $a_{min}$ | Minimum age considered in the model, 4 years; this was chosen as the growth curve and catch-at-length distributions suggest minimal catch of sole of ages less than 4. |
| $a_{max}$ | Maximum age considered in the model, 9 years; the combination of growth curve and catch-at-length distributions suggest few sole caught of greater ages than this. |
| $l_{min}$ | Length minus-group; quantities and data below 25cm are pooled into this minus group so that no length group contains less than 2% of the total observations (summed across the years). |
| $l_{max}$ | Length plus-group; quantities and data above 42cm are pooled into this plus group. Note that the 40 and 41cm groups have additionally been pooled into a single group, so that this combined group meets the 2% threshold. |
| $g_4, g_5, \alpha$ | Estimable parameters for the $g$ function (Equation 2). |
| $\phi$ | The variance parameter for the growth-curve, fixed at 0.05 (Equation 7). |

**Table 2:** Estimates and standard deviations for $\epsilon_y$ (Equation 3).

| Year | Estimate | Standard error |
|---|---:|---:|
| 1994 | -0.090 | 0.169 |
| 1995 | -0.365 | 0.150 |
| 1996 | -0.094 | 0.147 |
| 1997 | 0.306 | 0.131 |
| 1998 | 0.284 | 0.135 |
| 1999 | 0.292 | 0.133 |
| 2000 | 0.025 | 0.135 |
| 2001 | -0.096 | 0.134 |
| 2002 | -0.007 | 0.134 |
| 2003 | -0.092 | 0.143 |
| 2004 | -0.280 | 0.145 |
| 2005 | -0.298 | 0.145 |
| 2006 | -0.243 | 0.133 |
| 2007 | 0.100 | 0.131 |
| 2008 | 0.254 | 0.130 |
| 2009 | -0.046 | 0.135 |
| 2010 | -0.236 | 0.135 |
| 2011 | -0.180 | 0.133 |
| 2012 | 0.015 | 0.134 |
| 2013 | 0.201 | 0.138 |
| 2014 | 0.577 | 0.149 |
| 2015 | -0.027 | 0.176 |

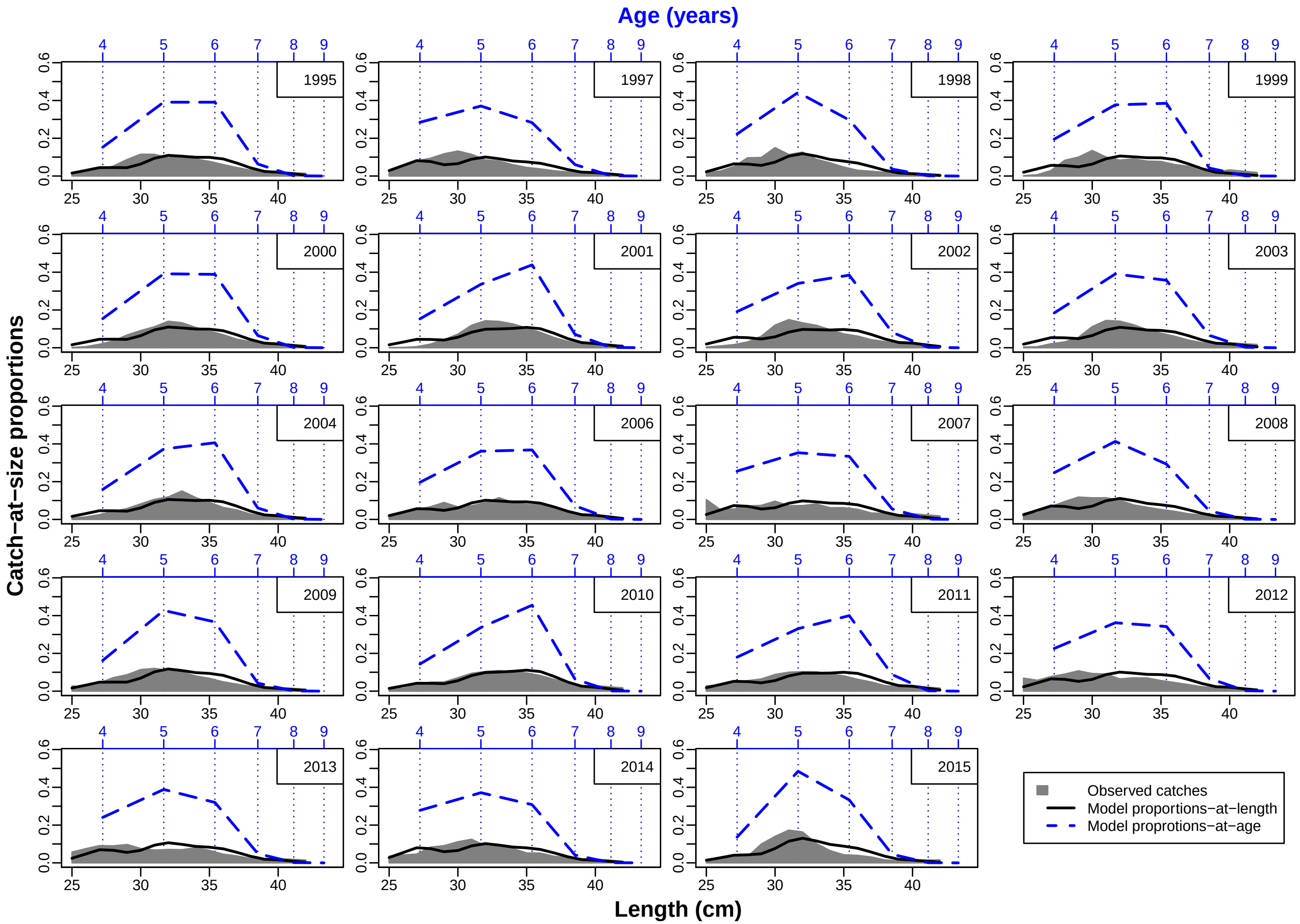


**Figure 1:** Fit to catch-at-length proportions. The grey shaded area shows the observed catch-at-length proportions, while the solid black lines show the model-predicted proportions-at-length (Equation 5). The lower (black) horizontal axis represents length in cm, while the upper (blue) horizontal axis represents age, where each age is located at the expected length from the growth curve equation. The dashed blue lines show the model-estimated proportions-at-age. Note that the observed and model-predicted proportions-at-length have been pooled below the length of 25cm (the minus group) and above the length of 42cm (the plus group), so that no length class has less than 2% of the total number of observations. Furthermore, observations from length groups 40 and 41cm have been combined into a single group to also meet this 2% threshold.

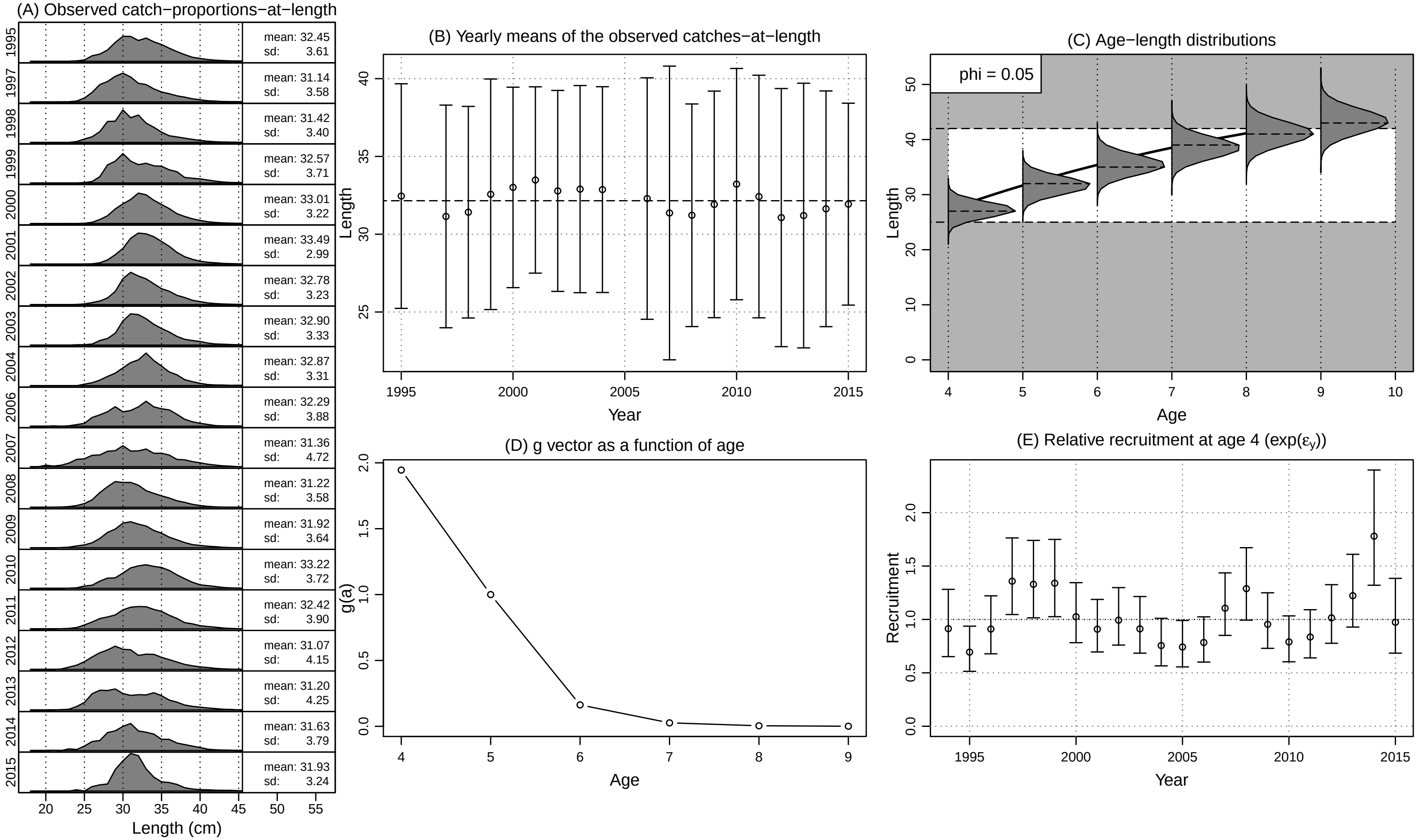


**Figure 2:** Selected model outputs. Panel (A) shows the (unpooled) observed catch-at-length proportions from Figure 1, and also shows the mean and standard deviation of the distribution for each year in the top right legend. Panel (B) then plots these means against time and shows the 95% probability intervals given by twice the standard deviation. Panel (C) shows the length distribution assumed in the model for each age group. The greyed-out area indicates the lengths that fall into the plus and minus groups. Panel (D) shows the estimated $g$ vector from Equation 2. Panel (E) shows the estimated recruitment values, given by the exponential of the residuals (Equation 3), along with their 95% confidence intervals given by twice the exponential of the standard errors estimated for $\epsilon_y$ (the $\epsilon_y$ estimates and standard errors are given in Table 2).