Single Hauls 1

This note gives an introduction to the basic tools for estimating the selectivity of trawl gear from data obtained by comparative (indirect) experiments.

Models

A **model** is a simplified description of a more complex system in a way that encompasses central aspects. These are mutually interacting and equipped with notions that allows for studying the system.

A **mathematical model** is a model that uses mathematical notions in its description of systematic behaviour in the system.

A **statistical model** is a mathematical model where deviations from the systematically behaviour of the system is described by use of appropriate random distributions.

A practical implication is that we will make a distinction between a curve and a model.

■

SELECT model (Share Each Length Class Total)

Length class ℓ fish arrive to the gear according to a Poisson process, with some mean value (rate-parameter). In statistical notation we write

$$\Lambda_\ell \sim Po(\lambda_\ell)$$

First consider a covered codend experiment and one length class only: ℓ
The fish may
- **either** be retained by the codend
- **or** escape through the codend and end up in the cover

The probability that a length class ℓ fish is retained by the codend is denoted π_ℓ. The probability that a length class ℓ fish ends up in the cover is therefore $1-\pi_\ell$. From SPR Theorem 1 it follows that the number of fish in the two compartments are described by Poisson distributions

- codend: $N_{\ell,\text{codend}} = \pi_\ell \cdot \Lambda_\ell \sim Po(\pi_\ell \cdot \lambda_\ell)$
- cover: $N_{\ell,\text{cover}} = (1-\pi_\ell)\cdot \Lambda_\ell \sim Po((1-\pi_\ell)\cdot \lambda_\ell)$

The total number of length class ℓ fish is $N_{\ell,+} = N_{\ell,\text{codend}} + N_{\ell,\text{cover}}$. (In this particular case $N_{\ell,+} = \Lambda_\ell$). Assume that the actual number of fish caught in the codend and cover was $n_{\ell,+}$ was $n_{\ell,+}$. By conditioning on the total catch we get (SPR Theorem 2)

$$\left(N_{\ell,\text{codend}} \,|\, N_{\ell,+} = n_{\ell,+}\right) \sim bi\left(\varphi_\ell ; n_{\ell,+}\right)$$

where φ_ℓ is the expected proportion of the total catch caught by the codend:

$$\varphi_\ell = \frac{\lambda_\ell \cdot \pi_\ell}{\lambda_\ell \cdot \pi_\ell + \lambda_\ell \cdot (1-\pi_\ell)} = \frac{\pi_\ell}{\pi_\ell + (1-\pi_\ell)} = \pi_\ell$$

(Again, this is a particular feature of the covered codend case).
We can now estimate π_ℓ by ML-technique:

$$l\left(\pi_\ell ; n_{\ell,\text{codend}}, n_{\ell,\text{cover}}\right) = c + n_{\ell,\text{codend}} \cdot \log(\pi_\ell) + n_{\ell,\text{cover}} \cdot \log(1-\pi_\ell)$$

If we maximise this function with respect to π_ℓ we get

$$\hat{\pi}_\ell = \frac{n_{\ell,\text{codend}}}{n_{\ell,\text{codend}} + n_{\ell,\text{cover}}} = \frac{n_{\ell,\text{codend}}}{n_{\ell,+}}$$

■

Next we consider a covered codend experiment and all length classes.
It is obviously intractable to fit a parameter for each length class. We will therefore be looking for means of summarising the selectivity across all length class by use of a few parameters.

A plot of the observed proportions clearly indicates that the retention probabilities grow with increasing length in a systematic way with some (random) deviations.

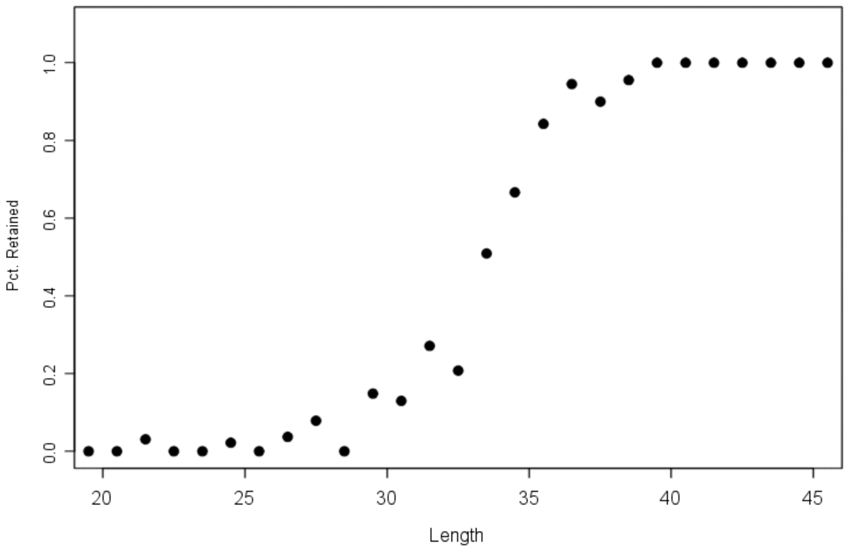

Figure 1: Observed proportions retained in the codend - Covered codend experiment

The sigmoid shape of the observed proportions suggests that the selectivity be described by a logistic function. There are however alternatives, which will be treated later.

$$r(\ell;\alpha,\beta)=\frac{\exp(\alpha+\beta\cdot\ell)}{1+\exp(\alpha+\beta\cdot\ell)}$$

The objective is now to find ML estimates of α and β. This is achieved by inserting $r(\ell;\alpha,\beta)$ into the log-likelihood function in place of π_ℓ and summing over all length classes:

$$\begin{aligned} l(\pi_\ell;n_{\ell,\text{codend}},n_{\ell,\text{cover}}) &= \sum_\ell n_{\ell,\text{codend}}\cdot\log(r(\ell;\alpha,\beta))+n_{\ell,\text{cover}}\cdot\log(1-r(\ell;\alpha,\beta)) \\ &= \sum_\ell n_{\ell,\text{codend}}\cdot(\alpha+\beta\cdot\ell)-n_{\ell,+}\cdot\log(1+\exp(\alpha+\beta\cdot\ell)) \end{aligned}$$

The maximum is however not expressible in a closed form. Consequently the MLE is found by use of a general optimiser. These are built in to various general purpose numerical software packages, such as Excel, Splus, genstat a.o.
An alternative approach to the estimation is to recognize that this is a GLM (Generalized Linear Model) for binomial data. Most statistical software packages have tools for fitting GLMs, with a range of different link-functions. The general SELECT model is however not covered by such tools. McCullagh & Nelder (1989) gives an accessible introduction to GLMs.

■

We now turn to the application of the SELECT method to data collected from experiments with paired gear (twin trawls, alternate hauls or trouser trawls). These gears typically consist of an assumed non-selective control codend and a selective test codend.
It is frequently observed that one codend is more efficient in catching fish than the other. This could be referred to variation in local abundance or by other unknown causes. The important point is however to recognise the potential difference and allow that to be reflected in the model. We will therefore introduce an intensity parameter for the test codend and denote it by p. The parameter is also often called the split parameter. It reflects the conditional probability that a fish goes into the test codend given that it contacts the combined gear.
By $r(\ell)$ we denote the retention probability for a length class ℓ fish to be retained by the test codend. Following the same approach as before, but omitting the middle steps we get the expected counts in the two codends:

- test: $p\cdot r(\ell)\cdot\lambda_\ell$
- control: $(1-p)\cdot\lambda_\ell$

from which the expected proportion in the test codend is derived:

$$\begin{aligned} \varphi(\ell) &= \frac{p\cdot r(\ell)\cdot\lambda_\ell}{p\cdot r(\ell)\cdot\lambda_\ell+(1-p)\cdot\lambda_\ell} \\ &= \frac{p\cdot r(\ell)}{p\cdot r(\ell)+(1-p)} \end{aligned}$$

Note that $\varphi(\ell)$ converge to p as ℓ gets larger.
Setting $r(\ell)$ to be the logistic function the expression becomes

$$\varphi(\ell) = \frac{p \cdot \exp(\alpha + \beta \cdot \ell)}{1 - p + \exp(\alpha + \beta \cdot \ell)}$$

Again this is inserted into the log likelihood function and the estimates of the parameters are found by maximisation.

■

IMPORTANT: $\varphi(\ell)$ is called the associated curve. It is this curve that is fitted to the data. Visual assessment of a fit should therefore be based how well $\varphi(\ell)$ fits the observed proportions. In the covered codend case $\varphi(\ell)$ coincides with $r(\ell)$. This is not the case for paired gear data.

Figure 2. below plots the proportions retained in the test codend for four sets of trouser trawl data. It is noted that the values show considerable fluctuation in the upper part of the length scale. This seems to be a frequently observed property of such data. Consequently the split parameter is often vaguely determined; i.e. the estimated standard deviation is relatively large.

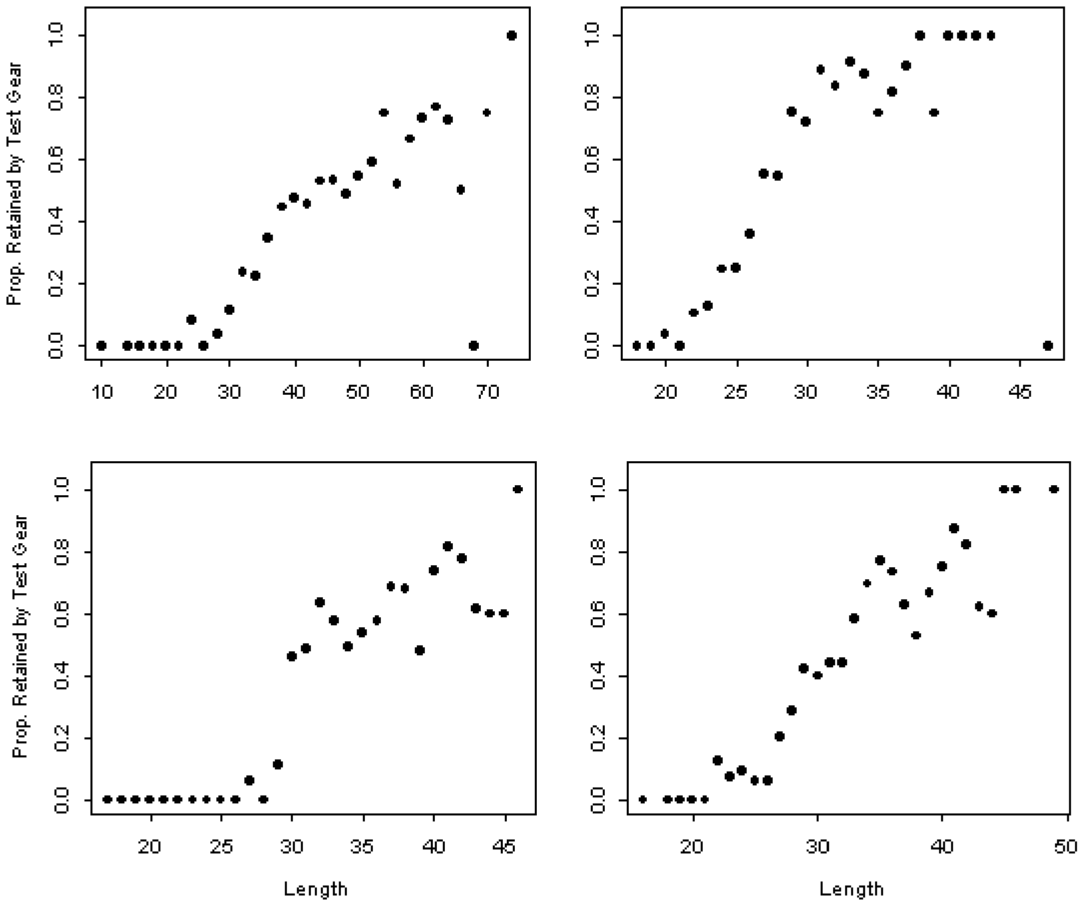

Figure 2: Observed proportions retained in the test codend - Paired gear experiments

Functional form of the selection curve:

The set of potential selectivity curves is not limited to fixed number of functional forms. In theory any increasing function can be used (but not necessarily estimated). However the following table represent a range of different shapes that has been found adequate for most cases. Furthermore all, except Richards curve, can be used as link functions, whenever the GLM case applies.

Curve	Functional form	L50	SR
Logit	$\frac{\exp(\alpha+\beta\cdot\ell)}{1+\exp(\alpha+\beta\cdot\ell)}$	$-\frac{\alpha}{\beta}$	$\frac{2\cdot\log(3)}{\beta}$
Probit	$\Phi(\alpha+\beta\cdot\ell)$*	$-\frac{\alpha}{\beta}$	$\frac{2\cdot\Phi^{-1}(0.75)}{\beta}$
Log-Log	$\exp(-\exp(-(\alpha+\beta\cdot\ell)))$	$\frac{-\log(\log(2))-\alpha}{\beta}$	$\frac{\log\left(-\frac{\log(4)}{\log(3/4)}\right)}{\beta}$
C-Log-Log	$1-\exp(-\exp(\alpha+\beta\cdot\ell))$	$\frac{\log(\log(2))-\alpha}{\beta}$	$\frac{\log\left(-\frac{\log(4)}{\log(3/4)}\right)}{\beta}$
Richards Curve	$\left(\frac{\exp(\alpha+\beta\cdot\ell)}{1+\exp(\alpha+\beta\cdot\ell)}\right)^{1/\delta}$	$\frac{\log\left(\frac{0.5^{\delta}}{1-0.5^{\delta}}\right)-\alpha}{\beta}$	$\frac{\log\left(\frac{0.75^{\delta}}{1-0.75^{\delta}}\right)-\log\left(\frac{0.25^{\delta}}{1-0.25^{\delta}}\right)}{\beta}$

* Φ is the Probability function for the standard normal distribution

Figure 3 plots the first four curves, with the values of L_{50} and *SR*. It is seen that logit and the probit are almost identical, except that probit is slightly more heavy tailed. Both of these curves are symmetrical, whereas log-log and c-log-log are both asymmetrical. Log-log has a heavy tail on the upper part and conversely c-log-log has a heavy lower tail.

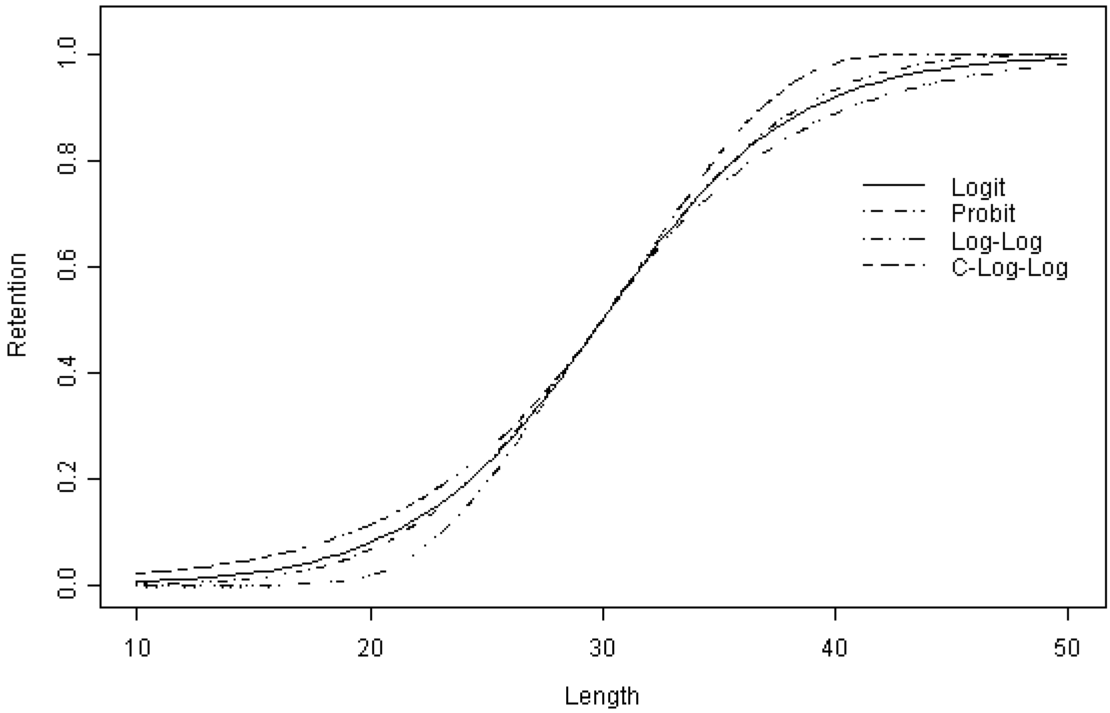

Figure 3: Curves for four different functions. L50=30, SR=9

Richards curve is an extended logit function by having an additional parameter $\delta > 0$. For $\delta = 1$ it is identical to the logit whereas it is asymmetrical for all other values. Figure 4 shows how the shape of the curve varies with various values of δ.

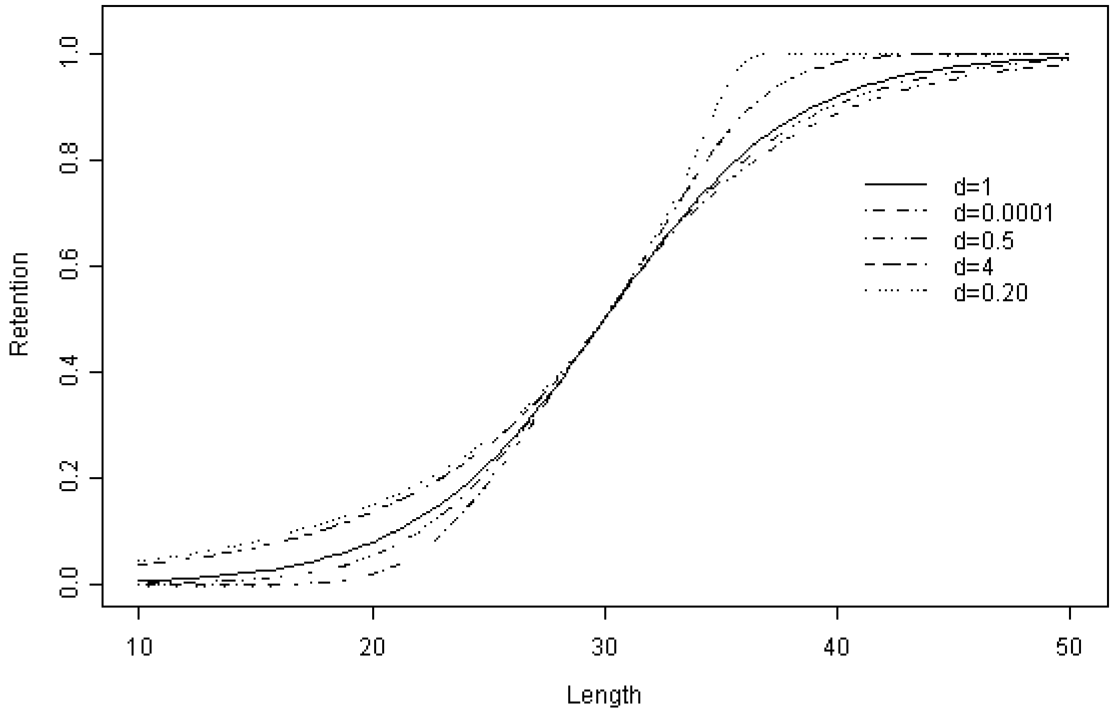

Figure 4: Richards curves with varying d-parameter. L50=30, SR=9

■

Scheme for general use of the SELECT method:

1. Assign a retention function to each selective device of the combined gear
2. Assign split-parameters (intensity) for each "fork"
3. Calculate, by use of the retention functions and split-parameters, the retention probabilities for each compartment
4. Calculate for each compartment the conditional probability that a fish was caught in this compartment given it was caught by the combined gear. These are the expected proportions.
5. Insert these conditional probabilities into the multinomial log-likelihood and find it's maximum
6. Derive the estimated selection curves from the estimated parameters.

This scheme may not always work. In particular if the model is over-parameterised it may be not be possible to obtain unique selection curves. Sometimes such cases can be remedied by additional constraints; e.g. the principle of geometrical similarity.

■

Conclusion: Proper Proportions