CCSBT-ESC/1009/29

What might be learned for Management Procedure purposes from a simple Moment based and Kalman Filter tuned model of the SBT.

John Pope
NRC (Europe) Ltd
The Old Rectory, Staithe Road,
Burgh St Peter, Beccles
Suffolk, UK. NR34 0BT
Email popejg@aol.com

Abstract

A minimally realistic moment based model of SBT is described and its fitting by means of the Kalman filter explored. Some preliminary results are shown but these are still at an early stage of development. Feedback on the approach would be welcome.

Introduction

This is a draft concept paper. It will certainly need developing further to be useful. It introduces the possibility of moment based assessment or a moment based OMP being applied to the SBT. Moment based assessment (Fournier and Doonan. 1987, Pope 2003) allows a parsimonious description of a fish stock that nevertheless allows a good deal of reality to be preserved. In the case of SBT it allows us to model the fact that the stock exists in a number of size based phases (a recruitment phase, young fish in the GAB phase, the Oceanic phase and the spawning phase) and that biomass progressively moves through these phases in a way we understand through our knowledge of growth and likely mortality rates.

Pope 2003 shows the linear annual recurrence relationships between the moments of the stock and hence between biological factors such as numbers and biomass. These can be very easily incorporated into a state space model of the fishery and this begs updating by some predictor corrector procedure such as the Kalman filter.

Thus the intention of this paper is

1. To provide a minimally realistic state space model of the SBT that describes the progression of SBT biomass though its size based phases.
2. To fit this model using the Kalman filter
3. To consider if this model might either act as an OMP in its own right or might alternatively help shed light on the best operation of simpler OMPs.

Model and Methods

The Protomoment model (Pope 2003) describes fish populations in terms of moments Ψ_i of their size distribution rather than in numbers at age or length, i.e as

$$\psi_i = \sum_{All\ lengths\ l} N_l l^i \qquad 1$$

Where l is length and N_l is number at length. Typically i = 0 to 4 since then Ψ_i provide the basis for calculating statistics of the mean, standard deviation, skewness and kurtosis of the size distribution. Clearly since Ψ_0 . = total numbers and Ψ_3 may be converted to total biomass under the assumption of isometric growth some of the Ψ_i do have clear biological meanings. Moreover, while Ψ_i for i=1,2 and 4 have less obviously useful biological interpretations (for example Ψ_1 is the distance the stock would extend if lined up nose to tail!) they are nevertheless useful for providing linear approximations to measures of interest such as spawning stock biomass, commercial catch per unit effort data (CPUE) or Aerial survey indices that are drawn from size dependent subsets of the total population. Thus the Ψ_i provide a useful and fairly adequate description of fish populations.

An additional advantage is that the Ψ_i can be updated using linear recurrence relationships. Using the well known linearization of annual growth and population change given by the Ford-Walford equation

$$L_{t+1} = L_t \exp(-K) + L_\infty(1-\exp(-K)) \qquad 2$$

Where K and L_∞ are the parameters of the Von Bertalanffy growth function.

and by the reverse Pope Cohort equation

$$N_{t+1} = N_t \exp(-M) - C_t \exp(-M/2) \qquad 3$$

Where N_t are stock numbers of a size (or age) group in year t (above some age of recruitment r) and C_t are the equivalent catch numbers and M the natural mortality rate.

Multiplying the RHS and LHS of equations 3 by powers (i=0:4) of the RHS and LHS of equation 2 and summing over all lengths we immediately see that with equation 2 raised to power 0 that

$$\Psi_0(t+1) = \Psi_0(t)\exp(-M) - X_0(t)\exp(-M/2)) + R(t+1). \qquad 4$$

Where R(t+1) are the number of recruits in year t+1.

We can also see with power 1 that

$$\Psi_1(t+1) = \exp(-K)\{\Psi_1(t)\exp(-M) - X_1(t)\exp(-M/2)\}$$

$$+ L_\infty(1-\exp(-K))\{\Psi_0(t)\exp(-M) - X_0(t)\exp(-M/2))\}$$

$$+ R(t+1)l_r \qquad 5$$

Where $X_i = \sum_{All\ lengths\ l} C_l l^i$ 6

and where l_r is the fish length at the time of recruitment.

Similar linear equations can be written for Ψ_i i=2 to 4 by squaring ,cubing or raising equation 2 to the i^{th} power and by multiplying by equation 3 and summing over all lengths

We thus arrive at the linear matrix update of the protomoments

$$\Psi_I = G^*\{\Psi_i \exp(-M) - X_i \exp(-M/2)\} + R(t+1)\,Lr \qquad 7$$

Where G is 5*5 the growth matrix whose terms $g_{i,j}$ are given by
$g_{i,j}$ = o if j>i,
$g_{i,j} = (i-1)!/(i-j)!*(j-1)!L_\infty(1-\exp(-K))^{(i-j)}*\exp(-K)^{(j-1)}$, and
Lr is the vector of lr to the powers 0 to 4.

In principle natural mortality that varies with age (as is usually postulated for SBT)might be introduced using approaches suggested in Pope 2003 but in practice it is far easier to use constant M and we do so here. M values of 0.1, 0.15 and 0.2 were considered.

In order to use the Ψ_i to provide predictors of those size dependent variables such as the SSB, the long line CPUE and the Aerial Survey indices that are used in SBT assessments we may make linear transformation that approximate maturity ogives and selection curves. The table below gives the adopted linear transformations SSBLT, CPUELT and ASLT that provide these three measures while Figure 1 shows the approximations to the size based functions these achieve. {N.B it is the ability of these transforms to estimate SSB etc. that is our main concern rather than their ability to precisely match a maturity or selection curve. This is an easier task since typically fish populations exist over a range of sizes and thus the negative values seen in figure 1 typically balance out with values from other parts of the size range.} Thus

$$SSB(t+1) = SSBLT^*\Psi_i \qquad 8$$

Note that LTSSB is the row vector of the transformation.

Text table Linear Transformations of Protomoments to provide predictors					
	PM0	PM1	PM2	PM3	PM4
SSBLT	0.00E+00	5.74E+03	-1.54E+02	6.93E-01	3.91E-03
CPUELT	5.37E+00	-2.18E-01	2.93E-03	-1.52E-05	2.68E-08
ASLT	0.00E+00	-1.89E+04	5.90E+02	-3.85E+00	6.85E-03

Given the linearity in the system (equations 7 and 8) the protomoment model is well suited to being framed as a state vector model. This suggests a Kalman filter approach to tuning the model to the forms of abundance data available for the SBT.

However to form a complete state vector model we need someway to predict recruitment.

Here we assume a Beverton and Holt S/R relationship

$R(t) = 1/\{(a/SSB)+b\}$ 9

Alas this does introduces non linearity into the system which we overcome using the extended Kalman filter which uses the exact (non) linear state equations (including equation 9) but updates the variance matrices and forms the Kalman gain matrix using the Jacobian of the update equations.

The state vector (SV) we adopt for the Kalman filter protomoment model is thus formed of the 11 element column vector,(' denoting the transposes of vectors)

$SV = [\Psi_I', a; b, q_{CPUE}, q_{AS}, R, SSB]'$ 10

Where q_{CPUE} and q_{AS} are the catchability (calibration) terms for the CPUE and Aerial surveys respectively and these together with the stock recruitment terms a and b are taken to be remain constant between years as they are updated by the process equations but are allowed to change in the light of data. i.e. in Kalman talk a(t|t-1)=a(t-1|t-1) etc.

The formulation of the Kalman filter follows the Wikipedia pages and notation (http://en.wikipedia.org/wiki/Kalman_filter

In particular X_i is treated as the control variable and for the historic period is estimated from the catch history of the SBT. (as may be seen from figure 10 the total catch weight derived from cf*X_3 does not yet match the catch record and needs to be improved in later versions of this report). At present, when in predictive mode, X_i is taken as the product of a constant harvest rate diagonal matrix H_i with the Ψ_i

i.e. $X_i(t) = \Psi_i(t) * H_i(t)$. 11

H_i for the predictive period is estimated for the average values of recent years that are then modified by a constant multiplier designed to steer the SSB to 20% of its unexploited level by 2035. So far such updates have been made annually but the same approach might be adopted for 2 or 3 year quota periods. Other strategies than constant harvest rate over the 2011-2035 period could also be explored. In OMP mode the historic period would be extended as data accrued and the predictions out to 2035 revised in the light of the updated state vector.

For the Historic period 1951-2009 the Kalman filter first updates SV(t-1|t-1) (its value given data up to year t-1) to that for year t (SV(t|t-1). This first update is based only upon the process using the deterministic process defined by equations 7 thro. 9 together with the constant updates of q_{CPUE} and q_{AS}, a and b.

SV(t|t-1) is then modified by proportions of the innovations (the differences between the actual and predicted observations of CPUE(t) and AS(t)) specified by the Kalman Gain matrix to give SV(t|t).

Note that the data prediction equations are non linear

$$CPUE(t) = q_{CPUE} * LTCPUE * \Psi_i(t)$$
$$AS(t) = q_{AS} * LTAS * \Psi_i(t)$$

And thus their Jacobian is adopted in the calculation of variance and of the Kalman gain matrix.

In prediction mode only the deterministic update can be made. In order to achive the target 20% of the unfished SSB by 2035 the multiplier of the harvest rate is modified by an iteratively estimated multiplier.
The covariance matrices of the initial state, the deterministic update and the data matrix were estimated using assumptions of constant coefficients of variation of the various elements. These were chosen to favour the process rather than the data inorder to keep stability of the critical parameter estimates a, b, q_{cpue} and q_{as}. Since these were initially unknown they were tuned so that their initial values matched the average values estimated over the period for which CPUE and AS data were available. So far this tuning has been handraulic but I believe it could be fairly simply achieved by an iterative process. Other approaches to estimating initial values of a, b, q_{cpue} and q_{as} might be a backward smoothing using the reversed Kalman filter or by endeavoring to reduce the bias and variance of the innovations.

Using assumptions about how the harvest rates are linked it is possible to simply solve the system for the steady state SSB and yield at various harvest rates and selection factors. Simple matrix equations for these purposes can be found in Pope 2003.

Results

Basis of fits

Results shown are those found using the average a, b, q_{cpue} and q_{as}. parameter value tuning described above and a natural mortality rate M of 0.2.

Fit to data

Figures 2 shows the fit of CPUE(t|t-1) and CPUE(t|t) to the observed CPUE series. Given the choice of covariance matrices that set observation error higher than process error the fitted lines are relatively stiff and while the (t|t) points move further toward the data than the (t|t-1) points they do not follow slavishly. Figure 3 shows simlar results for the shorter Aerial survey (these were given equivalent precision to the CPUE series)..

Figure 4 shows the innovations (differences between estimated and actual data values) fior each year for which CPUE and/or AS data existed. These seem reasonably balanced about zero though there is perhaps evidence of trend in the early years of the CPUE data which might be eliminated with alternative tuning approaches for the a, b, q_{cpue} and q_{as}.

Plots of a relative index of the 4 key parameters a, b, q_{cpue} and q_{as} are shown in figure 5. These suggest negative correlation between q_{cpue} and q_{as} and possibly the a parameter of the S/R relationship. The b parameter (inverse Rmax) is rather stable through the time period. The parameters of the S/R relationship suggest an steepness of 0.47 both at the beginning and end of the historical period. Higher steepness values may be obtained with alternative approaches to tuning the initial a, b, q_{cpue} and q_{as}.

Assessment estimates

Figure 6 shows the SSB both for the historical period and for predictions out to 2035. The later were made both to achieve the target and with zero harvest rates. In this realization the target is only achieved with near zero catches so the zero H line is overlaid. Again the process depicted is rather stiff due to the relative choice of process and observational error made.

Figure 7 shows the recruitment estimated in the historical period and for the projections to achieve the SSB target in 2035. Clearly the model has reappraised recruitment between 1980 and 1990 presumable resulting from the dip in the a and b parameters seen in figure 5 in that period.

Figure 8 shows the estimates of harvest rate by protomoment. In general harvest rate on numbers H_0 is considerably higher than the harvest rate on biomass H_3. However, all Hs are fairly correlated and peak in the mid 1980’s and again in the period 2000-2010. They become zero in the prediction period post 2010 because this reduction was needed to meet the SSB target in this pessimistic realization. Figure 9 shows these same H results expressed relative to the H_3 value. Figure 10 shows the catches estimated from the X3 estimates. Note that these do not yet match the OMP groups catch series and will need to be modified.

Steady state yield and steady state SSB.

Figure 12 shows the steady state yield for a range of harvest rates and selections. Selections here are taken as the ratio of H0 to H3 (other Hi) being interpolated from these values. Note that since $H_i = X_i / \Psi_i$, then

selection $= H1/H3=(X1/(X3*cf))* ((\Psi_3*cf)/ \Psi_1)$

=Average fish wt in the sea/Average fish weight in the catch.

This then is a rather understandable measure of selection. However the high values seen for SBT may suggest the model is accumulating too many "paper" old fish. The figure suggests that yield is higher at lower selections but note that these lower values may not be achievable. The general form of cross sections of this surface for given levels of selection is of a Fox type shape.

Figure 13 shows the SSB at various levels of harvest rate and selection. When harvest rate is zero it is the same of course regardless of the selection but for positive harvest rates the SSB achieved is obviously higher when larger fish are harvested (ie when the selection ratio is low).

Figure 13 attempts (rather unsuccessfully in this realization) to depict both the temporal trajectory of catches across the harvest rate –selection grid and to show this in relation to the yield surface. The intention of this figure was to illustrate that one of our problems with estimating productivity of this stock might be changing selection

Discussion –Conclusions

The present paper is intended as a first description of a possible form of MP. Clearly to become operational it will need to be developed in a way that can be tested against the full OMP test data grid.

This first development has been encouraging in that the model has proved easy to develop to this stage. Apart from three non-linear equations {those needed for the stock recruitment relationship and for estimates to be made of the CPUE and the Aerial survey indices} the model is entirely linear and most constants can be based upon known biological processes. Thus many constructs can be developed with a few matrix equations. The steady state yield and SSB calculations are an example of this. For example the following MATLAB code estimates both steady state yield and steady state SSB for a given harvest rate matrix HA.

```
IGH=eye(size(G))-G*em+G*HA*em2;
PMper=inv(IGH)*Lrec;
PMrec=PMper*(1-a/(SSBLT*PMper))/b;
yield(h1,h3+1)=CF*PMrec(4,1)*HA(4,4)/1000;
SSBg(h1,h3+1)=CF*SSBLT*PMrec/1000;
```

{NB eye is the identity matrix, inv is the matrix inverse, em is Exp(-M) and em2 is Exp(-M/2).}

Thus the workings of the model are often mathematically explicit.

Fitted with a Kalman filter the model produces robust and seemingly sensible results. Those shown in the figures are somewhat pessimistic but this is a consequence of the particular tuning approach used which converged towards relatively low steepness. Using a similar approach but adopting a lower initial ‘a’ parameter value leads to solutions with higher steepness. Figure 14 shows the SSB from one such run with initial steepness of 0.58 and final steepness of 0.55. The trick to keeping the equivalent tuning seems to be to have a smaller initial SSB with a more productive stock. Alternative tunings might be based upon reducing the bias and variance of the innovation estimates.

The ease of running the model should mean that it may prove useful for exploring matters of concern to the group. Examples might be the consequences of early and late pain strategies or factors influencing stock production.

I do not at the present time have scope to develop this as a full candidate OMP (and indeed you may prefer not to have any more) but if people feel that this might be useful I could endeavor to produce some thing that works for the ESC. If developed it will just have to be named JCL (Johnny Come Lately)!

References

Fournier, D. A., and Doonan, I. J. 1987. A length-based stock assessment method utilizing a generalized delay-difference model. Canadian Journal of Fisheries and Aquatic Sciences, 44: 422-437.

Pope, J. 2003. Golden ages or magic moments?. Natural Resource Modeling, 16: 439-464.

Figures

Figure 1 approximate ogives resulting from the linear transformations used to estimate predictors of the Aerial survey index, the CPUE index and the SSB from the population protomoments.

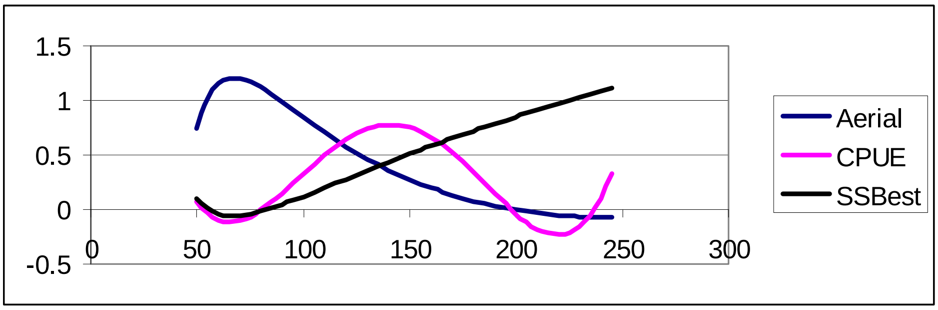

Figure 2 CPUE observations and the predictors based upon previous years data (k|k-1) and process and the estimate when the current years data is included (k|k).

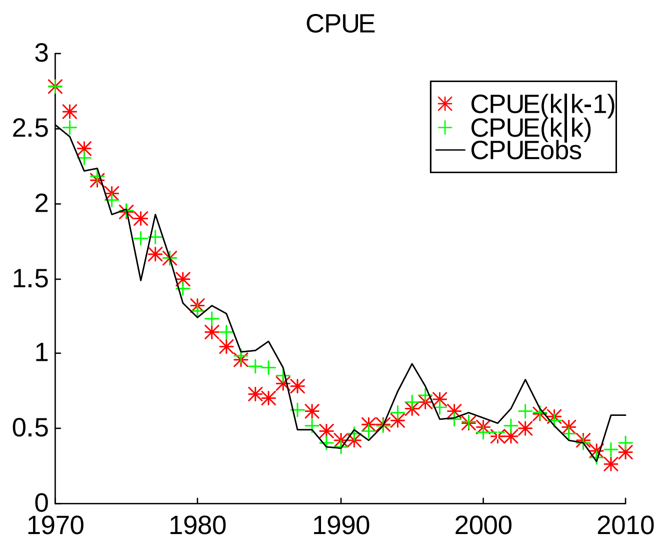

Figure 3 Aerial survey observations and the predictors based upon previous years data and process (k|k-1) and the estimate when the current years data is included (k|k)

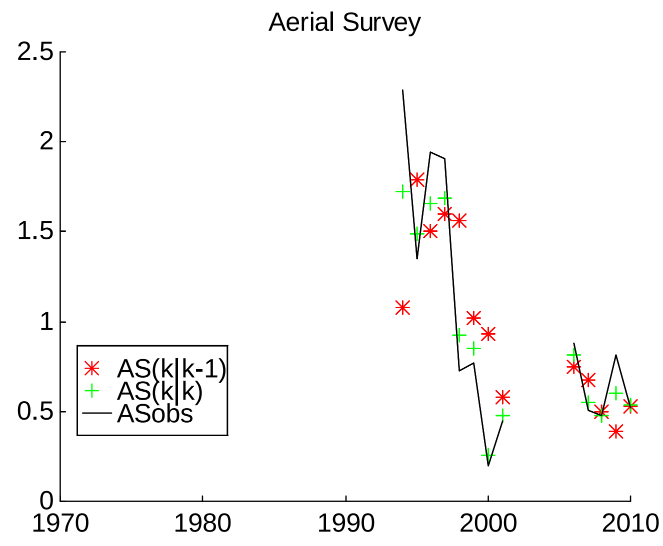

Figure 4 The innovations of the CPUE and the Aerial survey - (N.B.Innovations are Kalman Filter talk for the difference of the prediction of the data based upon previous years data and the process update and the data for the year)

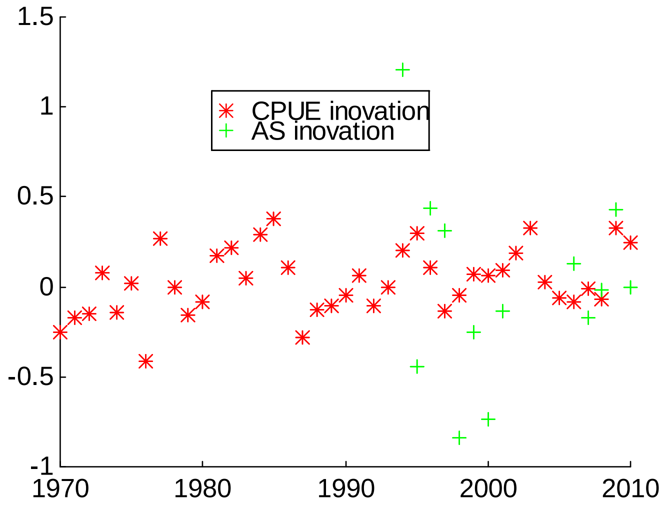

Figure 5 Relative changes in estimates of catchability and stock recruitment parameters. NB these were tuned to set initial values equal to the average for the periods with data.

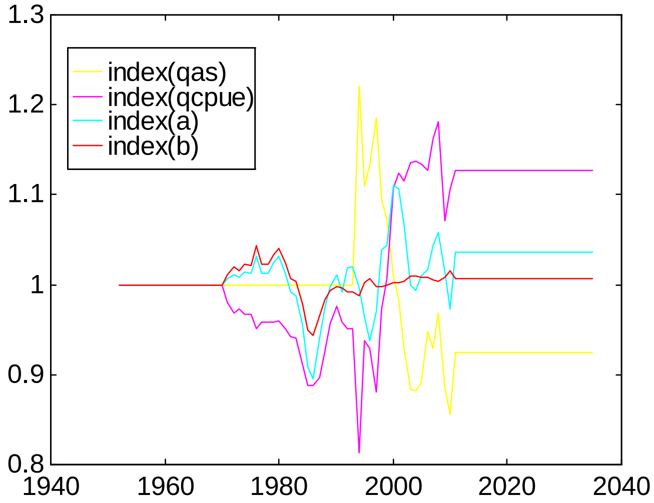

Figure 6 Estimates of SSB from process and data update estimates together with predictions out to 2035 with zero fishing mortality (hidden) and with mortality scaled down to meet a target of 20% of unfished SSB. (units tonnes)

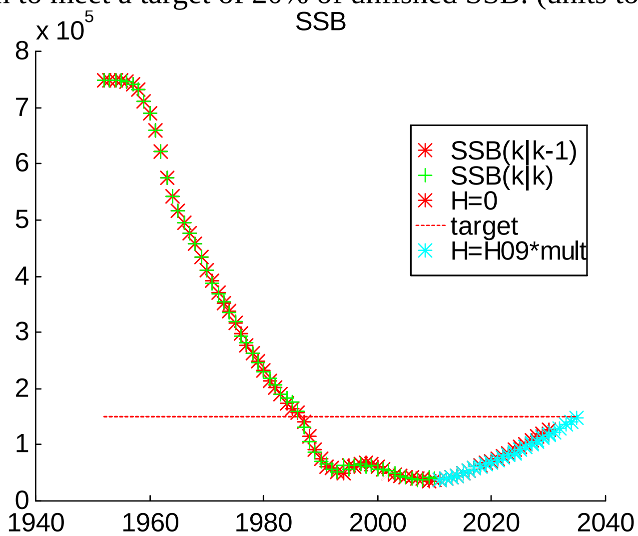

Figure 7 Estimates of recruitment from process and data update estimates together with predictions out to 2035 with zero fishing mortality and with harvest rate scaled down to meet a target of 20% of unfished SSB. (units 1000?check)

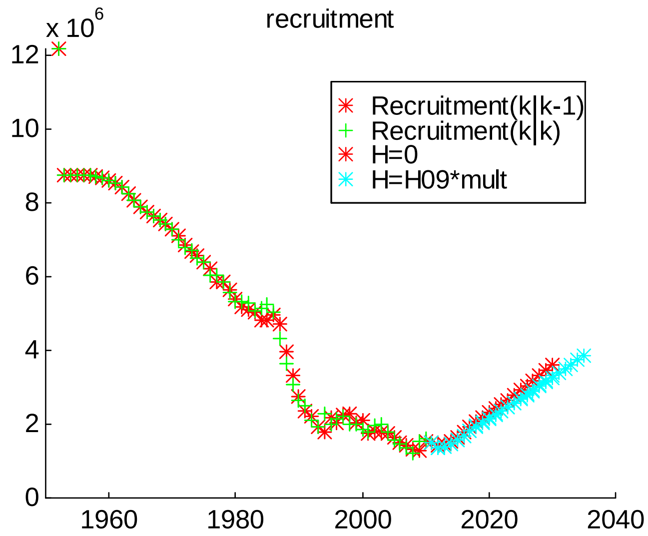

Figure 8 Estimates of harvest rate per protomoment including those used to make the predictions to 2035.

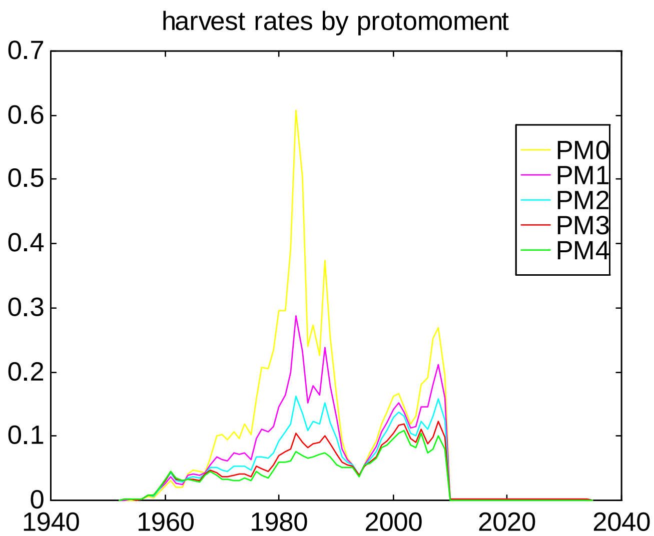

Figure 9 Harvest rates by year relative to the value for the harvest rate of biomass. (i.e. relative to the harvest rate of Protomoment 3)

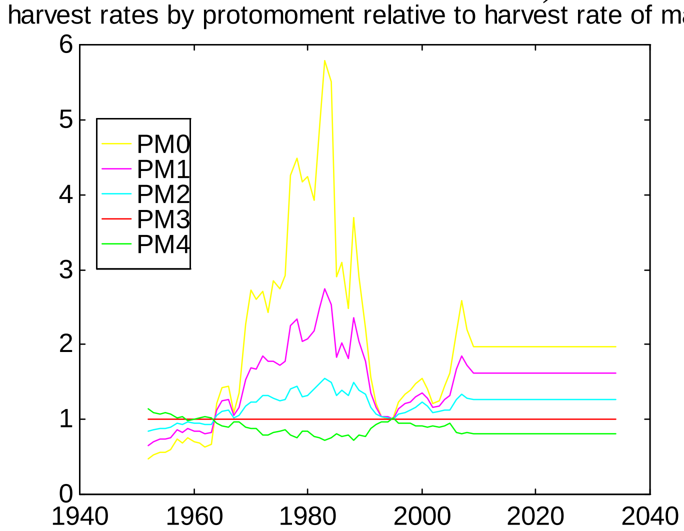

Figure 10 Total Catch including predicted catches to meet 2035 target (NB these historic catches used so far do not accord with the OMP record and will need amending)(units tonnes)

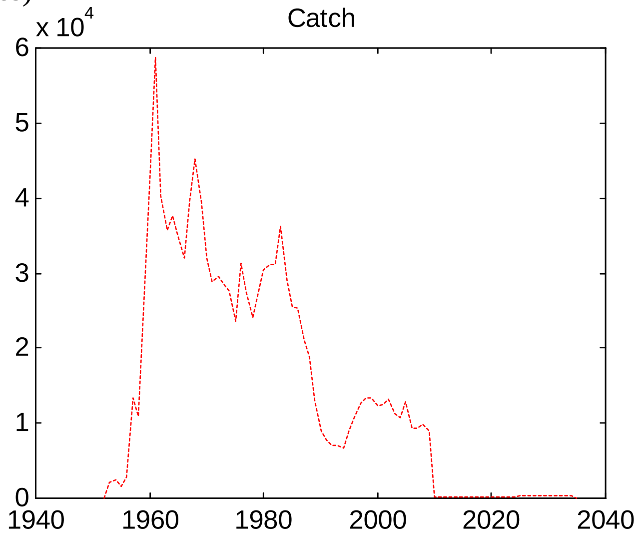

Figure 11 Yield surface (1000t) for different harvest rates of biomass (harvest rate) and different ratios of harvest rate of numbers to harvest rate of biomass(selection)- the latter estimate is also equivalent to the average weight in the sea/ average weight in the catch.

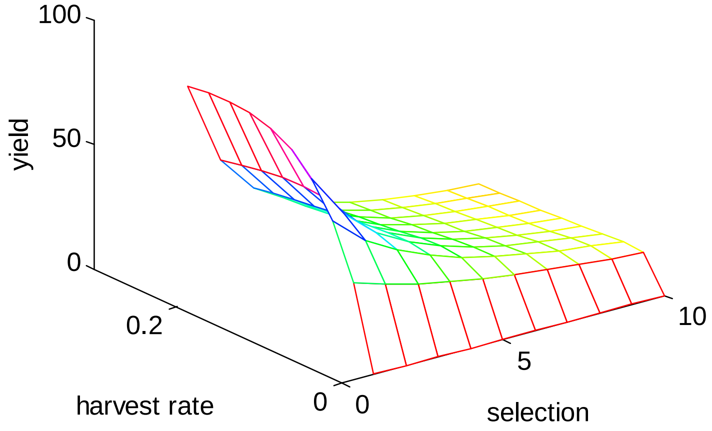

Figure 12 SSB surface (1000t) for different harvest rates of biomass (harvest rate) and different ratios of harvest rate of numbers to harvest rate of biomass(selection)- the latter estimate is also equivalent to the average weight in the sea/ average weight in the catch.

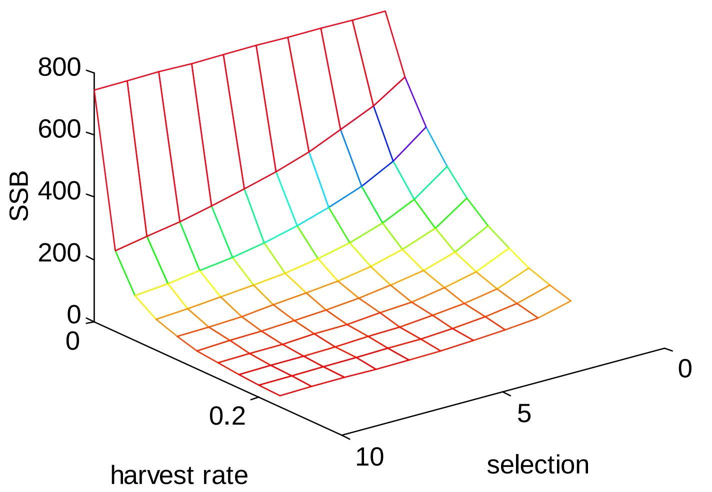

Figure 13 Time trajectory (telephone wires) of estimated catches (rather approximate) (tops of telephone poles) relative to the interpolated yield surface.

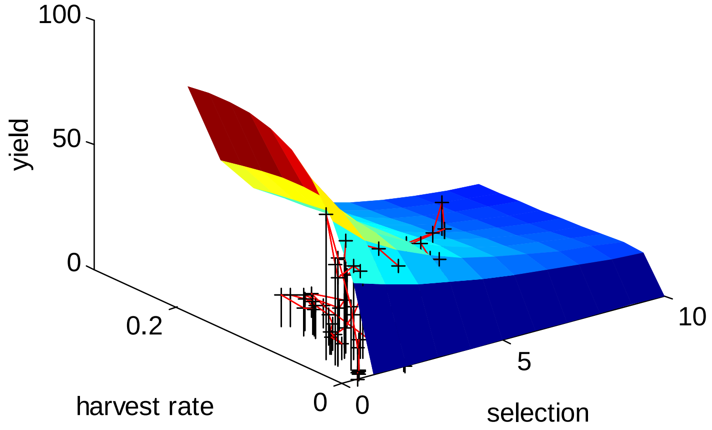

Figure 14 SSB from a tuned run made with an initial steepness of 0.58 and a final steepness of 0.55 (units tonnes)