NWWG 2002 **Working document # 22**

STatistical Age Model (STAM) of Icelandic cod using Excel

Einar Hjörleifsson

This documents describes the setup of a SIMPLE statistical catch at age model in MS Excel with data from the Icelandic cod used as an example. In principle this model does a forward projection of the stock in numbers, estimating catch at age and minimizes the the latter with the observed catch at age. Age-based survey index is also estimated by minimization with the observed age-based survey index. It is not considered to be a full fledged assessment package but more as an educational tool to the novice. The model uses catch-at-age data and age-based survey abundance indices and is based on the following equations:

$$s_a = \frac{1}{1+e^{-Ln(19)\frac{a-a_{50}}{a_{95}-a_{50})}}}$$

$$F_{ay} = s_a F_y$$

$$S_{ay} = e^{-(M+F_{ay})}$$

$$N_{a+1,y+1} = N_{ay} e^{-(M+F_{ay})} = N_{ay} S_{ay}$$

$$\hat{C}_{ay} = \frac{F_{ay}}{M+F_{ay}} N_{ay} \left(1-e^{-(M+F_{ay})}\right) = \frac{F_{ay}}{M+F_{ay}} N_{ay} (1-S_{ay})$$

$$\hat{I}_{ay} = e^{\ln \alpha_a + \beta_a \ln N_{ay}}$$

Parameter	Explanation
s_a	Selectivity of age a
a_{50}	age at which selectivity is 50%
a_{95}	age at which selectivity is 95%
F_y	Fishing mortality rate of the oldest ages in year y
F_{ay}	Fishing mortality of age a in year y
S_{ay}	Survivorship of age a in year y
N_{a+1}	Numbers at age a+1 in year y+1
N_{ay}	Numbers at age a in year y
M	Assumed natural mortality, set to 0.2
C_{ay}	Observed catch of age a in year y
C_{ay}-hat	Predicted catch of age a in year y
I_{ay}	Observed survey index of age a in year y
I_{ay}-hat	Predicted survey index of age a in year y
α_a	Survey scaling factor of age a
β_a	Survey scaling factor (slope) of age a

The fitting procedure was based on minimizing

$$SSR_C = \sum_{ay} \frac{\left[\ln\left(C_{ay}\right) - \ln\left(\hat{C}_{ay}\right)\right]^2}{2\sigma_a^2}$$

$$SSR_I = \sum_{ay} \frac{\left[\ln(I_{ay}) - \ln(\hat{I}_{ay})\right]^2}{2\rho_a^2}$$

$$SSR_T = a_C SSR_c + a_I SSR_I$$

The weighting factor (a_c and a_I) in each case was set to one. To reflect different accuracy by age in the estimation of catch-at-age and survey-index-at-age the residuals by age in the SSR_C and SSR_I were weighted according to:

Age ->	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Catch (sigma)			0,32	0,22	0,16	0,13	0,12	0,11	0,12	0,14	0,17	0,24	0,36	0,60	0,60
Survey (rho)	0,39	0,30	0,24	0,22	0,20	0,21	0,22	0,26	0,32	0,43					

In order to force the model to fit the observed yield within each year an additional minimization factor (penalty factor) was set in:

$$SSR_Y = \sum_y \frac{\left[\ln(Y_y) - \ln(\hat{Y}_y)\right]^2}{2\sigma_a^2}$$

$$SSR_T = a_C SSR_c + a_I SSR_I + a_Y SSR_y$$

where the a_Y factor was set to 1000.
The input data were into the model where:
1) catch-in-number matrix was based on ages 3 to 14 for years 1985 to 2001, Ca for ages 1 and 2 were assuemd 0 (thus 204 input values).
2) aged survey indices for ages 1 to 9 for years 1985-2002 (162 input values).
3) auxillary data on the CV by age groups for Cay and Uay - data obtained from hoski@hafro.is (see intext table above).
4) auxillary data such as corresponding weight at age in the catch and in the survey and maturity at age in the catch.

The number of estimated parameters were:
1) 13 estimates of N_{a85} for a = 2, 3, … , 14
2) 18 estimates of N_{1y} for y = 1985, 1986, …. , 2002
3) 9 estimates of α_a for a=1, 2, …. , 9
4) 7 estimates of β_a for a = 1, 2, …. , 7. I.e. power model was used for age groups 1-7, simple linear model for age groups 8 and 9 was used (β=1). The judgement on where to use power model was based on residual plots.
5) 17 estimates of Fy for y = 1985, 1986, …. , 2001
6) estmates of a_{50} and a_{95}
or a total of 66 parameter estimates. The input parameter

The accompanying Excel spreadsheet is named WD22_model.xls, with summary statistics and diagnostic plots provided in spreadsheet named WD22_plots.xls. When working through the formulation in the worksheet it is probably easiest to follow the order of the equations given above:
1. selectivity
2. fishing mortality
3. survivorship
4. numbers-at-age
5. exploitable number at age
6. prediced catch at age
7. catch at age residuals
8. catch at age residuals squared
9. prediced suvey index, residuals squared

The minimization in the current setup is performed via calculation in cells AU43:BI59 for catch in numbers, summed in cell B2. For the survey the minimization is performed via cells BN43:CB60, summed in cell B3. Solver is then uses to minimize cell B7 by changing values in B10:B79. It is important to have some reasonable initial values in cell ranges B10.B79. One way to start is by taking values from the good old VP-analysis. The model presented here is one of the simplest one of this type but could easily be expanded. The diagnostics are shown in a separte worksheet as well as some summary statistics of the model.

BOOTSTRAPPING [NOT COMPLETED AS OF YET IN THE ACCOMPANYING XLS-sheet, will before the end of the meeting]:
To characterize the uncertainty in the estimates and generate percentile confidence intervals the residuals from the optimum model fit were combined with the expected catch-at-age data to form a bootstap catch-at-age sample. Thus, the bootstrap samples were determined as:

$$C_{ay}^b = \hat{C}_{ay} \left(\frac{C_{ay}}{\hat{C}_{ay}} \right)^{boot}$$

The residuals applied to each age group in the current set up come from within each age group. For the survey the whole residual matrix was drawn from the residuals from a single year.
The bootstrap is set up as follows:
1. Find the optimised solution by using Solver (remember to do this on the initial/original data).
2. Copy the optimised Cay-hat data in cells AA3:AO18 and past them as values into AU43:BI98
3. In cell AW123 calculate Cay-obs/Cay-hat by setting in formula =ia3/ac3 [this is for the 3 year olds in 1985]. Copy this formula down and to the right, ending in cell BH138. Copy cell range AW123:BH138 and paste in the same locations as values.

4. Your Cay-boot values should now already be calculated for you in cell range AW163:BH178. Check your Excel manual on explanation about how the formula "works".
5. Copy the optimum estimated survey indexes in cell range AA43:AP59 and paste them as values in cell range BN83
6. In cell BN123 calculate Iy-obs/Iy-hat by setting in formula =g43/aa43. Copy this formula down to BN139.
7. Your Iy-boot values should now already be calculated for you in cell range BN163:BV179. Check your Excel manual
8. Save your spreadsheet before proceeding
9. Push the Do Boot button at the top left in your worksheet. READ item 10 here below FIRST. If you are seriously attempting to do a bootstrap I recommend to exclude the short-term prediction, it seems like the Excel “Table” feature slows things a lot down.
10. Take a long coffee break (actually this could be more like 2-48 hours depending on your computer speed and the number of bootstrapping you have set up in the macro - one can change your number of boostrappings in the macro). The bootstrap results are stored in a separate worksheet. 95%, 50% and the median are plotted up in an accompanying Excel document.

A background on this stuff can be found in Haddon 2001. Just note that there is an error in his Cay-hat calculation in the Excel example in chapter 11.

FORWARD PROJECTION:

SHORT TERM PREDICTIONS: The results of the short term predictions is found in cells G236:K272. It uses the estimates of the incoming recruits that are estimated in the model and whatever setting is set for the F value in the assessment year (cell AR139).

LONG TERM PROJECTIONS: No long term projections are provided in this version but could easily be included once a stock recruitment function is fitted and some distributional error structure on the deterministic recruitment estimator.