Description of an assesssment and prognosis tool (ADCAM)

Höskuldur Björnsson

6th May 2002

NWWG - 2002
Working document number 33

The following paper represents a set of routines written in AD-model builder. It is not a complete model where everything can be changed by changing an input file but rather a set of functions that can be added, changed and linked, with variable degree of effort. The "program" is designed for testing harvest control laws and the plan is to incorporate features like variable M to be able to test the robustness of different harvest control laws.

1 Description of model

The equations describing evolution of stocks in the model are the traditional cohort equations. (y stands for year and a for age).

$$F_{y,a} = F_y S_{y,a} \tag{1}$$

$$N_{y+1,a+1} = N_{y,a} e^{-Z_{y,a}} \tag{2}$$

$$Z_{y,a} = F_{y,a} + M_{t,a} \tag{3}$$

$$\hat{C_{y,a}} = \frac{F_{y,a}}{Z_{y,a}}(1 - e^{-Z_{y,a}}) \tag{4}$$

$$\tag{5}$$

Migrations are added (or subtracted) for specified agegroups in specified years and their magnitude estimated.

The model is a simulation model where an objective (negative log-likelihood) function that is a measure of the discrepancy between model and data is minimized by changing selected parameters.

The objective function is a sum of a number of different terms.

$$\ell = \sum_i \ell_i \tag{6}$$

1.1 Discrepancy between modeleled and observed catch in numbers.

$$\ell_1 = \sum_{a,y} (\frac{\log{(C_{y,a} + \epsilon_C)} - \log{(\hat{C_{y,a}} + \epsilon_C)})^2}{2\sigma_{1a}^2} - \log \sigma_{1a} \tag{7}$$

ϵ_C (read from file) is to reduce the effect of small numbers. The standarddeviations σ_{1a} are estimated by the model by assuming that they follow a parabolic relationship. The use a mixture of normal distribution and a distribution with heavier tail as described in the Autodiff manual might be suitable in many cases.

$$\sigma_{1a} = \kappa_1 + \kappa_2(a - \kappa_a)^2 \tag{8}$$

where some of the the parameters κ_1, κ_2 and κ_a are estimated.

Correlation in measurement errors of different agegroups is also implemented. Then the equation changes to

$$\mathbf{\Gamma} = \log{(\mathbf{C_{y,a}} + \epsilon_\mathbf{C})} - \log{(\hat{\mathbf{C}}_\mathbf{y,a} + \epsilon_\mathbf{C})}) \tag{9}$$

$$\ell_1 = \sum_y 0.5 \log \det \mathbf{\Theta_1} + \mathbf{\Gamma^T} \mathbf{\Theta_1^{-1}} \mathbf{\Gamma} \tag{10}$$

Where $\mathbf{\Gamma}$ are the residuals, and Θ_1 is a covariance matrix in the multivariate normal distribution describing the measurement error. Parameter(s) in the matrix Θ_1 can be estimated.

1.2 Discrepancy beetween model and survey indices.

$$\ell_2 = \sum_{a,y} (\frac{\log{(I_{y,a} + \epsilon_S)} - \log{(\hat{I_{y,a}} + \epsilon_S)})^2}{2\sigma_{2a}^2} - \log \sigma_{2a} \tag{11}$$

$$\sigma_{2a} = \delta_1 + \delta_2 (a - \delta_a)^2 \tag{12}$$

where some or all of the parameters δ_1, δ_2 and δ_a are estimated parameters. ϵ_S is to reduce the effect of small numbers.

As with the residuals between observed and predicted catch in numbers, coorrelation between different agegroups can be assumed. Then the equation changes to

$$\mathbf{\Gamma} = \log{(\mathbf{I_{y,a}} + \epsilon_\mathbf{S})} - \log{(\mathbf{\hat{I}_{y,a}} + \epsilon_\mathbf{S}))} \tag{13}$$

$$\ell_2 = \sum_y 0.5 \log \det \mathbf{\Theta_2} + \mathbf{\Gamma^T \Theta_2^{-1} \Gamma} \tag{14}$$

Here $\hat{I_{y,a}}$ is the modelled survey index and $I_{y,a}$ the observed survey index and $\mathbf{\Gamma}$ the vector of survey residuals. In some of the model runs the matrix $\mathbf{\Theta_3}$ has been set so

$$\Theta_{2ij} = \sigma_{2i} \sigma_{2j} \kappa^{abs(i-j)} \tag{15}$$

κ is an estimated parameter which has been estimated in the range 0.3 to 0.7 for cod, haddock and saithe in the Icelandic March survey.

Often CV of the measurement error is included with the estimate of a survey index, the estimated CV being high when few stations are responsible for the majority of the survey index. The CV can be utilized in the model by adding the measurement variance to σ_{2a}^2

$$\sigma_{2a} = \delta_+ \delta_2 (a - \delta_a)^2 \tag{16}$$

$$\sigma_{3a} = \sqrt{\sigma_{2a}^2 + CV_{y,a}^2} \tag{17}$$

$$\mathbf{\Theta_{2ij}} = \sigma_{3i} \sigma_{3j} \kappa^{absi-j} \tag{18}$$

$$\tag{19}$$

Instead of using the log transformation relationships of the form could be used.

$$\sum \frac{((I_{y,a} - \hat{I_{y,a}})^2}{2\sigma_4^2 \hat{I_{y,a}}^p} + 0.5 \log \sigma_4^2 I_{y,a}^{\hat{p}} \tag{20}$$

Where p is a power describing the relationship between standard deviation in the index and stock abundance.

$$\sum \frac{((I_{y,a} - \hat{I_{y,a}})^2}{2\sigma_4^2 (p_1 + \hat{I_{y,a}}^{p_2})} + 0.5 \log (\sigma_4^2 (p_1 + \hat{I_{y,a}}^{p_2}) \tag{21}$$

Could also be used.

These relationships could show better behaviour where $I_{y,a}$ is very low but $\hat{I_{y,a}}$ just low, as often happens when lack of samples cause small values to become zero. (multinomial)

$\hat{I_{y,a}}$ is calculated from either of the following equations.

$$\hat{I_{y,a}} = \alpha_a N_{y,a} \tag{22}$$

$$\hat{I_{y,a}} = \alpha_a N_{y,a}^{\beta_a} \tag{23}$$

The first equation has for Icelandic cod been used for age 6 and older but the second or third for ages 1 to 5. The parameters α_a and β_a are estimated for each age group but a parametric selection function could also be used.

1.3 Catch in tonnes.

As described above catch in numbers at age is one component in the objective function to be minimized. This does in many cases guarantee that the modelled catch in tonnes is close to the landed catch but in some years this is not the case. In all cases

$$C_y = \sum_a W_{ay} C_{ay} C_y \neq \hat{C}_y = \sum_a W_{ay} \hat{C_{ay}} \tag{24}$$

To let the model follow the "real" landed catch the following term is added to the objective function.

$$\ell_3 = \sum_y (\frac{\log C_y - \log \hat{C}_y)^2}{2\sigma_3^2} + \log \sigma_3 \tag{25}$$

Where σ_3 is input from a file and is typically rather low (0.03 to 0.05). An interesting exercise is to use high value of σ_3 and see where the modelled catch deviates from the observed catch.

1.4 Stock-recruitment model

As the model is to be used in prognosis some recruitment model must be included. As the model is set up now recruitment is predicted from long term mean, spawning stock (or some kind of spawning stock) using one of a number of relationships. The Beverton-Holt and Ricker relationship are expressed in terms of SSB_{max} which does of course not exist for the Beverton-Holt relatioship but means here 5 times the SSB giving half of maximum recruitment.

1.4.1 Beverton and Holt

$$\hat{N}_1 = R_{max} \frac{SSB}{0.2 SSB_{max} + SSB} \tag{26}$$

$$\tag{27}$$

1.4.2 Ricker

$$\hat{N}_1 = R_{max} e^1 \frac{SSB}{SSB_{max}} \exp \frac{-SSB}{SSB_{max}} \tag{28}$$

1.4.3 Shannon index

Relationship based on the Shannon age diversity index with mean 0.

$$N_{mat,tot} = \sum_{age} N_{mat,age} \tag{29}$$

$$Sha = \log N_{mat,tot} - \sum_{age} \frac{N_{mat,age} \log N_{mat,age}}{N_{mat,tot}} \tag{30}$$

$$Sha = Sha - S\bar{h}a \tag{31}$$

First relationship based on the Shannon index is a Ricker curve multiplied by the index times an estimated number.

$$\hat{N} = R_{max} e^1 \frac{SSB}{SSB_{max}} \exp \frac{-SSB}{SSB_{max}} * e^{\kappa * Sha} \tag{32}$$

Where κ is an estimated parameter.

Alternative relationship based on the Shannon age diversity index where the spawing stock is multiplied by the Shannon index and the Ricker stock-recruitment applied to that stock.

$$SSB_1 = SSBe^{Sha} \tag{33}$$

$$\hat{N} = R_{max} e^1 \frac{SSB_1}{SSB_{max}} \exp \frac{-SSB_1}{SSB_{max}} \tag{34}$$

1.4.4 Egg production

Relationship base where egg production as proportion of biomass is assumed to a function of ungutted weight. This function is then applied to the weight in the spawning stock. The motivation behind this function is that for Icelandic cod eggs are much larger proportion of the body weight of large females than is the case with small females.

$$E_{year} = \sum_{age} N_{mat,a,y} W_{mat_a,y} \gamma_1 W_{mat_a,y} \tag{35}$$

Where E_{year} is an estimate of the egg production. The relationship between mature biomass and egg production is based on analysis of data from the Icelandic groundfish in March survey where eggs,gonads, liver, ungutted weight and gutted weight have since 1994 been registered for all cod sampled for otholiths. Mature fish from the groundfish survey was split into groups according to ungutted weight and the total egg mass in each group calculated as fraction of the total mature biomass of the group. A linear relationship was fitted and $gamma_1$ is the result of that fit. The relationship is then applied to the mean weight in the spawning stock calculated with catch in numbers, which is the only weight available back in time but is far from being the right weight to use in this context.

1.4.5 Constant recruitment

The last SSB-recruitment relationship implemented is constant recruitment.

In all the relationship linear trend on R_{max} can be estimated stopping 5 years from the last data year.

2 Residuals from the SSB-recruitment relationship.

Instead of constant CV from the SSB-recruitment relationship the CV is assumed to follow a relationship of the form

$$\sigma_{4y} = \alpha \frac{SSB_y}{500}^{\beta} \tag{36}$$

For Icelandic cod the parameter β is estimated to be in the range 0.5 - 0.7 indicating increased variability in recruitment with decreased spawning stock. The case $\beta = 0$ means that the CV is independent of spawning stock size.

For Icelandic cod the ratio $\frac{SSB_y}{500}$ was limited to the range 0.4 to 2.0 to avoid extrapolation of this relationship into unknown terretory.

To take account of autocorrelation in residuals the residuals from the SSB-recruitment relationship are assumed to come from a multivariate normal distribution with variance - covariance matrix

S where element ij of **S** is given by

$$S_{ij} = r^{i-j} \sigma_{4y_i} \sigma_{4y_j} \tag{37}$$

σ_{4y} is given by the equation above and r is the autocorrelation of the residuals in the SSB-recruitment relationship. Of the Icelandic stock cod, haddock and saithe autocorrelation of residuals was estimated to be most important for the saithe stock.

3 Changes in effort and selection

3.1 Selection described by a logit function

As described before the fishing mortality is given as product of a selection pattern which can change with time and effort.

In one version version the selection pattern is described by a logit function

$$F_{y,a} = F_y S_{y,a} \tag{38}$$

$$S_{y,a} = \frac{1}{1 + e^{\alpha_y (a - a_{50y})}} \tag{39}$$

the notation a_{50y} indicates that a_{50} is allowed to change with time. To put some limitation on a_{50} it is assumed to develop as random walk.

To increase flexibility the model allows different slope below and above a_{50}. The transition is done by a steep logit function to make the function differentiable.

The random walk in a_{50} is implemented by adding the following term to the likelihood function.

$$\ell_5 = \sum \frac{(\log a_{50y+1} - log a_{50y})^2}{2\sigma_{5a}^2} + \log \sigma_{5a} \tag{40}$$

and if the slope evolves as random walk the following term is added

$$\ell_{5a} = \sum \frac{(\log \alpha_{50y+1} - \log \alpha_{50y})^2}{2\sigma_{5s}^2} + \log \sigma_{5s} \tag{41}$$

When the slope α is different below and above a_{50} two such equations must be formulated, one for each slope.

The model can not estimate σ_{5a} and σ_{5s}, it then finds the solution $\sigma_{5a} = 0$ and $\sigma_{5s} = 0$ having infinite value of the likelihood function . Therefore the proportion of these values and the measurement error is input from a file but another way would have been to start a run with little penalty on changes in in selection and use model output to get reasonably values for σ_{5a} and σ_{5s} ;

In addition transient changes around a fixed mean could be implemented.

$$\ell_5 = \sum \frac{(a_{50y} - \bar{a_{50}})^2}{2\sigma_5^2} + \log \sigma_5 \tag{42}$$

An important research on the model would be to understand how the objective function should be written so $a_5 0$ could be estimated. Ideas for that work could probably be obtained from a Kalman filter work where standard deviation of process and measurement error can both be estimated ??????

Fishing effort F_y is estimated for every year. The same considerations apply to fishing effort as with a_{50} in the selection pattern that the fishing effort can follow a random walk model.

$$\ell_6 = \sum \frac{(\log F_{y+1} - \log F_y)^2}{2\sigma_6^2} + \log \sigma_6 \tag{43}$$

Estimation of σ_6 is questionable and the same considerations apply there as to the estimation of σ_5. When "real" change in effort occurs as was the case when the catchrule was the restriction of F evolving as random walk with standard deviation σ_6 should be relaxed (GG). Large changes in effort are often accompanied by large changes in selection so equation 42 should also be relaxed.

3.2 Nonparametric fishing mortality

Describing the fishing pattern by a parametric function like the logit function can in many cases be too limiting and if natural mortality is to be estimated limitations of the selection pattern can easily generate the estimated M. The logit function is of course far from the best function to use, and orthogonal polynomials are possibly better better.

The most flexible way is though to estimate the fishing mortality for every year and age as done in an example in the AD-model builder manual. The estimation is done in steps with a seperable model first estimated and then deviations from the seperable model.

To put some restrictions on the fishing mortality txhe fishing mortality for each age group is assumed to follow a random walk model adding the following term to the likelihood function.

$$\ell_5 = \sum_y \sum_a \frac{(F_{y+1,a} - F_{y,a})^2}{2\sigma_{5a}^2} + \log \sigma_{5a} \tag{44}$$

The standard deviations σ_{5a} can not be estimated but are fixed as proportion of the measurement error (equation 8).

$$\sigma_{5a} = \sigma_{1a} P_a \tag{45}$$

Where P_a are proportions read from the file, describing the ratio between process and measurement error. For Icelandic cod and haddock measurement and process error were assumed to be similar for the middle aged fishes, measurement error more important for the older fishes and process error for the smaller.

Correlation in the process error could also be implemented and would be expected to be positive between adjacent agegroups while the measurement error is probably negative.

Equation is not used when real aprori known changes in effort occur as happened with the Icelandic cod between 1993 and 1994.

The nonparametric selection has been used in all of the assessments presented at this meeting.

3.3 Prognosis

The likelihood componts L_1 to L_6 apply to historical data i.e catch in numbers, survey, recruitment and catch in tonnes. None of those data has to exist for all the years and an interesting exercise is to skip catch data some years and see what the model does or skip survey data altogether and look at results based on catch in numbers only.

Prognosis runs in exactly the same way as the estimation i.e from given annual fishing effort, selection pattern and recruitment the development of the population can be calculated. But how are those items estimated as there are not data.

For the selection pattern the mean of some recent years is used

Recruitment is according to the spawningstock- recruitment estimated from historical data. No other data are available on future yearclasses so the model will be followed exactly but in the Monte Carlo simulations random deviations from the relationship will occur.

The given Harvest control rule will give some desired catch C_{yrule}. The likelihood component

$$\ell_8 = \sum \frac{(\log C_{yrule} - \log \hat{C}_y)^2}{2\sigma_8^2} + \log \sigma_8 \tag{46}$$

will be sufficient to estimate F_y for future years. σ_8 in the equation is typically low or 0.03.

Assessment error is implemented by "estimating" a vector $\epsilon_y..\epsilon y + ns$ where ns is the number of simulation years. The likelihood component used to generate ϵ is

$$\ell_7 = \sum \frac{(\epsilon_y)^2}{2\sigma_7^2} + \log \sigma_7 \tag{47}$$

where σ_7 is the assessment error. In the estimation all the ϵ are but they get the right distribution in the bayesian simulations and are added to the catch according to the catch rule so the equiation for ℓ_8 above becomes.

$$\ell_8 = \sum \frac{\log(C_{yrule} + \epsilon_y - \log \hat{C}_y)^2}{2\sigma_8^2} + \log \sigma_8 \tag{48}$$

Autocorrelations and dependence of assessment error on stocksize have also been added in the same way as done in the spawning stock-recruitment relationship. These values could also be estimated from historical data.

4 Variable natural mortality

The model was designed as a tool to look at the effects of different harvesting strageties. An important thing to look at in this context is the robustness of harvest control laws to (possible) variations in natural mortality. To do this M can be allowed to evolve as random walk by estimating it annually and allowing it to evolve as random walk by adding the following term to the likelihood function.

$$\ell_9 = \sum_y \frac{(M_{y+1} - M_y)^2}{2\sigma_9^2} + \log \sigma_9 \tag{49}$$

Getting "resonable" posteriors of natural mortality has turned out to be difficult as available data do not contain much information on M. For Icelandic cod estimating the migrations from Greenland (11 pcs) reduces the negative log-likelihood function 3 times more than changing the value of M from 0.3 to 0.1. If the selection pattern used is too resctrictive estimated values of M are just to compensate for lack of flexibility in the selection pattern. It must though be emphasised that the goal here is not to estimate M

but to get reasonably “wide” posteriors of M but the posteriors of M are very narrow when it is used to compensate for too restrictive selection pattern.