Appendix C2: Statistical Catch-at-Age Analysis Methodology

The model equations and the general specifications of the SCAA methodology applied are described below, followed by details of the contributions to the (penalised) log-likelihood function from the different sources of data available and assumptions concerning the stock-recruitment relationship. Quasi-Newton minimization is used to minimize the total negative log-likelihood function (the package AD Model Builder, Otter Research, Ltd is used for this purpose).

B1. Population dynamics

B1.1 Numbers-at-age

The resource dynamics are modelled by the following set of population dynamics equations:

\[ N_{y,a+1} = N_{y,a} e^{-m_{a}/2} - \sum f C_{y,f}^a \]  \[ e^{-m_{a}/2} \] for \( 1 \leq a \leq m - 2 \)  

\[ N_{y,m} = \left( N_{y,m-1} e^{-m_{m-1}/2} - \sum f C_{y,f}^m \right) e^{-m_{m-1}/2} + \left( N_{y,m} e^{-m_{m}/2} - \sum f C_{y,f}^m \right) e^{-m_{m}/2} \]

where

- \( N_{y,a} \) is the number of fish of age \( a \) at the start of year \( y \) (which refers to a calendar year),
- \( R_y \) is the recruitment (number of 1-year-old fish) at the start of year \( y \),
- \( M_a \) denotes the natural mortality rate for fish of age \( a \),
- \( C_{y,f}^a \) is the predicted number of fish of age \( a \) caught in year \( y \) by fleet \( f \), and
- \( m \) is the maximum age considered (taken to be a plus-group).

B1.2. Recruitment

The number of recruits (1-year olds) at the start of year \( y \) is assumed to be related to the spawning stock size (i.e. the biomass of mature fish) by a Beverton-Holt or a modified (generalised) form of the Ricker stock-recruitment relationship, parameterised in terms of the “steepness” of the stock-recruitment relationship, \( h \), and the pre-exploitation equilibrium spawning biomass, \( SSB_0 \), and recruitment, \( R_0 \) and allowing for annual fluctuation about the deterministic relationship:

\[ R_{y,1} = \frac{4hR_0 SSB_y}{SSB_0 (1-h) + (5h-1)SSB_y} e^{\left(\gamma y - \sigma_y^2/2\right)} \]

for the Beverton-Holt stock-recruitment relationship and

\[ R_{y,1} = \alpha SSB_y \exp\left(-\beta (SSB_y)^{\gamma} \right) e^{\left(\gamma y - \sigma_y^2/2\right)} \]

with

\[ \alpha = R_0 \exp\left(\beta (SSB_0)^{\gamma} \right) \quad \text{and} \quad \beta = \frac{\ln(5h)}{(SSB_0)^{\gamma} (1-5^{-\gamma})} \]

for the modified Ricker relationship (for the true Ricker, \( \gamma = 1 \)).

50th SAW Assessment Report  839  Pollock; Appendixes
reflects fluctuations about the expected recruitment for year \( y \), which are assumed to be normally distributed with standard deviation \( R \) (which is input in the applications considered here); these residuals are treated as estimable parameters in the model fitting process. \( SSB_y \) is the spawning biomass at the start of year \( y \), computed as:

\[
SSB_y = \sum_{a=1}^{m} f_{y,a} w_{y,a} N_{y,a}
\]

where

- \( w_{y,a} \) is the mass of fish of age \( a \) at the beginning of the year (Table A6), and
- \( f_{y,a} \) is the proportion of fish of age \( a \) that are mature (Table A5).

In the fitting procedure, \( SSB_0 \) is estimated while \( h \) can be estimated or fixed. For the Beverton-Holt form, \( h \) is bounded above by 0.9 to preclude high recruitment at extremely low spawning biomass, whereas for the modified Ricker form, \( h \) is bounded above by 1.5 to preclude extreme compensatory behaviour.

### B1.3. Total catch and catches-at-age

The fleet-disaggregated catch by mass in year \( y \) is given by:

\[
C_y = \sum_{a=1}^{m} w^{\text{mid}}_{y,a} C_{y,a} = \sum_{a=1}^{m} w^{\text{mid}}_{y,a} N_{y,a} e^{-M \cdot a / 2} S_{y,a} F_y
\]

where

- \( w^{\text{mid}}_{y,a} \) denotes the mass of fish of age \( a \) landed in year \( y \) (Tables A7, A8 and A9),
- \( C_{y,a} \) is the catch-at-age, i.e. the number of fish of age \( a \), caught in year \( y \) by fleet \( f \),
- \( S_{y,a} \) is the commercial selectivity of fleet \( f \) (i.e. combination of availability and vulnerability to fishing gear) at age \( a \) for year \( y \); when \( S_{y,a} = 1 \), the age-class \( a \) is said to be fully selected, and
- \( F_y \) is the proportion of a fully selected age class that is fished, for fleet \( f \).

### B1.4. Initial conditions

For the first year \( (y_0) \) considered in the model, the stock is assumed to be at a fraction \( (\theta) \) of its pre-exploitation biomass, i.e.:

\[
SSB_{y_0} = \theta \cdot SSB_0
\]

with the starting age structure:

\[
N_{y_0,a} = R_{\text{start}} N_{\text{start},a} \quad \text{for} \quad 1 \leq a \leq m
\]

where

\[
N_{\text{start},1} = 1
\]

\[
N_{\text{start},a} = N_{\text{start},a-1} e^{-M \cdot a-1} (1 - \phi S_{a-1}) \quad \text{for} \quad 2 \leq a \leq m-1
\]

\[
N_{\text{start},m} = N_{\text{start},m-1} e^{-M \cdot m-1} (1 - \phi S_{m-1}) / (1 - e^{-M \cdot (1 - \phi S_m)}) \quad \text{(B12)}
\]

where \( \phi \) characterises the average fishing proportion over the years immediately preceding \( y_0 \).

### B2. The (penalised) likelihood function

The model can be fit to survey indices and catch-at-age as well as commercial catch-at-age data to estimate model parameters (which may include residuals about the stock-recruitment function,
through the incorporation of a penalty function described below). Contributions by each of these to the negative of the (penalised) log-likelihood (-nl) are as follows.

**B2.1 Survey relative abundance data**

The likelihood is calculated assuming that an observed index for a particular survey is log-normally distributed about its expected value:

\[ l_y^i = l_y^i \exp(\epsilon_y^i) \quad \text{or} \quad \epsilon_y^i = n(l_y^i) - n(l_y^i) \]

(B13)

where

\[ l_y^i \] is the survey index for year \( y \) and series \( i \),

\[ l_y^i = q^i \hat{B}_y^{\text{surv}} \] is the corresponding model estimate, where

\[ \hat{B}_y^{\text{surv}} = \sum_{a=1}^{m} S_{a}^{\text{surv}} N_{y,a} e^{-\frac{M_a}{4} \left(1 - \sum_{f} F_{y,f}^l / 4\right)} \]  

(B14)

for spring surveys,

\[ \hat{B}_y^{\text{surv}} = \sum_{a=1}^{m} S_{a}^{\text{surv}} N_{y,a} e^{-\frac{2}{2} \left(1 - \sum_{f} F_{y,f}^l / 2\right)} \]  

(B15)

for summer surveys,

\[ \hat{B}_y^{\text{surv}} = \sum_{a=1}^{m} S_{a}^{\text{surv}} N_{y,a} e^{-\frac{4}{4} \left(1 - 3 \sum_{f} F_{y,f}^l / 4\right)} \]  

(B16)

for fall surveys,

\[ \hat{B}_y^{\text{surv}} = B_y^{sp} \]  

(B17)

for the larval index, and

\[ q^i \] is the constant of proportionality (catchability) for survey series \( i \), and

\[ \epsilon_y^i \] from \( N(0, \sigma_y^i)^2 \).

The contribution of the survey indices to the negative of the log-likelihood function (after removal of constants) is then given by:

\[-nl^{\text{surv}} = \sum_{i} \sum_{y} \left[ n(l_y^i) + (\epsilon_y^i)^2 / 2(\sigma_y^i)^2 \right] \]

(B18)

where

\[ \sigma_y^i \] is the standard deviation of the residuals for the logarithm of index \( i \) in year \( y \), taken to be given by the survey CV.

The estimated CVs likely fail to include all sources of variability, and unrealistically high precision could hence be accorded to these indices. The procedure adopted takes account of an additional variance \( (\sigma_y^i)^2 \) which is treated as another estimable parameter in the minimisation process, and included by replacing \( \sigma_y^i \) by \( \sqrt{(\sigma_y^i)^2 + (\sigma_y^i)^2} \) in equation B18. This procedure is carried out enforcing the constraint that \( 0 \leq (\sigma_y^i)^2 \leq 2 \).

The catchability coefficient \( q^i \) for survey index \( i \) is estimated by its maximum likelihood value:
The contribution of the catch-at-age data to the negative of the log-likelihood function under the assumption of an “adjusted” lognormal error distribution is given by:

\[ n \hat{q}^i = \frac{\sum_y (\ln f_{y,a} - \ln \hat{B}_{y,a}^\text{surv})/\left( (\sigma_{y,a}^i)^2 + (\sigma_A^i)^2 \right)}{\sum_y 1/(\sigma_{y,a}^i)^2 + (\sigma_A^i)^2} \]  

(B19)

**B2.3. Commercial catches-at-age**

The contribution of the catch-at-age data to the negative of the log-likelihood function under the assumption of an “adjusted” lognormal error distribution is given by:

\[ -n L_{\text{CAA}} = \sum_f w_{\text{CAA}} \sum_y \sum_a \left[ n \left( \sigma_{\text{com}}^f / \sqrt{p_{y,a}^f} \right) + p_{y,a}^f \left( n p_{y,a}^f - n \hat{p}_{y,a}^f \right)^2 / 2(\sigma_{\text{com}}^f)^2 \right] \]  

(B20)

where

- \( p_{y,a}^f = C_{y,a}^f / \sum_a C_{y,a}^f \) is the observed proportion of fish caught in year \( y \) by fleet \( f \) that are of age \( a \),
- \( \hat{p}_{y,a}^f = \hat{C}_{y,a}^f / \sum_a \hat{C}_{y,a}^f \) is the model-predicted proportion of fish caught in year \( y \) by fleet \( f \) that are of age \( a \),

where

\[ C_{y,a}^f = N_{y,a} e^{-M_{y,a}/2} \sum_f S_{y,a}^f \]  

(B21)

and

\[ \sigma_{\text{com}}^f \] is the standard deviation associated with the catch-at-age data of fleet \( f \), which is estimated in the fitting procedure by:

\[ \sigma_{\text{com}}^f = \sqrt{\sum_y \sum_a p_{y,a}^f \left( n p_{y,a}^f - n \hat{p}_{y,a}^f \right)^2 / \sum_y \sum_a 1} \]  

(B22)

\( w_{\text{CAA}} \) is input (this allows for the contribution from these data to be up-or downweighted compared to that from the survey indices).

The log-normal error distribution underlying equation (B20) is chosen on the grounds that (assuming no ageing error) variability is likely dominated by a combination of interannual variation in the distribution of fishing effort, and fluctuations (partly as a consequence of such variations) in selectivity-at-age, which suggests that the assumption of a constant coefficient of variation is appropriate. However, for ages poorly represented in the sample, sampling variability considerations must at some stage start to dominate the variance. To take this into account in a simple manner, motivated by binomial distribution properties, the observed proportions are used for weighting so that undue importance is not attached to data based upon a few samples only.

Commercial catches-at-age are incorporated in the likelihood function using equation (B20), for which the summation over age \( a \) is taken from age \( a_{\text{minus}} \) (considered as a minus group) to \( a_{\text{plus}} \) (a plus group).

**B2.4. Survey catches-at-age**

The survey catches-at-age are incorporated into the negative log-likelihood in an analogous manner to the commercial catches-at-age, assuming an adjusted log-normal error distribution (equation B20) where:

- \( p_{y,a}^{\text{surv}} = C_{y,a}^{\text{surv}} / \sum_a C_{y,a}^{\text{surv}} \) is the observed proportion of fish of age \( a \) from survey \( \text{surv} \) in year \( y \),
- \( \hat{p}_{y,a}^{\text{surv}} \) is the expected proportion of fish of age \( a \) in year \( y \) in the survey \( \text{surv} \), given by:
\[ \hat{\beta}_{y,a}^{\text{spring}} = \frac{S_{\text{surv}} y,a N y,a e^{\frac{M_{3a}}{4}} \left(1 - \sum_{f} S_{y,a,F_y}^{f} / 4\right)}{\sum_{a'} S_{\text{surv}} y,a e^{\frac{M_{3a}}{4}} \left(1 - \sum_{f} S_{y,a,F_y}^{f} / 4\right)} \] 

for spring surveys, and

\[ \hat{\beta}_{y,a}^{\text{fall}} = \frac{S_{\text{surv}} y,a e^{\frac{3M_{3a}}{4}} \left(1 - 3 \sum_{f} S_{y,a,F_y}^{f} / 4\right)}{\sum_{a'} S_{\text{surv}} y,a e^{\frac{3M_{3a}}{4}} \left(1 - 3 \sum_{f} S_{y,a,F_y}^{f} / 4\right)} \] 

for fall surveys.

**B2.5. Survey catches-at-length**

The predicted proportions-at-age from equations B23 and B24, or similar equations for other surveys, may be converted into proportions-at-length using the von Bertalanffy growth equation, assuming that the length-at-age distribution remains constant over time:

\[ \hat{\beta}_{y,l}^{\text{surv}} = \sum_{a} \hat{\beta}_{y,a}^{\text{surv}} A_{a,l}^{\text{surv}} \]  

where

\[ A_{a,l}^{\text{surv}} \] is the proportion of fish of age \( a \) that fall in the length group \( l \) for survey \( \text{surv} \) (i.e. \( \sum_{l} A_{a,l}^{\text{surv}} = 1 \) for all ages \( a \) for survey \( \text{surv} \)).

The matrix \( A \) is calculated under the assumption that length-at-age is normally distributed about a mean given by the von Bertalanffy equation, i.e.:

\[ L_{a} \sim N \left[ L_{a} \left(1 - e^{-\kappa(a-1)}\right), \sigma_{a}^{2}\right] \]  

where \( N \) is the normal distribution, and \( \sigma_{a} \) is the standard deviation of length-at-age \( a \), which is modelled to be proportional to the expected length at age \( a \), i.e.:

\[ \sigma_{a} = \beta_{a} \left(1 - e^{-\kappa(a-1)}\right) \]  

where \( \beta_{a} \) can be fixed or estimated in the model fitting process.

The following term is then added to the negative log-likelihood:

\[ -n L_{\text{CAL}}^{\text{surv}} = \sum_{\text{surv}} w_{\text{CAL}} \sum_{y} \sum_{l} \left[ n \left( \sigma_{\text{len}}^{\text{surv}} / \hat{\beta}_{y,l}^{\text{surv}} \right) + \hat{\beta}_{y,l}^{\text{surv}} \left( n p_{y,l}^{\text{surv}} - n \hat{\beta}_{y,l}^{\text{surv}} \right)^{2} / 2\left(\sigma_{\text{len}}^{\text{surv}}\right)^{2} \right] \]  

where \( \hat{\beta}_{y,l}^{\text{surv}} \) is the observed proportion (by number) in length group \( l \) in the catch in year \( y \) for survey \( \text{surv} \), and \( \sigma_{\text{len}}^{\text{surv}} \) is the standard deviation associated with the length-at-age data for survey \( \text{surv} \), which is estimated in the fitting procedure by:

\[ \sigma_{\text{len}}^{\text{surv}} = \sqrt{\sum_{y} \sum_{l} \left( \hat{\beta}_{y,l}^{\text{surv}} \left( \ln p_{y,l}^{\text{surv}} - \ln \hat{\beta}_{y,l}^{\text{surv}} \right)^{2} / \sum_{y} \sum_{l} 1 \right)} \]  

for spring surveys, and

\[ \hat{\beta}_{y,l}^{\text{fall}} = \frac{S_{\text{surv}} y,a e^{\frac{3M_{3a}}{4}} \left(1 - 3 \sum_{f} S_{y,a,F_y}^{f} / 4\right)}{\sum_{a'} S_{\text{surv}} y,a e^{\frac{3M_{3a}}{4}} \left(1 - 3 \sum_{f} S_{y,a,F_y}^{f} / 4\right)} \]  

for fall surveys.
The \( w_{C_{\text{kal}}} \) weighting factor may be set at a value less than 1 to downweight the contribution of the catch-at-length data to the overall negative log-likelihood compared to that of the survey or catch-at-age data. The reason that this factor is introduced is that the \( p_{y,j}^{\text{sur}} \) data for a given year frequently show evidence of strong positive correlation, and so are not as informative as the independence assumption underlying the form of equation B28 would otherwise suggest.

**B2.6. Stock-recruitment function residuals**

The stock-recruitment residuals are assumed to be log-normally distributed. Thus, the contribution of the recruitment residuals to the negative of the (now penalised) log-likelihood function is given by:

\[
- nL_{SR}^{\text{pen}} = \sum_{y=y_1}^{y_2} \left[ \frac{\varepsilon_y^2}{2\sigma^2_R} \right]
\]

(B30)

where

- \( \varepsilon_y \) from \( N(0, \sigma_R^2) \), which is estimated for year \( y_1 \) to \( y_2 \) (see equation (B4)), and
- \( \sigma_R \) is the standard deviation of the log-residuals, which is input (a value of 0.4 is used for the Base Case assessment).

**B3. Model parameters**

**B3.1. Commercial fishing selectivity-at-age**

The commercial fleet-specific fishing selectivity, \( S_{a,f} \), is estimated directly for each age from age ‘minus’ to age ‘plus’. The estimated decreases from ages minus+1 to minus and ages plus-1 to plus are either assumed to continue exponentially to ages 0 and \( m \) (maximum age considered) respectively.

Time dependence may be incorporated into these specifications by estimating different selectivity parameters for specific time periods, so that \( S_{a,f} \rightarrow S_{y,a,f} \).

**B3.2. Survey fishing selectivity-at-age**

For the NEFSC spring and fall surveys, the fishing selectivity, \( S_{a}^{\text{surv}} \), is estimated directly for each age from age 1 to age 8. The selectivity is assumed to remain constant at the level estimated for age 8 for ages 9 and above.

For the NEFSC summer survey, the selectivity is assumed to take the form of an exponential decline up to some maximum age specified, after which it becomes zero:

\[
S_{a}^{\text{surv}} = e^{-\lambda(a-1)}
\]

(B31)

The Maine/New Hampshire spring and fall surveys, as well as the Massachusetts inshore surveys are taken as indices of recruitment for the Base Case as their catch-at-length distributions are dominated by lengths corresponding to 1-year-old fish, i.e.:

\[
S_{a}^{\text{surv}} = \begin{cases} 
1 & \text{for } a = 1 \\
0 & \text{for } a \neq 1 
\end{cases}
\]

(B32)

**B3.3. Natural mortality-at-age**

\( M_a = 0.2 \)

(B33)