

8 ADAPT

The data model of ADAPT supplements the catch in numbers-at-age and year with at least one abundance or biomass index series. This allows estimation of the entire set of equations. With the exception of natural mortality, this leaves no external parameters to be fixed, as VPA requires. Care should be taken that there is sufficient information available, so that all parameters are estimable. This is a problem when the available indices are of the biomass type. In most assessment problems, the availability of a biomass index (e.g. catch weight per unit effort) will not be sufficient to allow full estimation of the state of the stock. In this section, the “tuning series” are presented as a CPUE series obtained by research vessels (e.g. a bottom trawl survey). It is assumed that this survey will provide a valid index for all age groups and that all data have the same error distribution. It is however not assumed that the sampling efficiency (the catchability) is the same for all age groups. This assumption of fixed error is relaxed in Section 8.3 where relative weighting of the data series is discussed.

This procedure was presented by Gavaris (1988). Stefansson (1997) gives a practical guide to the method.

8.1 ADAPT WITH EXTERNAL WEIGHTING

The ADAPT framework assumes that all deviations between the model and the observations are due to measurement error. It does not make any explicit assumptions of the error structure, but assumes that the least-squares estimator is applicable. The distribution of measurement errors in biological investigations is often not symmetrical, and therefore it is customary to work on logarithmic values. This has been found to be more appropriate in most real stock assessments.

The estimation is to find the minimum of the least-squares sum. This sum now has at least two terms, the catch and CPUE:

$$\sum_{a \in \text{age}} \sum_{y \in \text{year}} \left[\ln \text{Catch}_{ay}^{\text{obs}} - \ln \text{Catch}_{ay}^{\text{mod}} \right]^2 + \sum_{a \in \text{age}} \sum_{y \in \text{year}} \left[\ln \text{CPUE}_{ay}^{\text{obs}} - \ln \text{CPUE}_{ay}^{\text{mod}} \right]^2 = \text{MIN}\{\text{model parameters}\} \quad (86)$$

The sum range over age and year may not be the same for the catch and the CPUE data. For the catch, the model in numbers by age and by year is:

$$\ln \text{Catch}_{ay}^{\text{mod}} = \ln P_{ay} + \ln F_{ay} + \ln \left(\frac{1 - \exp(-Z_{ay})}{Z_{ay}} \right) \quad (87)$$

with

$$Z_{ay} = F_{ay} + M_a$$

and for the CPUE data the model is:

$$\ln C_{pue}_{ay}^{mod} = \ln q_a + \ln \tilde{P}_{ay}$$

with

$$\ln \tilde{P}_{ay} = \ln P_{ay} - \alpha Z_{ay} + \ln \left(\frac{1 - \exp(-(\beta - \alpha)Z_{ay})}{(\beta - \alpha)Z_{ay}} \right) \quad (88)$$

and

$$Z_{ay} = F_{ay} + M_a$$

The parameters are the matrices: population in number P_{ay} or the fishing mortality F_{ay} . The natural mortality M_a and the parameters α and β relating the timing of the survey to the stock calculation are provided as external parameters. α and β are the start and end point in time of the observation given as a fraction of the year.

The problem can be reduced by assuming that the noise in the catch data is much less than the errors in the CPUE data and therefore the catch term in the sum of squares can be fulfilled exactly. This means that all fishing mortalities and all population sizes can be written as a function of the fishing mortality in the terminal year for all age groups and of the terminal age for all years. This reduces the parameters to the log catchabilities and the terminal log fishing mortality (or equivalently the terminal log population of survivors). A further reduction is to note that the least-squares estimator for the log catchabilities can be found explicitly as:

$$\ln q_a = \frac{\sum_y (\ln CPUE_{ay}^{obs} - \ln \tilde{P}_{ay})}{n} \quad (89)$$

where n is the number of years in the sum. After these reductions, the problem is to find the minimum of the sum-of-squares:

$$\sum_{a \in age} \sum_{y \in year} \left[\ln CPUE_{ay}^{obs} - \ln \left(\frac{\sum_y (\ln CPUE_{ay}^{obs} - \ln \tilde{P}_{ay})}{n} \right) - \ln \tilde{P}_{ay} \right]^2 = MIN \{ \text{Terminal } F's \} \quad (90)$$

The algorithm to find these unknown parameters consists of the following steps:

- (i) Initiate the unknown parameters, $\ln(F)$ or $\ln(P)$, with guessed estimates.
- (ii) Perform a VPA to estimate population and fishing mortality coefficients for all age groups and years.
- (iii) Calculate the $\ln(q_a)$ parameters.
- (iv) Calculate the sum-of-squares using the VPA solution and the calculated log catchabilities.

Steps (ii)-(iv) need to be imbedded in an iterative routine to find the minimum of the sum-of-squares.

The results are:

- Estimates of the abundance P_{ay} .
- Estimates of fishing mortality coefficients F_{ay} .
- Diagnostic information. This includes residuals calculated for the CPUE index and the standard errors of parameters, terminal fishing mortalities or terminal survivors, the coefficients of variation of these parameters, the correlation matrix of the parameters (Chapter 7).

The ADAPT procedure works best when catches are the dominating cause of mortality (i.e. $F > M$). This is the case for many important fish stocks (e.g. Halliday and Pinhorn 1996, Serchuk *et al.* 1996). With the introduction of the precautionary approach, hopefully there will be increasing numbers of examples where $F \gg M$ no longer applies. The estimation procedures are then more prone to noise in the abundance index data than if $F > M$. In practice, this is illustrated by the instability in the estimated stock sizes that occur in the assessments between two neighbouring years.

The ADAPT estimates are functions of the natural mortality coefficient. From VPA, it is known that the estimated $F + M$ is fairly constant locally around the chosen value of M . The VPA behaviour dominates the ADAPT calculations except for the most recent years where the results are very dependent on the input values for the oldest age and for the terminal year in the assessment. Even so, for the most recent year the fishing mortality estimate will still be lower with larger natural mortality values.

8.2 SEVERAL ABUNDANCE INDICES

Normally several abundance indices are available and the object function is expanded to include these:

$$\begin{aligned}
 Lsq = & \sum \left(\ln Catch^{obs} - \ln Catch^{teo}(F, N) \right)^2 \\
 & + \sum \left(\frac{\ln Cpue_1^{obs} - \ln Cpue_1^{teo}(F, N, q_1)}{\sigma_1} \right)^2 \\
 & + \sum \left(\frac{\ln Cpue_2^{obs} - \ln Cpue_2^{teo}(F, N, q_2)}{\sigma_2} \right)^2
 \end{aligned} \tag{91}$$

The weights that are required are the variances (σ^2) of the CPUE indices. These should be readily available from the surveys. In the formulation above these weights are expressed relative to the accuracy of the catch data, a problem that is often less tractable than the estimation of the variance from the surveys. This is because of the way the catch-at-age matrix is built from many different sources and in some cases

there is no direct age-length key available for a portion of the total catch. In these cases one of the available ALKs is used which is thought to best represent that catch, but it is difficult to assess the variance and bias that such a procedure may produce in the catch-at-age matrix. In practice, the problem is unresolved and the CPUE indices (or other stock indicators) are given equal weight, while the catches are assumed to be estimated with far less variance than the surveys and hence this contribution is given infinite weight, i.e. the catch equation is fitted exactly.

8.3 EXTENDED SURVIVOR ANALYSIS (XSA)

This is the standard procedure used in ICES (Darby and Flatman 1994, Shepherd 1999). The method is different from the approach presented above in that the procedure does not define an object function. Instead, XSA is based on an iteration procedure of the functional type. The method is of the same type as the ADAPT presented above, the data are catch-at-age in numbers by age and by year supplemented by stock abundance indices. However, the approach is restricted as only age dis-aggregated abundance indices can be used.

The basis of the method is the link between the population and the abundance index through the catchability q :

$$Cpue = qN^{\gamma} \quad (92)$$

where q , γ vary with abundance index and with age, but is constant with respect to time. The CPUE values are all corrected to refer to the stock at the beginning of the year using the usual formula:

$$Cpue_{ayf} = Cpue_{ayf}^{obs} / \left[\exp(-\alpha(F_{a,y} + M_a)) \frac{1 - \exp(-(\beta - \alpha)(F_{ay} + M_a))}{F_{ay} + M_a} \right] \quad (93)$$

where α and β are the start and end point in time of the observation given as a fraction of the year.

The XSA iteration starts with an initial guess of the number of survivors (population at the end of the year of the oldest age group included in the analysis) and M . The XSA then applies a standard VPA to the catch-at-age and year data to provide stock sizes N . Based on these stock sizes the catchability q and the exponent γ can be estimated by linear regression (Equation 67):

$$\ln N_{ay}^{VPA} = \frac{1}{\gamma_{af}} \ln Cpue_{ayf} - \frac{\ln q_{af}}{\gamma_{af}} \quad (94)$$

where the subscripts a = age, y = year and f = fleet. The regression is hence over years for fixed age and fleet.

When the catchability and the exponent in the CPUE–stock relation has been determined, the next step is to correct the stock estimate by:

$$\ln N_{ayf}^{corr} = \frac{\ln Cpue_{ayf} - \ln q_{af}}{\gamma_{ay}} \quad (95)$$

so that each abundance index estimates the stock in numbers by age and by year. These estimates are then averaged to provide a new starting point for a new VPA. This average is based on calculating the number of survivors of the oldest age group included in the catch-at-age analysis

$$\ln N^{survivors} = \ln \left[\frac{Cpue_a}{q_a} \right] - F_{a,cum} - M_{a,cum} \quad (96)$$

where the fishing mortality F and natural M are cumulative over age (a) until the oldest age included in the analysis. For a given cohort, there will be a number of such estimates of survivors. These come from different age groups observed in the same abundance index and from different indices (e.g. commercial CPUE and research vessel surveys). The XSA combines these weighted estimates into a single estimate of the survivors of that cohort. This estimate is then introduced into a VPA of the catch in numbers by age and by year thereby obtaining stock in numbers and fishing mortality. This concludes the iteration loop. The next iteration loop begins by using these estimates to calculate the catchabilities (q_a) by age and by index type. The whole process is repeated until convergence. However, convergence is not guaranteed and there are examples where the iteration diverges.

The weights used for the survivor estimates are the inverse prediction variance around the regression carried out to estimate the catchabilities, multiplied by $F_{a,cum}$.

8.4 DOWN WEIGHTING OF OLDER DATA IN THE ANALYSIS

The segments of the time series may be of a different relevance. Fisheries develop and therefore catchabilities estimated from data that stretch far back in time may be of little use for an assessment that focuses on projecting the future. It therefore appears reasonable to introduce a down-weighting of older data. This can be done by simply restricting the analysis to the most recent 10-15 years of data or a more gradual down weighting with a time taper. In a common implementation of XSA, explicit weighting of the residuals with the year is introduced as a time taper (Darby and Flatman 1994).

$$W_{year} = \left[1 - \left(\frac{\text{Taper range} - \text{year}}{\text{Taper range}} \right)^p \right]^p \quad (97)$$

for year = 0,...,Taper range.

The taper range in the Lowestoft package is set at a default value of 20 years, so only the last 20 years catch or CPUE are included in fitting the model. The default taper is a tricubic type ($p=3$). The introduction of such a time taper modifies the object function to:

$$\begin{aligned}
& \sum_{a \in \text{age}} \sum_{y \in \text{year}} W_{\text{year}} \left[\ln \text{Catch}_{ay}^{\text{obs}} - \ln \text{Catch}_{ay}^{\text{mod}} \right]^2 \\
& + \sum_{a \in \text{age}} \sum_{y \in \text{year}} W_{\text{year}} \left[\ln \text{CPUE}_{ay}^{\text{obs}} - \ln \text{CPUE}_{a,y}^{\text{mod}} \right]^2 = \text{MIN}\{\text{model parameters}\}
\end{aligned} \tag{98}$$

8.5 REGULARISATION

The basic idea of regularisation is to assume that the exploitation pattern and the fishing mortality do not change abruptly from one year to the next. The estimation equation is therefore expanded, but includes a penalty for changes in fishing mortality. These methods include the shrinkage implemented in the XSA version of the Lowestoft/ICES software.

In many fisheries assessment problems, it is reasonable to assume that certain variables vary slowly (e.g. the fishing mortality). The basis for such assumptions is that an effort increase may require additional fleet capacity to be built that takes 1-2 years. The technical formulation of this assumption is to add an extra term to the least-squares equation. This means that the final least-squares expression to be minimised becomes:

$$\begin{aligned}
& \sum_{ay} \left(\ln C_{ay}^{\text{obs}} - \ln C_{ay}^{\text{teo}} \right)^2 + \sigma \sum_{ay} \left(\ln \text{Cpue}_{ay}^{\text{obs}} - \ln \text{Cpue}_{ay}^{\text{teo}} \right)^2 + \dots \\
& + \lambda \sum_{ay} \left(\ln F_{a,y+1} - \ln F_{ay} \right)^2 = \text{MIN}
\end{aligned} \tag{99}$$

The regularisation parameter (λ) controls how much variation between years is expected in the fishing mortality. The parameter should be supplied by the user and is given in units of inverse variance relative to the weight applied to the catch-at-age data. The example of the regularisation above is given in logarithmic terms, but this term can also be arithmetical.

Making estimates conditional on past values may not only represent the behaviour of the fishery more closely, but also improve the statistical behaviour of the fit. This can be seen more clearly by considering how the estimates change as λ in Equation 99 increases. As λ gets larger, there is an increasing "cost" to differences between sequential F 's, so as λ approaches infinity Equation 99 can only be minimised by making $F_{ay} = F_{a,y+1}$, so all F_{ay} 's are equal for all y . In essence, we would only be estimating one parameter, a best-fit F_a . Conversely, as λ approaches zero, we return to estimating an independent fishing mortality for each age in each year. As λ takes on values between these two extremes, the effective number of parameters being estimated varies continuously. In general, the fewer parameters there are (higher weight λ), the more reliable the estimates, but the poorer the model fits the data. This, and similar time series techniques, can allow a stock assessment to balance improving the model fit against estimate reliability.

Shrinkage as introduced in the XSA package belongs to this class of regularisation. In that implementation, the fishing mortalities after the estimation are regressed on a moving average:

$$F_{ay} = (1 - \theta)\tilde{F}_{ay} + \theta \frac{\sum_{j=1}^J F_{a,y-j}}{J} \quad (100)$$

where J is the number of years used to calculate the moving average, \tilde{F}_{ay} is the proper estimate from the minimisation algorithm and θ is a weight parameter ($0 < \theta < 1$) controlling how much weight should be given to the “shrinkage”. The population estimate of the last year can be treated similarly.