ESTIMATION OF NATURAL MORTALITY USING CATCH-AT-AGE AND AGED MARK-RECAPTURE DATA: A MULTI-COHORT SIMULATION STUDY COMPARING ESTIMATION FOR A MODEL BASED ON THE BARANOV EQUATIONS VERSUS A NEW MORTALITY EQUATION

S.G. Candy
Australian Antarctic Division
Department of Sustainability, Environment,
Water, Population and Communities
203 Channel Highway
Kingston, Tasmania 7050
Australia
Email – steve.candy@aad.gov.au

Abstract

An estimation strategy for natural mortality, M, using multiple cohorts and multiple years of catch-at-age and aged mark-recapture data was tested using simulation. Alternative fishing selectivity functions of age of dome-shaped versus sigmoidally shaped were applied. Two alternative estimation models were developed both using a Poisson likelihood for annual number of recaptures-at-age and model the population numbers-at-age by annual difference equations obtained by integrating an ordinary differential equation (ODE) for within-year population dynamics. The 'fully parametric' BODE model is based on the Baranov ODE while the 'semi-parametric' constant catch ODE (CCODE) model uses a new total mortality ODE with constant within-year catch per unit time and does not estimate annual fishing mortality rates (i.e. the F's) or fishing selectivity function parameters. It removes the actual, considered known, catch-at-age numbers directly from the population. Estimation for the BODE model requires an extra component to the log-likelihood which defines the process error in predicted catch-at-age numbers.

Simulations assumed 1 000 releases per year over 12 years with recruitment average of 2 million with annual coefficient of variation (CV) of 0.3 and annual catch of 500 000. Simulations which passed catch-at-age numbers to the estimation algorithm after perturbation by observational error were also carried out for each model in order to investigate the effect on estimation of M. Simulations carried out without observational error showed that when all parameters were jointly estimated and selectivity was dome-shaped, estimation of M was unreliable for both models but more so for the BODE model. The reason for this is explained by the confounding of selectivity parameter estimates with that for M. In contrast, when sigmoidally shaped selectivity was simulated, and the functional form of selectivity was correctly specified in the BODE model, both models gave close to unbiased and reasonably precise (CVs of 0.07 to 0.14) estimates of M, but the BODE model estimate was substantially more precise. However, when a minor misspecification of the functional form of selectivity was fitted by the BODE model, in comparison the CCODE model gave superior accuracy. When realistic observational error in catch-at-age numbers was included in simulations and combined with the sigmoidally shaped selectivity function, the bias and imprecision in estimates of M increased by no more than 2% for the CCODE model with no increase detectable for the BODE model. With these caveats, both models can be used to estimate this notoriously difficult parameter with the profile likelihood a useful indicator of the degree of success of estimation, even if some bias remains.

Introduction

The annual natural mortality rate, M, is a very influential parameter for estimating the productivity of a stock when considered in combination with annual number of recruits, growth rate in body size and the age-at-maturity ogive. Hence, obtaining as accurate an estimate of M as possible, along with estimates of its uncertainty bounds, is a key objective for assessments of stock status. For age-structured integrated assessments for the Patagonian toothfish (*Dissostichus eleginoides*) (Hillary et al., 2006; Candy and Constable, 2008) a pre-specified value of M has been applied when estimating other model parameters and then calculating allowable catch under the set of decision rules applied in these CCAMLR fisheries (Constable and de la Mare, 1996).

Although natural mortality is likely to vary between age class and sex and vary both temporally and spatially, it is extremely difficult to successfully model this level of complexity. Typically, M is assumed a single constant for the stock. Since, unlike fishing mortality, natural mortality is unobserved and is a result of a complex array of factors, an estimate of M can only be inferred from estimated changes in population numbers-at-age after accounting for age- and year-specific fishing mortality. Quinn and Deriso (1999, page 360) note that although age-based (such as the above integrated assessments), length-based and delay-difference population models have naturally mortality explicitly embedded with them, the estimation of M (even as a single constant) using these models is usually not successful because the objective function typically does not vary much with M. A number of methods of estimating M for commercially fished stocks with associated single datasets or combinations of datasets have been proposed, such as the method of Leigh et al. (2006) that uses multi-year tagging studies providing exact times of release and recapture or the method of Wang et al. (2009) that uses annual catch-at-age numbers combined with catch-per-unit-effort data and the 'extended survivors' method of Shepherd (1999). Leigh et al. (2006) use data for a single cohort and do not employ catch-at-age numbers, while Wang et al. (2009) do not employ aged mark-recapture data for parameter estimation. The two alternative population models and their corresponding estimation methods proposed here both rely on annual numbers in the catch and annual numbers of releases and recaptures for multiple cohorts and multiple years of fishing with each dataset disaggregated to age class. The motivation for this study was the need for a reliable estimate of M for the Heard and McDonald Islands (Division 58.5.2) Patagonian toothfish fishery, and a concerted effort to age historical collections of otoliths in recent years had allowed such datasets to be constructed as described in Candy et al. (2011). This study describes these proposed models and their estimation, and investigates their utility for estimating M, via simulation, in order to better inform their application to the above fishery as described in Candy et al. (2011).

Both population models require input of cohort-specific recruitment by combining the mean (i.e. the expected value across years) recruitment of age-1 fish, defining the parameter K_0, with year-class strength (YCS) parameters. Both then model the age-structured population for each fishing year by 'tracking' all cohorts in the population over time. The initial age structure at the first year of fishing and for subsequent years is obtained by combining values of fishing and natural mortality with values of cohort-specific recruitment. This requires the annual recursive use of difference equations for population numbers-at-age obtained by integrating a first-order ordinary differential equation (ODE) for within-year population dynamics. The difference in the population models derives from differences in their ODE. The model based on the Baranov (Baranov, 1918; Quinn and Deriso, 1999) ODE (BODE model) is based on fishing mortality expressed as a rate per unit of the population and therefore requires estimates of annual fishing mortality rates (F's) combined with estimates of the parameters of a fishing selectivity function. The form of this function is assumed to be known and the same function and parameter values are assumed to apply across fishing years. The population model based on a new total mortality equation (constant catch ODE (CCODE) model) uses an ODE with constant within-year catch per unit time for each age class, and as a result allows catch to be removed directly from the population.

The corresponding estimation models to the BODE and CCODE population models fit the number of recaptures by age class using a Poisson likelihood and use maximum likelihood estimation. However, the BODE model requires an extra component to the likelihood that compares actual and estimated catch-at-age numbers.

The BODE model is similar to the model of Polacheck et al. (2006) with the main differences being (i) the estimation of a fishing selectivity function, whereas Polacheck et al. model fishing mortality rates for all combinations of fishing year and age class so that their model of numbers caught is 'saturated' with parameters, (ii) the assumption of a lognormal distribution for catch-at-age numbers compared to the normal (i.e. Gaussian) assumption used by Polacheck et al. (2006), and (iii) a fundamental difference in the definition of the likelihood for the number of recaptures-at-age.

The CCODE model does not require estimation of annual fishing mortality rates. Instead, parameters are estimated exclusively from the recapture-at-age numbers conditional on release-at-age numbers and estimation of fishing selectivity as a function of age is not required. Therefore, the CCODE model can be considered a 'semi-parametric' model relative to the 'fully parametric' BODE model.

The properties of the maximum likelihood estimates (MLEs) for the BODE and CCODE models were studied using simulation with the simulation model applied corresponding to underlying population dynamic equations for each of these models.

In practice, catch-at-age numbers are estimated, typically, using a census of total catch weight, a weight versus length regression, a random sample of measured lengths, a sub-sample of these which are aged, and the application of age–length keys (Quinn and Deriso, 1999). The effects of the resultant observational (i.e. estimation) errors in catch-at-age numbers on the bias and precision of the estimate of *M* were investigated for a subset of the simulations by incorporating realistic levels of observational error in the data used for estimation for each of the models. This approach can be used to carry out a bias adjustment to estimates of *M* if the bias is of practical significance.

Modelling population dynamics using total mortality and catch equations

In order to estimate *M* from aged mark-recapture data, it is necessary (as shown in the next section) to model the dynamics of the population both prior to and during fishing years. The dynamics before fishing are required because the age-structure first encountered by the fishery is a result of the previous history of recruitments (assumed here of age-1 recruits) combined with the effect of the assumed constant *M* in the absence of fishing. For a single cohort the numbers-at-age, N_a, can be obtained by integrating the following simple first-order ODE over time, t', within the fishing year defined from year t to $t +1$ (considering for the moment only a single fishing year and assuming each age class is fully selected by the fishery)

$$\frac{dN_{t',a}}{dt'} = \left(-F^{(\delta t')} - M^{(\delta t')}\right)N_{t',a} \tag{1}$$

where $F^{(\delta t')}$ and $M^{(\delta t')}$ are the constant unit-time, $\delta t'$, instantaneous fishing mortality and natural mortality rates respectively. Equation (1) is the Baranov (1918) ODE and when integrated from time t to $t +1$ where there are $\delta t'^{-1}$ time units for each time unit on the t time scale (e.g. t may have units of years and t' units of days so that $\delta t' = 1/365$) in this interval, gives

$$N_{t+1,a+1} = N_{t,a} \exp\left(-F^{(\delta t')} / \delta t' - M^{(\delta t')} / \delta t'\right)$$
$$= N_{t,a} \exp\left(-F - M\right) \tag{2}$$

where $F = F^{(\delta t')} / \delta t'$ and $M = M^{(\delta t')} / \delta t'$.

Using the annual time scale and equation (2) gives the usual annual Baranov population dynamic equation which, when combined with a catch equation (Baranov, 1918; see also Quinn and Deriso, 1999), given that annual fishing mortality, F, is allowed to vary by year of fishing and across age classes by combining F with a fishing selectivity function, S_a, gives (see Branch, 2009a, 2009b):

$$N_{t+1,a+1} = N_{t,a} \exp\left(-S_a F_t - M\right) \tag{3}$$

$$C_{t,a} = \frac{S_a F_t}{M + S_a F_t} N_{t,a} \left\{1 - \exp\left(-S_a F_t - M\right)\right\} \tag{4}$$

where $N_{t,a}$ is the number in the population for a given year t and age a and $C_{t,a}$ is the number in the catch for that given fishing year and age.

The total mortality equation, or numbers-at-age difference equation, is therefore given by

$$N_{t,a} - N_{t+1,a+1} = N_{t,a} \left\{1 - \exp\left(-S_a F_t - M\right)\right\}. \tag{5}$$

Equations (4) and (5) are denoted the BODE model since they are derived from the Baranov ODE.

Note that it is usual to assume that both the form and parameter values for the function for S_a do not vary across fishing years. For years prior to fishing, the value of F_t is simply set to zero. The scaled selectivity function, S_a', with Figure 1 giving examples of dome-shaped and sigmoidally shaped functions combined with a lower threshold at age 3, is usually specified so that for one or more ages it has a maximum of 1. However, in the above equation it is appropriate to define $S_a = AS_a' / \sum_{a=1}^{A} S_a'$, where there are A age-classes considered for the population (i.e. so that S_a has a mean of 1), rather than as $S_a = S_a'$ (see Appendix 1). To estimate F_t from equation (4) using known numbers caught in each age class, assuming M and S_a are fixed, requires a recursive numerical solution since an exact algebraic solution is not available.

An alternative to the Baranov catch equations is proposed here and is denoted the CCODE model since its ODE incorporates constant absolute fishing mortality by age class per unit time. It is derived as follows:

Let this first-order ODE be given by

$$\frac{dN_{t',a}}{dt'} = -M^{(\delta t')} N_{t',a} - c_a^{(\delta t')} \tag{6}$$

where $c_a^{(\delta t')}$ is the unit-time catch of age class a which is assumed constant within year t with the corresponding total catch for year t in age class a given by $C_{t,a} = c_a^{(\delta t')} / \delta t'$. If the time unit is a day then $1/\delta t'$ is 365 if the full year is fished. The integral of the above ODE between time t and $t+1$ is given by

$$N_{t+1,a+1} = N_{t,a} \exp(-M) \\ -C_{t,a}\{1-\exp(-M)\}M^{-1}. \tag{7}$$

The population numbers-at-age difference equation is therefore given by

$$N_{t,a} - N_{t+1,a+1} = N_{t,a}\{1-\exp(-M)\} \\ +C_{t,a}\{1-\exp(-M)\}M^{-1}. \tag{8}$$

Using equation (8) as the difference equation for updating the cohort size at year t has the advantage that there is no need for a catch equation since $C_{t,a}$, the absolute fishing mortality in equation (8), is taken as a known constant. Therefore there is no need to estimate a fishing mortality rate, which must be defined given S_a with respect to the unknown $N_{t,a}$. Note that if the fishing period is defined to not be the whole year but the latter q fraction of the year, then this two-period equation can be defined as

$$N_{t+(1-q),a} = N_{t,a}\exp\left(-M^{(\delta t')}[1-q]/\delta t'\right)$$

$$N_{t+1,a+1} = N_{t+(1-q),a}\exp\left(-M^{(\delta t')}q/\delta t'\right) \\ -c^{(q\delta t')}(q/\delta t') \\ \left\{1-\exp\left(-M^{(\delta t')}q/\delta t'\right)\right\}\left(M^{(\delta t')}q/\delta t'\right)^{-1} \tag{9}$$

where $c^{(q\delta t')} = c^{(\delta t')}/q$. Combining these equations by substituting for $N_{t+(1-q),a}$ in the second gives

$$N_{t+1,a+1} = N_{t,a}\exp(-M) \\ -C_{t,a}\{1-\exp(-M')\}M'^{-1} \tag{10}$$

where $M' = qM$. So the form of the equation is the same for a partially fished year, however, the fractional reduction in M must be applied in the adjustment to the catch-at-age numbers given the fractional reduction from a full year to the length of the fishing season.

Both the BODE and CCODE are idealised models, however, given that the F's are small, the BODE model can be made to give a daily catch that does not vary much from the constant catch, $c_a^{(\delta t')}$, in a given year. This can be achieved by appropriately scaling $F^{(\delta t')}$ given $N_{t,a}$ and S_a given that the annual catch, equation (4), is well approximated by $S_a F_t N_{t,a}$ (Appendix 1). So although the daily catch can be made similar for the two ODEs, the crucial difference is that they lead to very different forms and parameterisations for the total mortality equation as seen by comparing equations (4) and (5) with equation (8).

To generalise the above models to a multi-cohort setting, the simple time notation, t, needs to be replaced by a year, y, notation including pre-fishing

and fishing years. In order to apply equations (3) and (4) recursively from the start of fishing at year $y = t_0$, then the starting numbers-at-age for the population $\underline{N}_{t_0} = \left\{ N_{t_0,1},...,N_{t_0,A} \right\}$ must be estimated. Note that given fish caught in year $y = t$ of age a, then the year they were age-1 and recruited to the population can be determined as $t - a + 1$. The starting population, $\underline{N}_{t_0}$, can be estimated if an estimate of mean age-1 recruitment, K_0, is available, combined with an estimate of the series of YCS, $\Gamma_{y_{pf}}$, for the pre-fishery (*pf*) years combined with the first year of the fishery where $\underline{y}_{pf} = \left\{ t_0 - A + 1,...,t_0 \right\}$. Since recruitment to age-class 1 obviously continues once fishing begins, then additional YCS estimates, Γ_{y_f} need to be estimated for fishing years $\underline{y}_f = \left\{ t_0 + 1,...,t_0 + n_f - 1 \right\}$ where n_f is the total number of fishing years. To assist in obtaining an average close to 1 for the combined set of YCS estimates, $\underline{\Gamma} = \left\{ \Gamma_{pf}, \Gamma_f \right\}$, for the $(A + n_f - 1)$ cohorts $\underline{y}' = \left\{ \underline{y}_{pf}, \underline{y}_f \right\}$, a penalty function is added to the log-likelihood as described later. Since $K_{y'} = K_0 \Gamma_{y'}$ this ensures that average annual recruitment of age-1 fish for these years is close to K_0. This assumes that fishing mortality is not high enough to affect recruitment. If this is not the case, then a stock-recruitment relationship (Quinn and Deriso, 1999) should be employed to model this effect on $K_{y'}$ which then requires extra parameters to be estimated. This extra complication is not considered further, and in practice for a stock for which the spawning stock is not seriously depleted by fishing, the inclusion of a stock-recruitment relationship does not affect model outputs to any practically significant degree. Therefore, to reconstruct the age-structured population for each fishing year, the BODE model requires simultaneous estimates of *M*, the n_f *F*-parameters, K_0, the parameters of the selectivity function, and the $A + n_f - 1$ YCS parameters.

Basing estimation of *M* on equation (7), the CCODE model has the advantage that such estimation only requires simultaneous estimates of *M*, K_0, and the $A + n_f - 1$ YCS parameters.

Modelling number of aged recaptures

Development of within-year dynamic equations for the release/recapture process could follow along the lines of the process models for population and catch numbers-at-age (i.e. equations (3) to (10)). However, the process is made difficult by

the requirement that releases should have adequate time to mix with the untagged population in order that sampling the population by commercial fishing to estimate the proportion of tagged fish at liberty (referred for the remainder as the 'tagged proportion') can be used to obtain unbiased estimates of the population size for each age class and year of fishing. Allowing an adequate time for mixing is more important for species that are largely sedentary, such as Patagonian toothfish. Therefore, approximate annual difference equations are used to model numbers of releases and recapture numbers by age class and year of fishing. Even though the simulation model and estimation procedure assume instantaneous and complete mixing of tagged fished with the untagged population, the difference equations can still be validly applied by modifying the number of releases and recaptures when this assumption is relaxed as described below. This requires a fixed minimum period to be specified from time of release to recapture in order to allow mixing to be sufficient that recapture rates are not significantly exaggerated by the fact that fishing is not a spatially random process.

The likelihood is developed first for the number of recaptures. Representing end of each year by $\underline{y} = \left\{ t_0 + 1,...,t_0 + n_f \right\}$, the response random variable is defined as the number of recaptures of $R_{y_i,a}$ in *i*th year of fishing years $\underline{y}$ and recapture age *a*. The $R_{y_i,a}$ can be disaggregated according to the contribution of each year of release so that $R_{y_i,a} = \sum_{v=1}^{i} r_{y_v,y_i,a}$ where $r_{y_v,y_i,a}$ is the number of recaptures of fish released in year y_v (for $v \leq i$) that are recaptured in year y_i when they are age a in year y_i. The total number of age a fish caught in year y_i given by $C_{y_i,a}$ is considered known without error and is used to determine the conditional distribution of $R_{y_i,a}$. The number of tagged fish available to be caught in year y_i is given by $G_{y_i,a}$. However, $G_{y_i,a}$ is the sum of previous fishing years' releases that are still at liberty (i.e. have not died due to natural or fishing mortality up to year y_i) so this number by age class is not known exactly and in fact depends on the parameter *M*. The definition of expected number of recaptures-at-age and $G_{y_i,a}$ used discrete annualised values. The calculation of $G_{y_i,a}$ is given in Appendix 2.

If the unknown total number of fish in the population at age a (including both tagged and untagged fish) is $N_{y_i,a}$ then the hypergeometric likelihood

for the observed recaptures, considering that $R_{y_i,a}$ has a hypergeometric distribution for each separate year and age class conditional on $G_{y_i,a}$ and $N_{y_i,a}$, is given by

$$L_R = \prod_{y_i,a} \binom{G_{y_i,a}}{R_{y_i,a}} \binom{N_{y_i,a} - G_{y_i,a}}{C_{y_i,a} - R_{y_i,a}} \binom{N_{y_i,a}}{C_{y_i,a}}^{-1} . \quad (11)$$

The above likelihood can be reasonably approximated by a binomial distribution in terms of expected value and second central moment for $R_{y_i,a}$, since both distributions have the same expected value and if $C_{y_i,a}$ is small relative to $N_{y_i,a}$ then both have approximately the same variance. Therefore, using the approximation of a binomial with expected tagged proportion given by $\tau_{y_i,a}$ where

$$\tau_{y_i,a} = G_{y_i,a} N_{y_i,a}^{-1} \quad (12)$$

the binomial likelihood is given by

$$L_R = \prod_{y_i,a} \binom{C_{y_i,a}}{R_{y_i,a}} \tau_{y_i,a}^{R_{y_i,a}} \left(1 - \tau_{y_i,a}\right)^{C_{y_i,a} - R_{y_i,a}} . \quad (13)$$

Further simplification of the binomial likelihood to a Poisson likelihood is justified since the expected tagged proportion is typically very small, which gives

$$L_R = \prod_{y_i,a} \frac{1}{R_{y_i,a}!} \exp\left(-C_{y_i,a}\tau_{y_i,a}\right)\left(C_{y_i,a}\tau_{y_i,a}\right)^{R_{y_i,a}} . \quad (14)$$

Note that the expected Poisson number of recaptures-at-age in year y_i can be expressed equivalently as either $G_{y_r,a}P_{y_r,a}$, where $P_{y_i,a} = C_{y_i,a}N_{y_i,a}^{-1}$, or as given in equation (14) by $C_{y_i,a}\tau_{y_i,a}$. The Poisson variance for $R_{y_i,a}$ is given by its expected value, $C_{y_i,a}\tau_{y_i,a}$, however, the variance for $R_{y_i,a}$ conditional on $C_{y_i,a}$ as a binomial is slightly less since it scales the Poisson variance by $\left(1 - \tau_{y_i,a}\right)$. Since $\left(1 - \tau_{y_i,a}\right)$ is typically very close to 1, then these two variances are approximately the same.

The kernel of the –2 log-likelihood corresponding to equation (14) is given by

$$-2\ln\left(L_R\right) =$$
$$-2\sum_{y_i,a}\left\{R_{y_i,a}\ln\left(C_{y_i,a}\tau_{y_i,a}\right) - C_{y_i,a}\tau_{y_i,a}\right\} . \quad (15)$$

In simulation studies this last form of the likelihood for number of recaptures was found to result in successful maximum likelihood estimation. Due to the instability of the binomial likelihood when values of $\tau_{y_i,a}$ were vanishingly small for age classes with zero or very small values of selectivity, S_a, the Poisson $-2\ln(L_R)$ given by equation (15) was preferred. Note that zero observed recaptures combined with very low expected recaptures due to unpopulated young (<4 years) and old (>25 years) age classes in the catch due to zero or very low selectivity in these age classes (Figure 1), contribute a negligible amount to $-2\ln(L_R)$ in equation (15) and therefore make an insignificant contribution to estimation.

Since $N_{y_i,a}$ is unknown, its value is required to be estimated using the model parameters of M, K_0, and $\Gamma = \left\{\Gamma_{pf}, \Gamma_f\right\}$ and, additionally for the BODE model, the n_f F-parameters, and the parameters of the selectivity function. For the following, $N_{y_i,a}$ is the number of fish in the population at the end of fishing year y_i in age class a.

Appendix 3 gives the calculation of $N_{y_i,a}$ for each of the BODE and CCODE models.

Modelling catch-at-age numbers

The catch-at-age number for the BODE model is estimated as

$$\hat{C}_{y_i,a} = \frac{\hat{S}_a\hat{F}}{\hat{S}_a\hat{F} + \hat{M}} \hat{N}_{y_i,a}\left\{1 - \exp\left(-\hat{S}_a\hat{F}_i - \hat{M}\right)\right\} \quad (16)$$

noting that this estimate requires estimates of M, the selectivity parameters required to give $\hat{S}_a$, and the F's estimated from the previous iteration of the maximum likelihood fitting algorithm.

The likelihood component for the catch-at-age numbers for the BODE model is assumed that for a lognormal distribution for $C_{y_i,a}$ as follows

$$L_C =$$

$$\prod_{y_i,a} \frac{1}{\sqrt{2\pi\sigma_C^2}} \exp\left[-0.5 \left\{ \begin{array}{c} \ln\left(C_{y_i,a}\right) \\ -\ln\left(\hat{C}_{y_i,a}\right) \end{array} \right\}^2 / \sigma_C^2 \right]. \quad (17)$$

Therefore the additional parameter σ_C is the coefficient of variation (CV) of $C_{y_i,a}$ about its prediction and is required to be estimated by the BODE model. If $C_{y_i,a}$ is not subject to observational error, then there is only one error involved in equation (17) and that is process error given by $\varepsilon_{y_i,a} \equiv \ln\left(C_{y_i,a}\right) - \ln\left(\hat{C}_{y_i,a}\right)$. If catch-at-age numbers are subject to observational error with the true catch given by $C_{y_i,a}$ and the observed catch by $\tilde{C}_{y_i,a}$, then observational error is given by $\xi_{y_i,a} \equiv \ln\left(\tilde{C}_{y_i,a}\right) - \ln\left(C_{y_i,a}\right)$. It is not possible to model the $\xi_{y_i,a}$ without constructing a comprehensive 'error budget' by modelling all data inputs and calibration errors involved in constructing $\tilde{C}_{y_i,a}$ which would require much more complicated likelihoods than equations (15) and (17). A less rigorous but much simpler approach, that can also be applied to the CCODE model, is to model the effect of a simplified observational error process on estimates of *M* by effectively simulating the $\xi_{y_i,a}$ to obtain $\tilde{C}_{y_i,a}$ and then use these numbers to replace $C_{y_i,a}$ in the estimation algorithms. This is the approach adopted for both BODE and CCODE models as described later.

Note that for the CCODE model there is no corresponding likelihood component for catch-at-age numbers since the actual catch-at-age numbers are used to calculate $N_{y_i,a}$ (Appendix 3). Therefore, the annual fishing mortality rates and the fishing selectivity-at-age given by S_a are not part of the CCODE model parameterisation.

Estimation methods

The parameters of each model were fitted by minimising the –2 log-likelihood (*L*) using the `nlminb` function in the R-package (R Development Core Team, 2006), where this requires minimising *L* ($=-2\ln L_R$) for the CCODE model and $L\left[=-2\left(\ln L_R + \ln L_C\right)\right]$ for the BODE model with

respect to the parameter set for each model. The R_0 parameter was estimated on the log scale. A penalty function to help assist in achieving an average of the estimated YCS parameters that is close to 1, given by $10000\left\{mean\left(\hat{\Gamma}\right) - 1\right\}^2$ was added to the above –2 log-likelihood.

When all model parameters were simultaneously estimated, point estimates for *M* were obtained both by joint minimisation of *L* giving MLEs and for comparison by profiling across a grid of values for *M* while estimating all other model parameters simultaneously to give the profile maximum likelihood estimate (PMLE) of *M*. These two methods are equivalent in well-behaved maximum likelihood estimation (McCullagh and Nelder, 1989), in the sense that minimisation of *L* should not get 'trapped' in potential local minima when simultaneous estimation of all parameters is carried out. The profiling method can protect against finding local minima and further allows approximate 95% confidence bounds for *M* to be easily obtained for the CCODE model.

The upper and lower bounds for parameters used in the `nlminb` function are given later. The starting value for *M* passed to `nlminb` when it was the only parameter estimated (i.e. all other parameters defined for the respective BODE or CCODE models were set at their nominal simulation values) was randomly chosen at a point between the bounds. When all parameters were estimated, the starting values for these were obtained as random values independently chosen between their respective bounds except for the YCS parameters which were set to start at 1.0. When *M* was profiled while jointly estimating the remaining parameters, the above applied but obviously the initial value of *M* was retained at the incremental value on the grid chosen for the profiling.

The simulation model

The simulation models for each of the BODE and CCODE models are each described for a series of 12 fishing years starting in calendar year 2010, with this number of fishing years chosen to reflect the length of time commercial fishing has been carried out in Division 58.5.2. Given ages up to 35 were simulated, this gave year classes or cohorts (*c*) of 1 to 46 corresponding to calendar years of 1975 to 2020 for which YCS was estimated.

First a population age structure was generated by calculating the probability density for each of 1 001 values of age, a', taken uniformly between age 0 and age 35 (i.e. consecutive age values 35/1 001 years apart) using the exponential density function $\Pr(Age = a') = Q(a') = Me^{-Ma'}$. The raw density values, $Q(a')$, were scaled by dividing by their sum to give $Q'(a_i') = Q(a_i') / \sum_{i=1}^{1001} Q(a_i')$. Using the 35 integer age classes given by $\{(0,1],...,(34,35]\}$, denoting the vector $\underline{a}$ as the upper limit of each class, the class membership of each of the 1 001 ages was determined and used to determine class probabilities. This numerical integration method of determining the age structure was used in preference to the method that uses the simple analytical integral since the former allows more accurate calculation of class probabilities when fishing selectivity is incorporated as described next.

Since non-trivial fishing selectivity was imposed, $Q'(a)$ was then multiplied by S_a' to give $Q''(a)$ where S_a' was determined from the 'double-normal' (DN) (i.e. dome-shaped) selectivity function combined with zero selectivity for ages of a_0 and less. This gives

$$
\begin{aligned}
S_a' &= 2^{-[(a-\lambda)/\sigma_L]^2} \quad ; \ a_0 < a \le \lambda \\
&= 2^{-[(a-\lambda)/\sigma_U]^2} \quad ; \ a > \lambda \\
&= 0 \quad\quad\quad\quad\ \ ; \ a \le a_0
\end{aligned}
\tag{18}
$$

where λ is a cut-point parameter corresponding to the age at which S_a' is 1, and σ_L and σ_U are parameters denoting the standard deviations of the scaled normal density functions specifying the lower and upper arms of the function respectively. For the base-line simulations and dome-shaped selectivity (equation 18), the parameters a_0, λ, σ_L and σ_U were set to 3, 5, 2 and 8 respectively.

For sigmoidally shaped (see below) selectivity, S_a' was determined from the 'lower-normal' (LN) selectivity function combined with zero selectivity for ages of a_0 and lower by simply modifying equation (18) to give

$$
\begin{aligned}
S_a' &= 2^{-[(a-\lambda)/\sigma_L]^2} \quad ; \ a_0 < a \le \lambda \\
&= 1 \quad\quad\quad\quad\ \ ; \ a > \lambda \\
&= 0 \quad\quad\quad\quad\ \ ; \ a \le a_0.
\end{aligned}
\tag{19}
$$

For the base-line simulations and sigmoidally shaped selectivity in equation (19), the parameters a_0, λ and σ_L were set to 3, 8 and 5 respectively. Note that equation (19) demonstrates a classical sigmoidal shape if λ is sufficiently large relative to a_0 and σ_L is sufficiently small relative to $\lambda - a_0$.

The integer-age population class probabilities were obtained by accumulating the sum of the $Q''(a)$ using their integer-age class membership to give Q_a where $a = 1,...,35$.

Figure 1 shows the values of S_a' obtained from each of equations (18) and (19) using the corresponding parameter values given above.

The second step of the simulation involved generating number of releases-at-age from simulated sampling of the population age structure. For each of the 12 years in a simulation run, release-at-age frequencies were generated for a sample of 100 hauls where across hauls the expected number of fish released was 10 giving 1 000 releases across age classes per year. For each haul the numbers per age class were drawn from the Dirichlet-multinomial distribution as described in Candy (2008), with proportions in age classes obtained from the selectivity function. The over-dispersion parameter ϕ was set to 1.1. The numbers of releases were aggregated across hauls to give $V_{y_i,a}$. The numbers of recaptures per release year, recapture year and age class, $r_{y_v,y_i,a}$, were drawn from a binomial distribution (see Appendix 4) with binomial sample size, $g_{y_v,y_i,a}$, equal to the number of tagged fish at liberty from release year, y_v, and age class, a, in the current fishing (i.e. recapture) year, y_i, where $G_{y_i,a} = \sum_{y_v \le y_i} g_{y_v,y_i,a}$, and probability, $P_{y_i,a}$, defined earlier.

Population numbers-at-age, $N_{y_i,a}$, for the BODE model were obtained using equation (A3.1) while for the CCODE model the $N_{y_i,a}$ were obtained using equation (A3.2). For the case of catch-at-age numbers assumed to be known without observational error, the simulated numbers were used

directly in estimation for both CCODE and BODE models. The total numbers of catch-at-age for each year were 500 000 for the CCODE model with these numbers disaggregated to age classes using the selectivity function to give the $C_{y_i,a}$. The catch-at-age numbers for the BODE model simulations were obtained using equation (4).

Observational error in catch-at-age numbers

To investigate the effect of observational error on estimation of *M*, after the simulated data were obtained and before the $C_{y_i,a}$ were passed to the estimation algorithms that maximise the likelihood, the $C_{y_i,a}$ were perturbed with observational error to obtain $\tilde{C}_{y_i,a}$. This perturbation was carried out in three stages. Firstly, the total catch numbers per year were drawn from a lognormal distribution with CV of either 0.1 or 0.2 with expected value the sum of the simulated values; $\sum_a C_{y_i,a}$. Secondly, the perturbed proportion in each age class for each year were obtained from separate multinomial distributions for each year with expected proportions given by $C_{y_i,a} / \sum_a C_{y_i,a}$ and 'effective' multinomial sample size of either 1 000 or 1 500. Note that the lower the effective sample size (ESS) the greater the degree to which the proportions are perturbed. These ESSs were realistic values for a catch of 500 000 fish per year obtained by comparison to the average ESS of approximately 2 500 for the approximately 1 million fish taken per year in the main trawl fishery in Division 58.5.2 as given by Candy and Welsford (2009) with the method of calculating the ESS for catch-at-age frequencies given in Candy (2009). The total of ESSs for years 1998 to 2008 calculated for this fishery was 27 785 (Table A1.5 of Candy and Welsford, 2009). Finally, perturbed numbers of catch-at-age, $\tilde{C}_{y_i,a}$, were obtained by multiplying the random lognormal value of total catch number by the above perturbed proportions. Figure 2 gives an example using a simulation of exact, $C_{y_i,a}$, and corresponding perturbed, $\tilde{C}_{y_i,a}$, numbers of catch-at-age when the CV was set to 0.2 and the ESS set to 1 000 using the CCODE model and LN selectivity (equation 19). Ignoring any zero values in $C_{y_i,a}$, the difference in the log of perturbed and log of exact catch-at-age numbers defines the observational error, $\xi_{y_i,a}$. As a

control (i.e. very close to zero observational error) a CV of 0.01 combined with an ESS of 20 000 was also applied.

Nominal parameter values

For all baseline simulations, annual recruitment was varied for each run of the simulation model as a lognormal variate about a mean of 2 million with CV (c_R) of 0.3 and with an initialisation period of 35 years.

For the BODE model all *F*'s were set to 0.03 while for the CCODE model total catch numbers were set for each fishing year to 500 000.

The parameters and their bounds were *M* (0.05, 0.3), K_0 (1.0 x 10^6, 3.0 × 10^6), Γ (i.e. YCS parameters) (0.1, 3.0), *F*'s (0.01, 0.1) and σ_C (0.2, 2.0). For selectivity function (18) the parameter ranges were λ (4, 8), σ_L (1, 8) and σ_U (4, 16). For selectivity function (19) the parameter ranges were λ (6, 12) and σ_L (3, 12). Note that the a_0 parameter was not estimated but assumed known for the fit of the BODE model.

For the simulation model, tag mortality rate, tag loss rate, detection rate, and *q* were assumed to be 0.0, 0.0, 1.0 and 1.0 respectively.

Each estimation model (BODE or CCODE) was fitted to its corresponding simulation model as described above. The simulation models were coded in the R-package. Except where noted otherwise, the simulations were carried out without the incorporation of observational error in catch-at-age numbers used in estimation.

Results

Table 1 gives summary statistics for estimates of *M* from 500 simulation runs using selectivity function (equation 18) for each nominal value of *M* of 0.10, 0.13, and 0.16 when estimation of *M* was carried out with all other model parameters fixed at their known simulated values (i.e. 'Fix non_*M*').

Figures 3 to 6 show results of estimation for one run of the base-line simulation model for selectivity function (equation 18) using the BODE model with a nominal value of *M* of 0.13. Figure 3 shows the profile –2 log-likelihood (*L*) for *M* along with the loess smoothed curve (R-function `loess`)

Table 1: Comparison of accuracy of BODE and CCODE models/estimation methods for runs of the corresponding simulation models with double-normal (DN) (dome-shaped) selectivity and M the only estimated parameter.

Estimation model	Nominal M	Annual number releases	Statistics for 500 MLEs of M				
			Mean	CV[a]	5–95% range in %error[b]	Mean bias (%)[c]	RMSE (%)[d]
BODE	0.10	1000	0.0939	0.0232	−6.92, −5.54	−6.07	6.50
Fix non_M[e]	0.13	1000	0.1248	0.0261	−4.35, −3.46	−4.00	4.78
	0.16	1000	0.1558	0.0051	−2.90, −2.33	−2.60	2.65
CCODE	0.10	1000	0.0974	0.0512	−8.45, 2.86	−2.65	5.77
Fix non_M[e]	0.13	1000	0.1301	0.0307	−3.24, 3.37	0.10	3.07
	0.16	1000	0.1631	0.0245	−0.79, 4.62	1.92	3.11

[a] Coefficient of variation
[b] %error $= 100\left(\hat{M} - M\right)/M$
[c] Mean of %error
[d] %RMSE (root mean square error) $= 100\sqrt{(\%bias\,/\,100)^2 + CV^2}$
[e] Estimate only M while setting other parameters to known simulation values.

fitted to the values of L. The PMLE of M obtained by taking the minimum of the fitted loess curve from Figure 3 is close to 0.1. However, the MLE of M obtained by jointly estimating all parameters was 0.224 with value of L of −18 301. Figures 4 to 6 show some important aspects of the fit using the jointly estimated MLEs. Figure 4 shows the observed and expected number of recaptures across years and age classes, showing the 1:1 line along with a Poisson regression fitted as a generalised linear model (GLM) (McCullagh and Nelder, 1989) through the origin. Note that this fitted GLM has nothing to do with the estimation of M and simply compares the 1:1 line to a line of best fit. Figure 5 shows simulated and estimated YCS parameters. The average of the YCS parameter estimates over all year classes shown in Figure 9 was 1.0145. Figure 6 shows the actual and estimated total catch numbers (i.e. summed across age classes) for each fishing year. The estimate of K_0 was 3 093 133 for a nominal simulated value of 2 118 444. The corresponding estimates of the F's were 0.0426, 0.0424, 0.0429, 0.0440, 0.0432, 0.0433, 0.04239, 0.0410, 0.04172, 0.0401, 0.0395 and 0.0395 while the estimates of parameters (λ, σ_L, σ_U, σ_C) were 5.91, 5.08, 10.86 and 0.309 respectively. Therefore, the nominal values of (λ, σ_U) of 5 and 8 were recovered quite well, however, σ_L, R_0 and the F's were substantially over-estimated with nominal values of 2, 2 million and 0.03, while Figure 6 shows that the estimated total catch numbers deviated systematically from the actual values. The profile −2 log-likelihood graph (Figure 3) indicates that estimation suffers

from a degree of instability. In a well-defined maximum likelihood estimation problem such a curve is usually much smoother (as seen below for the CCODE model) without the high degree of oscillation of values of L about the smooth (loess) fitted curve. Furthermore, it should achieve a PMLE of M (i.e. the value that minimises L) close to the MLE with a similar value of L in each case.

Figures 7 to 9 show the corresponding results for a single base-line simulation of the CCODE model ($M = 0.13$) for selectivity function (equation 18) and corresponding estimation to that of Figures 3 to 5. Figure 7 shows a PMLE of M of close to 0.146 while the MLE was 0.135 with a similar minimum of L of −18 797. The estimate of K_0 corresponding to the PMLE of M was 2 488 608, whereas the simulated value was 2 185 693. There was a very high positive correlation between profiled values of M and the corresponding estimates of K_0 of 0.99. Note that the profile is much smoother than that obtained for the BODE model and simulation. The dashed line in Figure 7 represents an approximate 95% confidence bounds around the PMLE of M giving bounds of (0.130, 0.164) which represents, in percentage terms, bounds of (−10.96, 12.33). These bounds were obtained by comparing the increase in L either side of the PMLE and comparing this to the 95% chi-square single degree of freedom deviate multiplied by an over-dispersion parameter. There was no evidence of over-dispersion (see below), so the over-dispersion parameter was set to 1.

Figure 8 shows the observed and expected numbers of recaptures across years and age classes, including the 1:1 line and the Poisson regression fitted as a GLM through the origin. Figure 9 shows the simulated and estimated YCS parameters. The average of the YCS parameter estimates over all year classes shown in Figure 9 was 1.0028. Note that in Figure 9, when accurate estimates of the simulated YCS parameter for given years cannot be obtained (e.g. there are too few fish captured and retained, and/or captured, released and recaptured in older age classes to give information on the earliest years of the series), then the estimates are shrunk towards their expected value of 1. The penalty function encourages this desirable behaviour, but even when the penalty function was excluded, this behaviour of estimated YCS parameters was retained.

Figure 10 shows the estimated linear regression between residual variance and mean expected number of recaptures for binned values using bin classes for expected number of recaptures of 5 units between 0 and 30 with upper bins of 30–40 and >40. The residual variance was calculated as the variance of observed minus expected number of recaptures for values in the each bin (recaptures shown in Figure 8). A gamma GLM with identity link weighted by the number of residuals in each bin minus 1 was fitted through the origin, and the slope of the regression gives an estimate of the over-dispersion parameter. The variance and corresponding mean for the 0–5 bin were excluded from the regression to eliminate the effect on the regression of excess zero observed recaptures combined with very low expected recaptures in this class due to unpopulated young (<4 years) and older (>25 years) age classes resulting from the zero or very low selectivity in these age classes (Figure 1). A comparison of the 1:1 line (solid line) in Figure 10, which represents Poisson variation, compared to the regression line indicates under-dispersion relative to that expected for a Poisson with given expected number of recaptures. A minor degree of under-dispersion would be expected since the number of recaptures-at-age was simulated as $B\left(G_{y_i,a}, P_{y_i,a}\right)$ and therefore has a variance equivalent to a Poisson variance scaled by the factor $\left(1 - P_{y_i,a}\right)$ where $P_{y_i,a}$ for fully selected age classes could be as high as 0.06.

Table 2 shows summary results of estimates of *M* for 500 simulation runs for DN selectivity function (equation 18) of each of the BODE and CCODE models and nominal *M* of 0.13. The results in Table 2 were obtained when (a) all parameters were estimated simultaneously, (b) the K_0 parameter was fixed at its simulated value (which takes a random value for each simulation run that is lognormally distributed about a mean of 2 million with CV 0.3) for the CCODE model, (c) for the BODE model, all other parameters except those of the selectivity function were estimated, so that the λ, σ_L and σ_U parameters were fixed at their nominal simulation values, and (d) all parameters estimated simultaneously for the BODE model but with the LN selectivity function (equation 19) fitted, representing a severely misspecified selectivity function.

Table 3 shows summary results of estimates of *M* for 500 simulation runs for LN selectivity function (equation 19) used for each of the BODE and CCODE models and nominal values of *M* of 0.10, 0.13 and 0.16. The results in Table 3 were obtained when all parameters were estimated simultaneously. Table 3 also shows the results for 500 simulation runs when the BODE model was fitted using a minor misspecification of the LN selectivity function. This misspecification involved simulating with equation (19) with a_0 set to 3 years, as in the base-line simulations, but fitting the BODE model with a_0 set to zero. This in effect extends the lower tail of the LN to age 1, giving higher selectivity for ages 1 to 3 than that simulated with the magnitude of this increase dependent on the estimates of λ and σ_L.

Table 4 shows summary results of estimates of *M* for 500 simulation runs for LN selectivity function (equation 19) used for each of the BODE and CCODE models and nominal value of *M* of 0.13 for a range of levels of observational error incorporated in the catch-at-age numbers used in maximum likelihood estimation. Table 4 also gives corresponding estimates of σ_C for the BODE model.

Figure 11 shows the simulated and estimated YCS parameters for a single simulation run and estimation for the CCODE model incorporating the LN selectivity function (equation 19) in the simulation model.

Discussion

Table 1 shows that for the most difficult scenario for successful estimation of *M* of dome-shaped

Table 2: Comparison of accuracy of BODE and CCODE models/estimation methods estimating combinations of all parameters for double-normal (DN) selectivity and nominal M of 0.13.

Model	Fixed[a] parameter	S'_a	Statistics for 500 MLEs of M					MLE of K_0 mean bias (%)[d]
			Mean	CV[b]	5–95% range in %error[c]	Mean bias (%)[d]	RMSE (%)[e]	
CCODE	K_0	DN[f]	0.1316	0.0871	−14.24, 14.14	1.25	8.80	na
BODE	$\lambda, \sigma_L, \sigma_U$	DN[f]	0.1228	0.0837	−16.23, 5.91	−5.51	10.02	5.77
BODE	None	DN[f]	0.1601	0.2858	−42.07, 53.85	23.13	36.77	13.75
BODE	None	DN[h]	0.1889	0.0978	17.33, 53.84	45.30	46.34	11.90
CCODE	None	DN[f]	0.1548	0.1005	0.09, 35.65	19.08	21.56	25.56

[a] Parameters set to known simulation values
[b] Coefficient of variation
[c] %error $= 100\left(\hat{M} - M\right)/M$
[d] Mean of %error
[e] %RMSE (root mean square error) $= 100\sqrt{\left(\%bias/100\right)^2 + CV^2}$
[f] Double normal, equation (18)
[h] Fitted selectivity function severely misspecified. Simulation used DN selectivity (equation 18) with BODE model fitted assuming lower-normal (LN) selectivity (equation 19).

Table 3: Comparison of accuracy of BODE and CCODE models/estimation for lower-normal (LN) selectivity.

Model	Nominal M	S'_a	Statistics for 500 MLEs of M					MLE of K_0 mean bias (%)[c]
			Mean	CV[a]	5–95% range in %error[b]	Mean bias (%)[c]	RMSE (%)[d]	
BODE	0.10	LN[f]	0.0950	0.0855	−17.82, 8.95	−5.00	9.91	0.51
BODE	0.10	LN[g]	0.0884	0.1501	−35.41, 12.34	−11.58	18.96	−7.30
CCODE	0.10	LN	0.0993	0.1485	−27.64, 20.58	−0.73	14.87	15.76
BODE	0.13	LN[f]	0.1266	0.0709	−12.06, 7.97	−2.64	7.57	−0.67
BODE	0.13	LN[g]	0.1167	0.1082	−26.26, 5.94	−10.25	14.91	−10.55
CCODE	0.13	LN	0.1250	0.1379	−30.41, 13.30	−3.85	14.32	13.24
BODE	0.16	LN[f]	0.1542	0.0725	−12.76, 8.29	−1.75	7.46	−0.13
BODE	0.16	LN[g]	0.1399	0.1024	−28.60, 0.85	−12.54	16.19	−14.86
CCODE	0.16	LN	0.1621	0.0958	−13.68, 12.95	1.31	9.67	22.51

[a] Coefficient of variation
[b] %error $= 100\left(\hat{M} - M\right)/M$
[c] Mean of %error
[d] %RMSE (root mean square error) $= 100\sqrt{\left(\%bias/100\right)^2 + CV^2}$
[f] Lower-normal selectivity equation (18) simulated and correctly specified for the fit of the BODE model.
[g] Minor misspecification of fitted selectivity function. Simulation used LN selectivity (equation 18) $a_0 \equiv 3$ but BODE model fitted assuming $a_0 \equiv 0$.

Table 4: Comparison of accuracy of BODE and CCODE models/estimation for lower-normal (LN) selectivity for nominal *M* of 0.13 and different levels of observational error in catch-at-age numbers.

| Model | CV[a] total catch | ESS[b] | Statistics for 500 MLEs of *M* | | | | | MLE of K_0 mean bias (%)[e] | MLE of σ_C mean |
			Mean	CV[c]	5–95% range in %error[d]	Mean bias (%)[e]	RMSE (%)[f]		
BODE[g]	0.00	-	0.1266	0.0709	-12.06, 7.97	-2.64	7.57	-0.67	0.368
BODE	0.01	20 000	0.1273	0.0698	-11.65, 8.50	-2.11	7.29	0.20	0.450
BODE	0.10	1 500	0.1289	0.0716	-12.83, 9.05	-0.82	7.21	0.85	0.684
BODE	0.10	1 000	0.1295	0.0802	-15.03, 10.05	-0.40	8.03	0.87	0.920
BODE	0.20	1 500	0.1290	0.0788	-12.89, 9.91	-0.80	7.92	2.35	0.703
BODE	0.20	1 000	0.1297	0.0732	-12.81, 10.04	-0.26	7.33	2.42	0.919
CCODE[g]	0.00	-	0.1250	0.1379	-30.41, 13.30	-3.85	14.32	13.24	-
CCODE	0.01	20 000	0.1265	0.1390	-29.89, 14.22	-2.71	14.16	13.55	-
CCODE	0.10	1 500	0.1262	0.1451	-34.12, 14.95	-2.93	14.80	14.47	-
CCODE	0.10	1 000	0.1239	0.1568	-36.89, 13.98	-4.73	16.40	13.38	-
CCODE	0.20	1 500	0.1251	0.1521	-35.23, 16.27	-3.74	15.66	14.60	-
CCODE	0.20	1 000	0.1246	0.1543	-34.80, 14.98	-4.12	15.97	14.86	-

[a] Coefficient of variation of observational error for total numbers caught, $\sum_a C_{y,a}$

[b] Effective sample size (ESS) for multinomial sample of catch-at-age proportions quantifying observational error in catch proportion-at-age.

[c] Coefficient of variation

[d] $\%error = 100 \left(\hat{M} - M \right) / M$

[e] Mean of %error

[f] %RMSE (root mean square error) $= 100 \sqrt{ \left(\%bias / 100 \right)^2 + CV^2 }$

[g] No observational error, repeated from Table 3 for ease of comparison.

selectivity, both BODE and CCODE models give accurate estimates of M when this was the only parameter estimated. The estimates of M for the BODE model were more precise than those of the CCODE model but had a greater bias. Overall, the accuracy as measured by the percent root mean square error (%RMSE, Table 1) was similar for the two models ranging from 3% to 7%.

To investigate the estimation of M when all parameters were simultaneously estimated, initially a single simulation run for each model with nominal M of 0.13 and DN selectivity (equation 18) was studied in detail. Both joint MLE and PMLE were obtained. The profile maximum likelihood approach ensures that the full range of M is investigated and that this is desirable for achieving a global minimum for L. The values of L when M was profiled were much better behaved (i.e. smooth and quadratic in shape) for the CCODE model (Figure 7) than the BODE model (Figure 3). The nominal value of M is at the lower end of the 95% confidence bound shown in Figure 7. The profile likelihood is therefore a useful indicator of the degree of success of estimation even if some bias remains.

The results in Table 2 for 500 simulation runs and a single nominal value of M of 0.13 for the BODE model, when DN selectivity (equation 18) was simulated and correctly specified in the estimation algorithm, show that if the selectivity function parameters are fixed at their simulated values, then approximately unbiased and reasonably precise estimates of M were achieved. However, when all parameters were simultaneously estimated, there was substantial positive bias and poor precision for the estimates of M. This became more extreme when the functional form of selectivity was misspecified by simulating with selectivity function (equation 18) but replacing this selectivity function with function (equation 19) in the estimation.

Similarly, for the CCODE model, when K_0 was fixed at its nominal simulated value and all other parameters estimated (i.e. M and the YCS parameters) the CCODE model/estimation gave close to unbiased and reasonably precise estimates of M. However, when all parameters, including K_0, were estimated the bias was also positive and substantial although improved compared to the BODE model

in terms of both bias and precision. There was a corresponding substantial positive bias in the joint estimate of K_0.

Table 3 shows results for nominal values of M of 0.10, 0.13 and 0.16 when LN selectivity (equation 19) was simulated. Both BODE and CCODE models gave close to unbiased and reasonably precise (CVs of 0.07 to 0.14) estimates of M for all three nominal values of M when all parameters for each model were simultaneously estimated. The precision of the BODE model was substantially better than the CCODE model when the selectivity function was correctly specified. However, for all three nominal values of M when the minor misspecification of the selectivity function was applied, the precision was only slightly better than that of the CCODE model while the bias was considerably greater, which, in combination, gave a slightly better %RMSE for the CCODE model estimates.

Table 4 shows results for a nominal value of M of 0.13 when LN selectivity (equation 19) was simulated and catch-at-age numbers were perturbed to account for observational error. For comparison, the corresponding results obtained when zero observational error was assumed have been reproduced from Table 3. It can be seen that for the BODE model, accounting for variations due to sampling 500 simulation runs, there is no detectable additional bias or imprecision in estimates of M across the range of values of CV and ESS. Note that the estimate of process error variance, σ_C^2, in equation (17) has 'absorbed' the additional variance associated with observational error. The estimates of σ_C in Table 4 demonstrate this with the reduction in ESS more influential than the increase in CV on the additional variance absorbed by the estimate of σ_C. This follows since annual estimates of F can largely absorb annual variation in total catch numbers arising from any CV imposed as a component of observational error.

For the CCODE model, the additional bias and imprecision due to observational error is detectable but still represents a relatively minor increase, with the maximum value of the increase of close to 2% when the CVs in Table 4 are expressed as percentages. The median of the estimates of M (not given in Table 4) was much closer to the nominal value of 0.13 than the mean with the difference due to the greater degree of negative skew of esti-

mates compared to the BODE model as seen in the 5% and 95% quantiles for percent error shown in Table 4.

Figure 11 shows that YCS parameter estimates account for more of the variation in the years well before the start of the fishery in 2010 when LN selectivity is simulated than the corresponding estimates for the DN selectivity function, since this last selectivity function only obtains a very small amount of information about these cohorts due to the very few fish above age 20 that are caught (Figure 1).

Polacheck et al. (2006) investigate the issue of estimation of *M* using catch-at-age and aged recapture data using a similar approach to that of the BODE model. Instead of the lognormal likelihood for the catch-at-age numbers they use a Gaussian log-likelihood which, for a given fishing year, is given by

$$\log(L_C) = \sum_i -\tfrac{1}{2}\left\{\log\left(\sigma_i^2\right) + \frac{1}{\sigma_i^2}\left(C_i - E(C_i)\right)^2\right\}$$

where the *i* subscript refers to the age class. However, they found that they could not estimate σ_i^2 even if a common-across-ages value, $\sigma_i^2 \equiv \sigma^2$, was assumed and in their application they fixed the coefficient of variation $CV = \sigma_i/E(C_i)$ to 0.3. Using a CV in this way makes the above component of the log-likelihood more similar to that corresponding to equation (17). The difficulty in estimating σ^2 is related to the way they 'saturated' the model for catch numbers-at-age by fishing year with fishing mortality rate parameters. Assuming constant fishing selectivity allowed the corresponding CV here (i.e. σ_C) to be successfully estimated by the BODE model, as shown in the example simulation/estimation presented where the estimate was 0.309. Additionally, the simulations incorporating observational error showed how this parameter 'absorbs' this additional error to that of process error (Table 4).

The BODE model requires a number of strong assumptions, in particular, that numbers caught-at-age can be adequately modelled using the *F*-parameter estimates and the selectivity function which in this last case is assumed to maintain the same functional form and parameter values across fishing years. The benefit of this approach is that, if these assumptions are satisfied and any variation

across years in the selectivity parameters is random, then to the degree that such random variation is small, information in the numbers caught-at-age on YCS parameter values can be exploited by this approach. This explains why the BODE model gave superior precision to the CCODE model when sigmoidally shaped (LN) selectivity was simulated. However, this advantage over the CCODE model was only realised when the form of selectivity was known and correctly specified in the BODE model fitting algorithm (Table 3).

Basing estimation of *M* on equation (7) (or (8)), the CCODE model has the advantage that such estimation only requires simultaneous estimates of *M*, K_0 and the $A + n_f - 1$ YCS parameters. It follows that, since the numbers caught-at-age are known and treated as known in the CCODE model, there is no need to have competing components in the log-likelihood, one for the catch data and one for numbers of recaptures, and avoids the resultant difficulty of correctly weighting each component. This allows profile likelihood confidence intervals for *M* to be easily constructed. The disadvantage of the CCODE model is that the estimate of *M* is highly positively correlated with the estimate of K_0, which, like selectivity, is another quantity that is difficult to independently estimate.

A general conclusion is therefore that the 'fully parametric' BODE model gives improved precision compared to the 'semi-parametric' CCODE model but performance quickly degrades when the selectivity function is misspecified. Since fishing selectivity is defined with reference to the unknown age structure in the population, yet is at the same time used to estimate this age structure, it is a difficult quantity to specify and validate as a function of age. The results in Table 3 showed that, when a minor misspecification of the functional form of selectivity was fitted using the BODE model, the above advantage over the CCODE model disappeared. Therefore, an advantage of the CCODE model is that it is robust to the unknown form of the selectivity function as long as the older aged fish are well selected (i.e. both models performed poorly when this was not the case).

Both models performed poorly when older age classes were not well selected, as quantified using the DN selectivity function (equation 18) and the relatively small value of σ_U (i.e. the older age classes have selectivity closer to 1 as σ_U is increased). For

such domed-shaped selectivity the inability for the gear to catch older (and therefore larger) fish suggests that they are not available to be caught by the gear which could be explained by these fish having emigrated from the study area (i.e. fishing ground). When the degree of this unavailability or 'emigration' was pre-specified by fixing the selectivity function parameters for the BODE model, or the K_0 parameter for the CCODE model, at their simulated values, estimation of M was reasonably accurate (Table 2). However, the property of both models to substantially over-estimate M when this type of selectivity was simulated, and all parameters simultaneously estimated, suggests that a lack of fish caught in older age classes was attributed by the estimation to higher natural mortality and less to what was actually simulated, which was a 'cryptic biomass' of older fish. This indicates that in the case of dome-shaped selectivity and the BODE model that selectivity parameter estimates are confounded with the estimate for M. For the CCODE model this confounding is between the estimates of K_0 and M.

In contrast, when sigmoidally shaped selectivity in the form of equation (18) was simulated, this problem did not occur and both models recovered the simulated value of M reasonably accurately, with small bias and quite high precision, with CVs ranging between 0.07 and 0.14 (Tables 3 and 4).

As noted earlier, the CCODE model does not account for the fact that in practice catch-at-age numbers are not known exactly. On the other hand, the BODE model assumes the estimation errors are purely 'process' errors (i.e. model lack-of-fit), however, these errors will be 'polluted' with observational error (Table 4). Therefore, the BODE model/ estimation as described does not allow a rigorous decomposition of the total variance into the appropriate contributions from the different sources of error. However, given a sufficiently large sample of random lengths and individual weights, combined with a sufficiently large sub-sample of aged fish, the simulations that included realistic levels of observational error indicated that the bias and imprecision in estimates of M attributable to observational error in catch-at-age numbers (Table 4) is not a significant issue relative to other sources of uncertainty, such as the form of selectivity function.

The values of ESS for catch-at-age proportions studied of 1 000 and 1 500 fish may seem large, however, a number of factors make these numbers realistic for the target fishery described in Candy et al. (2011). Table 1 in Candy et al. (2011) documents that over 11 years in this fishery some 158 616 fish were measured for length and 4 447 of these were aged. Further, using the method of Candy (2009), Candy and Welsford (2009) estimate the total of ESS over years 1998 to 2008 for this fishery at 27 785 (Table A1.5 of Candy and Welsford, 2009). Note also that the calculation of ESS has not been reduced by the effect of process error (Candy, 2008). This is because there is no process error involved in determining observational error variance for the catch-at-age numbers since it is calculated before any model of these numbers is fitted. To see this, note that if all fish in the catch were counted and correctly aged there would be zero observational error irrespective of model-based process error. Also, the combined effect on observational error of both the ESS and the CV that was placed on total catch numbers per year of 0.1 and 0.2 (assuming a lognormal distribution) should be considered. These values of CV are at the larger end of the scale, considering how total catch numbers per year are calculated (i.e. a census of catch weights, a large sample of lengths to determine mean length per cruise, and a typically very precise relationship between individual fish weight and length). Figure 2 shows that considerable variation in perturbed catch-at-age numbers (i.e. with observational error) about assumed-known catch-at-age numbers has been generated. Typically, the random length-frequency samples are large while the sub-sample sizes of aged fish are much smaller. The annual ESS for the catch-at-age proportions sits somewhere between the size of the ESS for the length samples and the size of the aged sample, given that age–length keys are applied. The stronger the relationship between length and age, the closer the catch-at-age ESS will be to that of the length-frequency ESS.

Even though the issue of such observational error should not be exaggerated, the BODE model showed less effect of these errors on M-estimation when these errors were modelled in this simplified approach. To generalise the BODE model to directly incorporate all the abovementioned sources of data and their sampling and/or measurement errors, is a much more complex proposition than that considered here. If sample sizes are not

adequate to allow these errors to be ignored, then a simulation-based bias adjustment as described above is a simple, but numerically intensive, solution (Carroll et al., 1995; see also Candy, 2002).

As mentioned in the Introduction, the assumption of a constant *M* across all age classes, as applied in the simulation studies, is unrealistic. However, as discussed earlier, the –2 log-likelihood (equation 16) concentrates most of the information on *M* from the well-populated age classes, in terms of catch, release and recapture numbers. Therefore, although a constant *M* is assumed, it is the value most applicable to these well-populated age classes. As a result, the potential for under-estimation of *M* for the youngest age classes that are not vulnerable to fishing would lead to some bias in estimates of population size. However, if there are few such age classes before they become vulnerable to fishing, the size of this effect should be relatively minor.

In addition to the above caveats, in practice, only the length of released fish that are not recaptured can be measured. So although the simulation study assumed numbers-at-age that were released were known without error, in practice these numbers must also be estimated using age–length keys. Other complexities in applying these models in practice include the difficulty in accounting for tag-induced mortality, tag loss rates (mentioned previously), tag reporting rates, and the potentially substantial effect of IUU catch. Candy et al. (2011) applied the estimation methods developed in this study to actual data from the main trawl ground for the Heard and McDonald Islands Patagonian toothfish fishery (Candy and Constable, 2008) for fishing years 1998 to 2008.

Acknowledgements

References

Baranov, F.I. 1918. On the question of the biological basis of fisheries. *Nauch. Issledov, Iktiol. Inst. Izv.*, 1: 81–128.

Branch, T.A. 2009a. Differences in predicted catch composition between two widely used catch equation formulations. *Can. J. Fish. Aquat. Sci.*, 66 (1): 126–132.

Branch, T.A. 2009b. Corrigendum: Differences in predicted catch composition between two widely used catch equation formulations. *Can. J. Fish. Aquat. Sci.*, 66 (9): p. 1631.

Candy, S.G. 2002. Empirical binomial sampling plans: model calibration and testing using William's method III for generalized linear models with overdispersion. *JABES*, 7 (3): 373–388.

Candy, S.G. 2008. Estimation of effective sample size for catch-at-age and catch-at-length data using simulated data from the Dirichlet-multinomial distribution. *CCAMLR Science*, 15: 115–138.

Candy, S.G. 2009. Incorporating sample variation and random reader error into calculation of effective sample size in the application of age length keys to estimation of catch-at-age proportions. Document *WG-SAM-09/8*. CCAMLR, Hobart, Australia.

Candy, S.G. and A.J. Constable. 2008. An integrated stock assessment for the Patagonian toothfish (*Dissostichus eleginoides*) for the Heard and McDonald Islands using CASAL. *CCAMLR Science*, 15: 1–34.

Candy, S.G. and D.C. Welsford. 2009. Update of the integrated stock assessment for the Patagonian toothfish (*Dissostichus eleginoides*) for the Heard and McDonald Islands (Division 58.5.2). Document *WG-FSA-09/20*. CCAMLR, Hobart, Australia.

Candy, S.G., D.C. Welsford, T. Lamb, J.J. Verdouw and J.J. Hutchins. 2011. Estimation of natural mortality for the Patagonian toothfish at Heard and McDonald Islands using catch-at-age and aged mark-recapture data from the main trawl ground. *CCAMLR Science*, 18: 29–45.

Carroll, R.J., D. Ruppert and L.A. Stefanski. 1995. *Measurement Error in Nonlinear Models*. Chapman and Hall, London: 305 pp.

Constable, A.J. and W.K. de la Mare. 1996. A generalised model for evaluating yield and the long-term status of fish stocks under conditions of uncertainty. *CCAMLR Science*, 3: 31–54.

Hillary, R.M., G.P. Kirkwood and D.J. Agnew. 2006. An assessment of toothfish in Subarea 48.3 using CASAL. *CCAMLR Science*, 13: 65–95.

Leigh, G.M., W.S. Hearn and K.H. Pollock. 2006. Time-dependent instantaneous mortality rates from multiple tagging experiments with exact times of release and recovery. *Environ. Ecol. Stat.*, 13 (1): 89–108.

McCullagh, P. and J.A. Nelder. 1989. *Generalized Linear Models*. Second Edition. Chapman and Hall/CRC, USA.

Polacheck, T., J.P. Eveson, G.M. Laslett, K.H. Pollock and W.S. Hearn. 2006. Integrating catch-at-age and multiyear tagging data: a combined Brownie and Petersen estimation approach in a fishery context. *Can. J. Fish. Aquat. Sci.*, 63 (3): 534–548.

Quinn T.J. II and R.B. Deriso. 1999. *Quantitative Fish Dynamics*. Oxford University Press, Oxford: 542 pp.

R Development Core Team. 2006. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. www.R-project.org.

Shepherd, J.G. 1999. Extended survivors analysis: An improved method for the analysis of catch-at-age data and abundance indices. *ICES J. Mar. Sci.*, 56 (5): 584–591.

Wang, Y.B., Q. Liu, C. Yu and L. Lian. 2009. Estimation of instantaneous rate of natural mortality from catch-at-age data and an abundance index based on extended survivors method. *Fish. Res.*, 97 (1–2): 127–133.
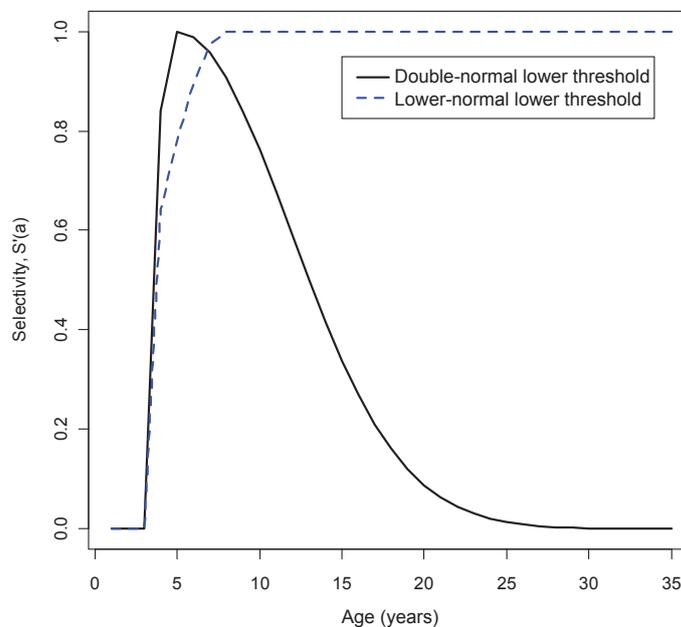
Figure 1: Double-normal (DN) (solid line) and lower-normal (LN) (dashed-line) selectivity functions, S'_a, with a lower threshold of age 3 above which selectivity is greater than zero. Curves correspond to parameter values used in baseline simulations.
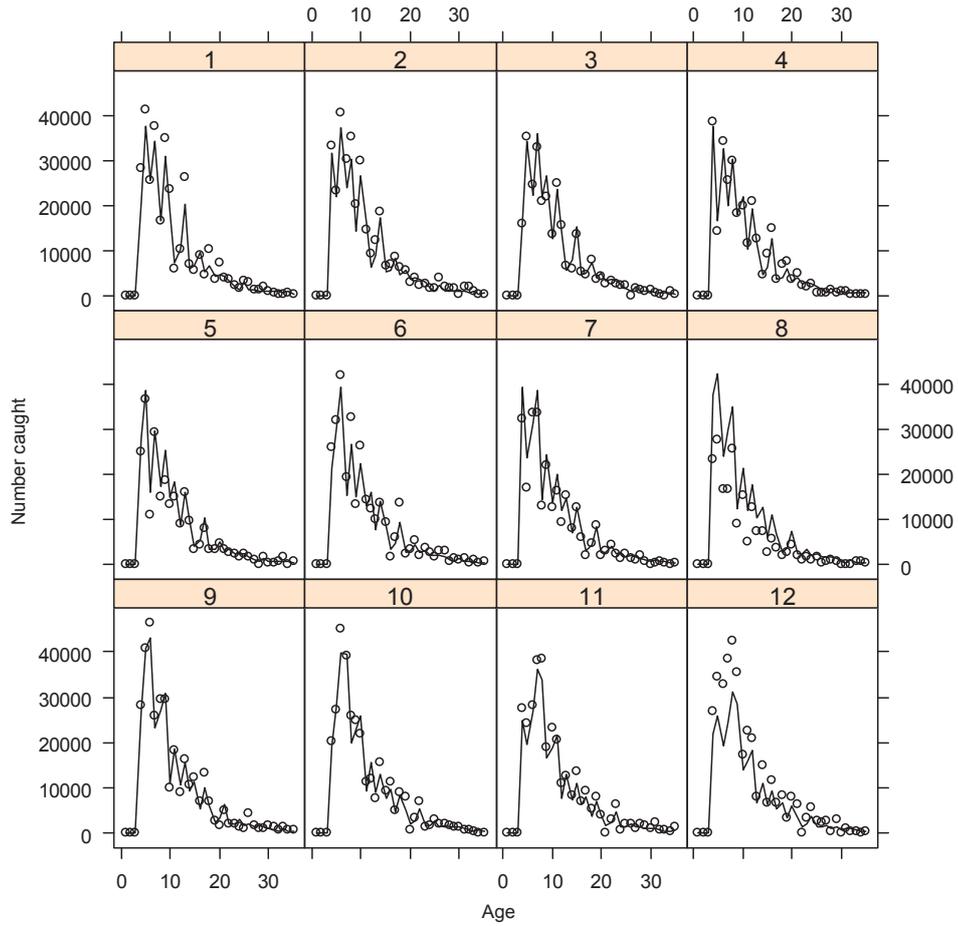
Figure 2:   Example of simulated numbers of catch-at-age for each fishing year (line) and corresponding numbers perturbed by observational error (points) with CV of 0.2 and effective sample size (ESS) of 1 000 using the CCODE model and lower-normal (LN) selectivity (equation 19).
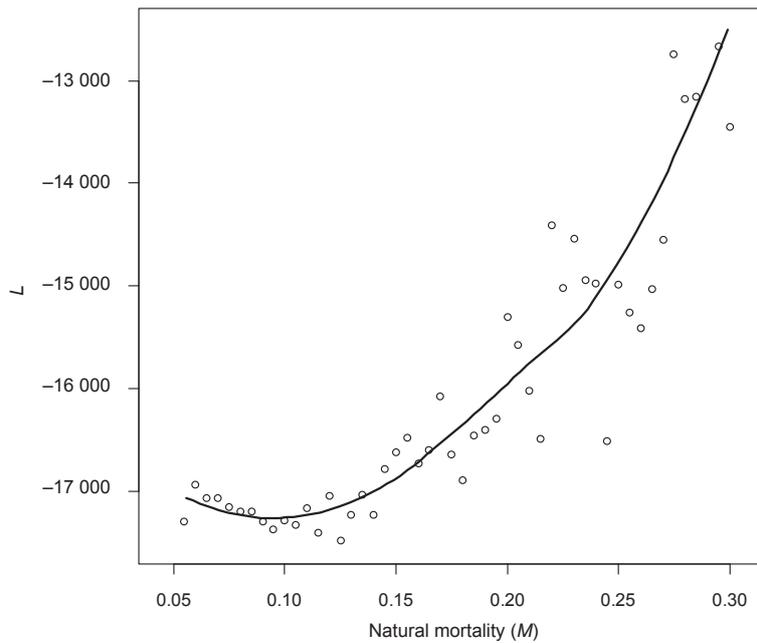

Figure 3:   Profile –2 log-likelihood for *M* for an individual base-line simulation (*M* = 0.13) with double-normal (DN) selectivity (equation 18) for the BODE model showing the loess smoothed curve (R-function `loess`) fitted to the values of *L*.
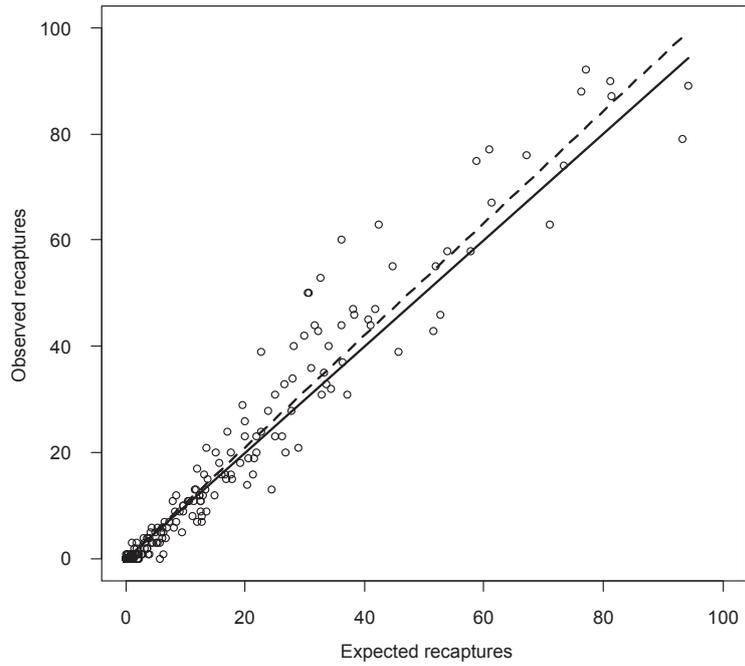
Figure 4: Observed and expected (i.e. estimated) number of recaptures across years and age classes for an individual base-line simulation ($M = 0.13$) with double-normal (DN) selectivity (equation 18) for the BODE model at the maximum likelihood estimate (MLE) of 0.224. The 1:1 line (solid line) and fitted GLM regression line (dashed line) are shown.
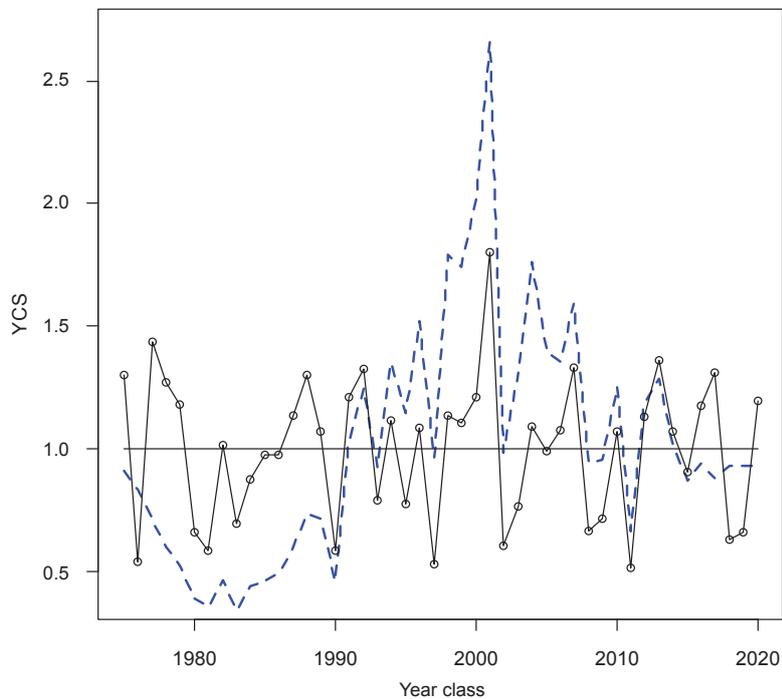

Figure 5: Estimated year-class strength (YCS) with year classes shown as calendar years for an individual base-line simulation ($M = 0.13$) with double-normal (DN) selectivity (equation 18) for the BODE model at the maximum likelihood estimate (MLE) estimate of 0.224 showing simulated values (circles, solid line) and estimated values (dashed line).
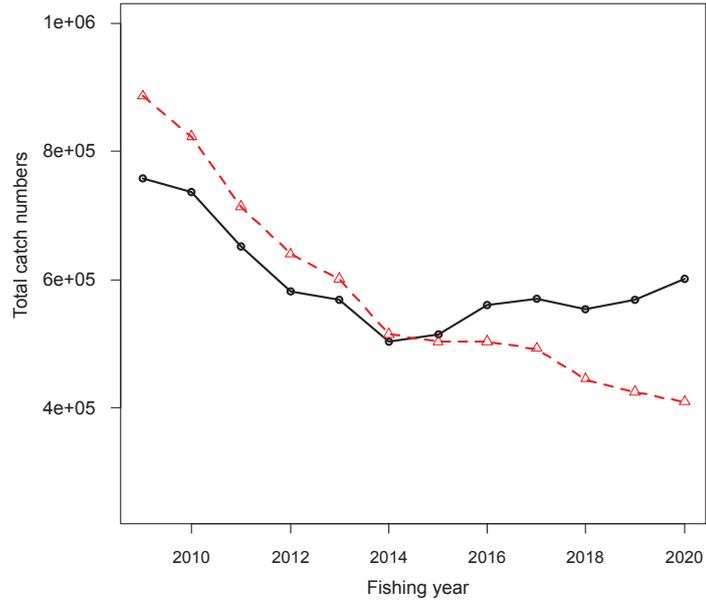
Figure 6: Actual (circles, solid line) and estimated (triangles, dashed line) total catch number for an individual base-line simulation (*M* = 0.13) with double-normal (DN) selectivity (equation 18) for the BODE model at the maximum likelihood estimate (MLE) of 0.224.
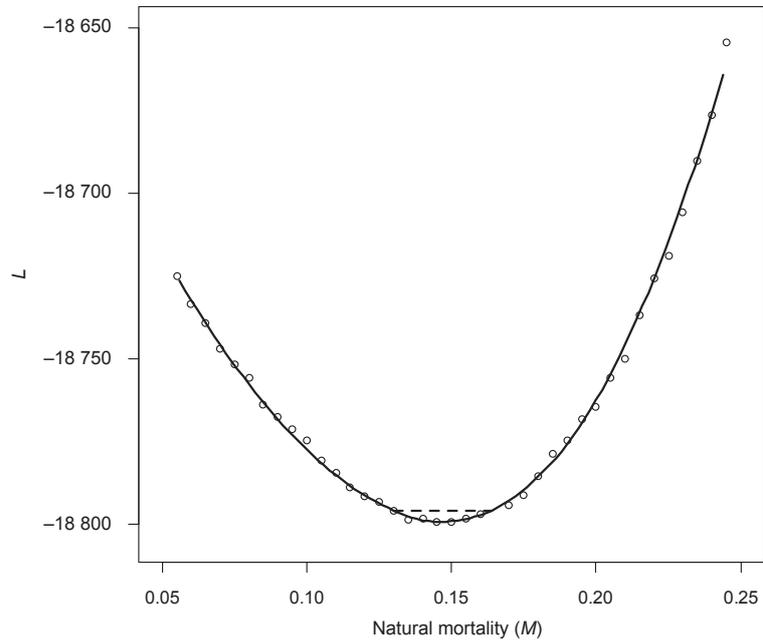
Figure 7: Profile −2 log-likelihood for *M* for an individual base-line simulation (*M* = 0.13) with double-normal (DN) selectivity (equation 18) for the CCODE model showing the loess smoothed curve (R-function `loess`) fitted to the values of *L*. The dashed line delineates the approximate 95% confidence limit of the maximum likelihood estimate (MLE) of *M* assuming Poisson dispersion.
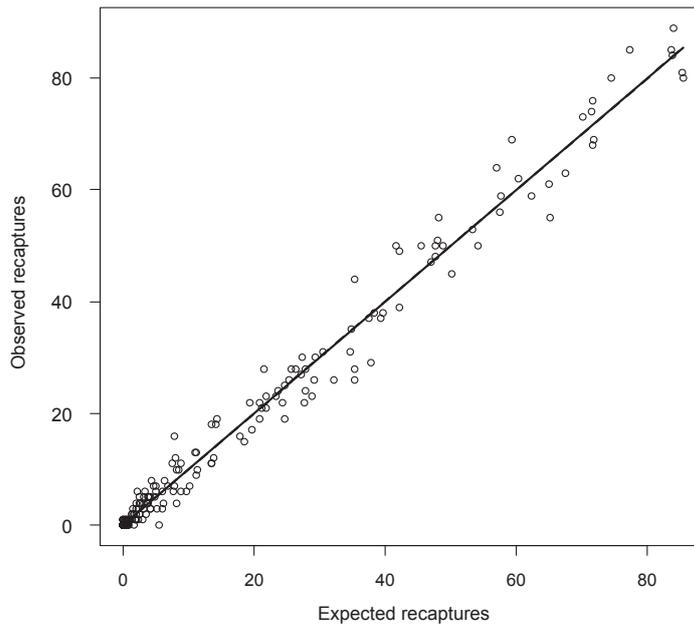
Figure 8:   Observed and expected (i.e. estimated) number of recaptures across years and age classes for an individual base-line simulation ($M = 0.13$) with double-normal (DN) selectivity (equation 18) for the CCODE model at the maximum likelihood estimate (MLE) of 0.135. The 1:1 line (solid line) and fitted GLM regression line (dashed line) are shown.
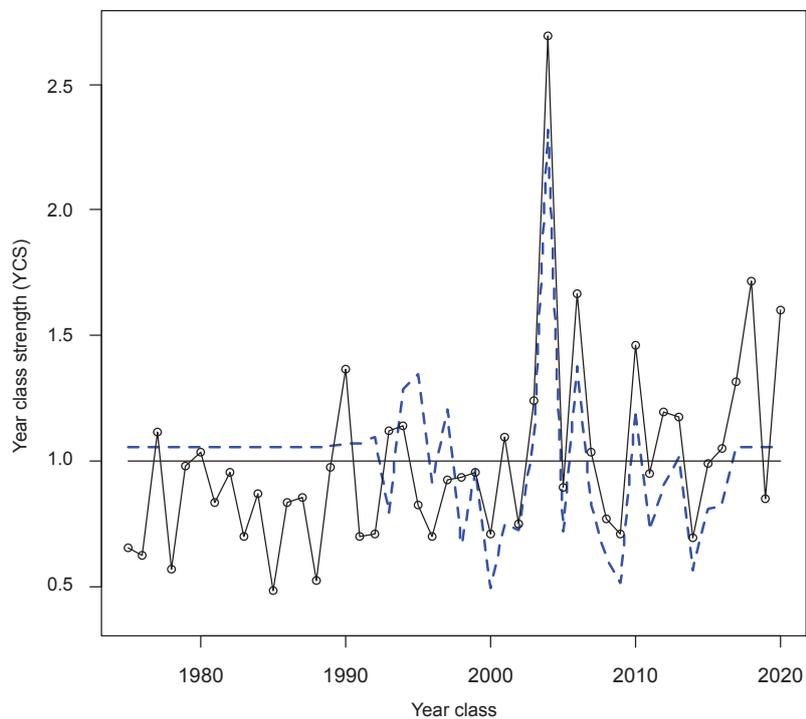

Figure 9:   Estimated year-class strength (YCS) with year classes shown as calendar years for an individual base-line simulation ($M = 0.13$) with double-normal (DN) selectivity (equation 18) for the CCODE model at the maximum likelihood estimate (MLE) of 0.135 showing simulated values (circles, solid line) and estimated values (dashed line).
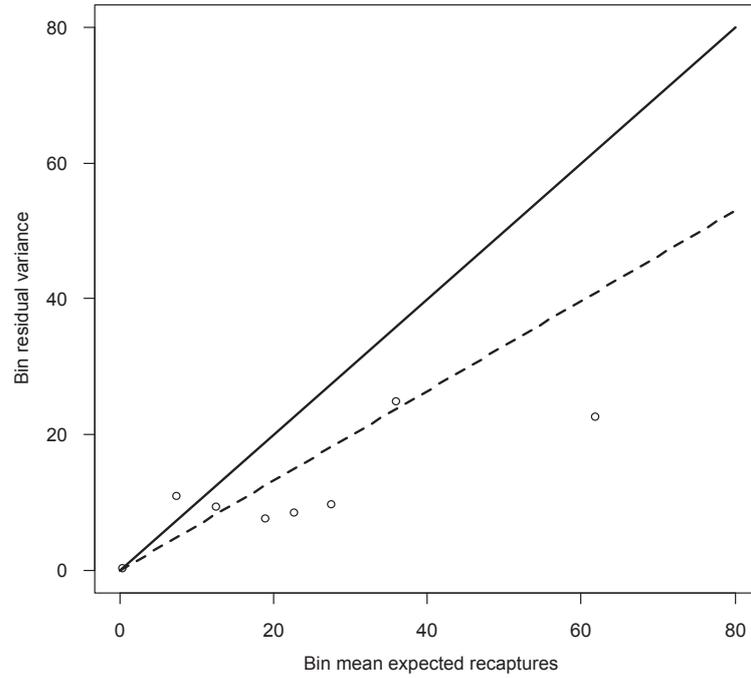
Figure 10:   Estimated linear regression (dashed line) between residual variance and mean expected number of recaptures for binned values (points) showing 1:1 line (solid line) for an individual base-line simulation (M = 0.13) with double-normal (DN) selectivity (equation 18) for the CCODE model at the maximum likelihood estimate (MLE) of 0.135. The 1:1 line represents Poisson variation.
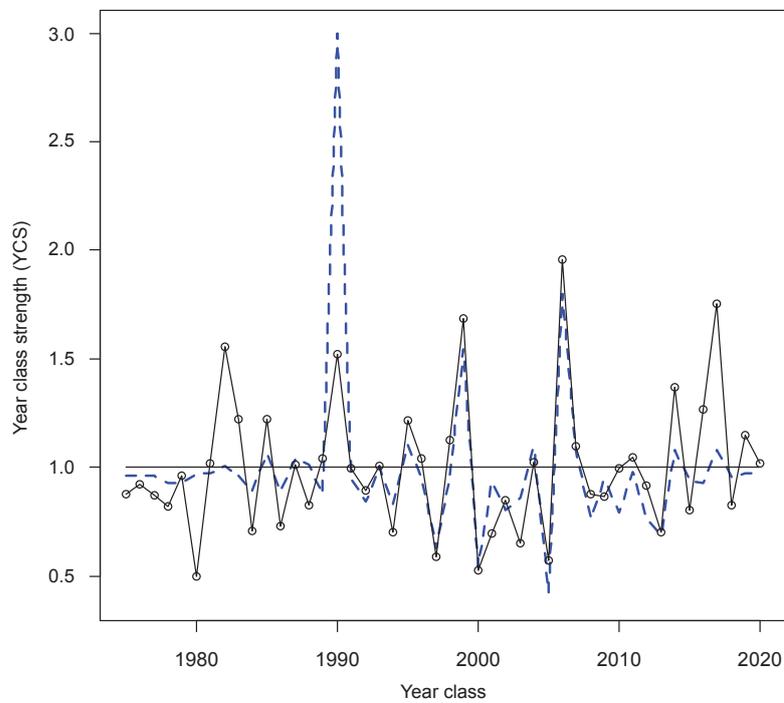

Figure 11:   Estimated year-class strength (YCS) for an individual simulation (M = 0.13) with lower-normal (LN) selectivity (equation 19) for the CCODE model showing simulated values (circles, solid line) and estimated values (dashed line).

SCALING THE SELECTIVITY FUNCTION

It is appropriate to define $S_a = A S_a' / \sum_{a=1}^{A} S_a'$, where there are A age-classes considered for the population so that S_a has a mean of 1 rather than a maximum of 1, with the latter case given by $S_a = S_a'$. This is because F_t is applied to the total population and not to the vulnerable population given by $\sum_a S_a' N_{t,a}$. That a mean of S_a of 1 is a more sensible constraint than $\max(S_a) = 1$, can be seen by considering the approximation of $\frac{S_a F_t}{M + S_a F_t} \{1 - \exp(-S_a F_t - M)\}$ by $S_a F_t$. This approximation is good for $S_a F_t$ ranging from 0.0 to 0.05, for example for $M = 0.14$ and $S_a F_t = 0.05$ then $\frac{S_a F_t}{M + S_a F_t} \{1 - \exp(-S_a F_t - M)\}$ is given by 0.046 while for $S_a F_t = 0.01$ the corresponding value is 0.0093. Therefore the catch can be approximated by $S_a F_t N_{t,a}$ and therefore if $\max(S_a)$ is 1 so that $S_a = S_a'$ and $\sum_a C_{t,a} = F_t \sum_a S_a' N_{t,a}$ then F_t is the fishing mortality rate relative to the fraction (≤ 1) of the population vulnerable to fishing mortality which is not how F_t is defined in equations (1) and (4). Note that scaling the S_a so that their mean is 1 has the effect that S_a will depend on the maximum number of age classes considered. Therefore, for the upper age classes that are sparsely populated, the point at which a maximum age, A, is chosen for input to the model may seem unnecessarily arbitrary. However, when the F_t's are estimated they will subsume any small differences in scaling due to different choices of A within a reasonable range (i.e. the range of A should cover all age classes expected to be present in the population). Given this caveat, the estimated catch will be unaffected by the value used for A.

Appendix 2

CALCULATION OF NUMBER OF FISH AVAILABLE FOR RECAPTURE

The calculation of $G_{y_i,a}$ is given as follows:

$$G_{y_i,a} = V_{y_i,a}, \quad i=1$$

$$G_{y_i,a} = \sum_{v=1}^{i}\left(V_{y_v,a} - r_{y_v,y_{i-1},a}\right)\exp\left\{-M\left(y_i - y_v\right)\right\}, \quad i=2,...,n_f \tag{A2.1}$$

where $V_{y_i,a}$ is the number of releases of aged *a* fish in fishing year y_i. In order for equation (A2.1) to give a sensible results, the $r_{y_v,y_i,a}$ must be set to zero for the nonsensical values in the case of $v > i$. This is required to define $r_{y_v,y_{i-1},a}$ when $v = i$ in equation (A2.1) in order for this equation to be expressed as simply as given above. Only recaptures in the previous year are removed from $V_{y_i,a}$ in equation (A2.1). However, fish released in year y_i are considered available for recapture in the same year. This assumes instantaneous and complete mixing of tagged fish with the untagged fish population. In practice, to allow time for adequate mixing, the number of releases by age class which are recaptured within a fixed time interval, which is preset as the minimum required to allow such mixing, should be removed from $V_{y_i,a}$ and $r_{y_v,y_i,a}$ before $R_{y_i,a}$ and $G_{y_i,a}$ are calculated. For example, Candy et al. (2011) apply a fixed time interval of 60 days. Note also in equation (A2.1) that natural mortality is applied only to the component of released fish that have not been recaptured by year y_i (i.e. $V_{y_v,a} - r_{y_v,y_{i-1},a}$) and that the subscripted age *a* for $G_{y_i,a}$ and $r_{y_v,y_{i-1},a}$ refers to the age-at-recapture while for $V_{y_i,a}$ it refers to the age of release. These are necessary logical constraints for constructing $G_{y_i,a}$.

The above value of $G_{y_i,a}$ given by equation (A2.1) assumes zero tagging-induced mortality, zero tag-loss rate and 100% detection of tags within catches. Equation (A2.1) needs to be modified if any of these assumptions are unrealistic. If a single dart tag is used, equation (A2.1) is easily modified by replacing *M* by $M + T$ where *T* is the annual tag-loss rate. However, double-tagging is more complex to take into account (see, e.g., Appendix D of Leigh et al., 2006 or Appendix 1 of Candy and Constable, 2008).

CALCULATION OF NUMBER OF FISH AT AGE BY FISHING YEAR IN THE POPULATION

For the BODE model, given K_0 and the combined set of YCS parameters, $\Gamma = \left\{ \Gamma_{pf}, \Gamma_f \right\}$ with YCS years of $\underline{y}' = \left\{ \underline{y}_{pf}, \underline{y}_f \right\}$ and fishing years $\left(y_1, \underline{y}_f \right)$ where y_1 is the first year of fishing then $N_{y_i,a}$ is obtained by summing over cohorts, indexed by c with corresponding year class y_c' (i.e. were recruited at age-1 in year y_c'), that are present in fishing year $y_i = \left(y_1, \underline{y}_f \right)_i$ to give

$$N_{y_i,a} = \sum_{c=1}^{a+i-1} K_0 \Gamma_{y_c'} \delta\left(y_i \in y_c' + a - 1 \right) \exp\left\{ -S_a \sum_{h=1}^{i} F_h - M(a-1) \right\}; \ i = 1,...,n_f \qquad \text{(A3.1)}$$

where $\delta(.)$ is the Dirac delta function which takes the value 1 if the argument is true and zero otherwise. Equation (A3.1) can be expressed as a difference equation but since a product of exponentials can be expressed as the exponential of the sum of arguments, the $N_{y_i,a}$ can be projected forward to any future fishing year in a single step. It can be seen from equation (A3.1) that a decrease in the estimates of both the F's and M can be compensated for by a decrease in the estimate of K_0 or, alternatively, a general increase in the estimate of S_a (via the estimates of the parameters of the selectivity function) and vice versa.

For the CCODE model the numbers remaining in each cohort, c, where $c = 1,...,A + n_f - 1$, in each fishing year, $N_{y_i,c}'$, where these cohorts can logically be present, must be calculated and then converted to the numbers-at-age in each fishing year, $N_{y_i,a}$. Therefore, for fishing year and cohort combinations and corresponding age of cohort, firstly for the year when a given cohort is first fished $N_{y_i,c}'$ is given by

$$N_{y_i,c}' = K_0 \Gamma_{y_c'} \exp\left\{ -M(a-1) \right\} - C_{y_i,a} \left\{ 1 - \exp(-M') \right\} / M'$$
$$i = 1,...,n_f; \ c = i,...,(A+i-1); \ a = (y_i - y_c' + 1); \ y_i \geq y_c' \qquad \text{(A3.2a)}$$

while for fishing year and cohort combinations where the cohort is fished in subsequent years

$$N_{y_i,c}' = N_{y_i-1,c}' \exp(-M) - C_{y_i,a} \left\{ 1 - \exp(-M') \right\} / M'$$
$$i = 2,...,n_f; \ c = i,...,(A+i-1); \ a = y_i - y_c' + i \ ; \ y_i \geq y_c'. \qquad \text{(A3.2b)}$$

Then converting to number of individuals in the population in fishing year y_i and age a gives

$$N_{y_i,a} = N_{y_i,c}'$$
$$i = 1,...,n_f; \ c = i,...,(A+i-1); \ a = y_i - y_c' + i \ ; \ y_i \geq y_c'. \qquad \text{(A3.2c)}$$

SIMULATING THE NUMBER OF RECAPTURES

The random variable with sample realisation $r_{y_v,y_i,a}$ is assumed to be distributed as binomial $B\left(g_{y_v,y_i,a}, P_{y_i,a}\right)$ and correspondingly the random variable with sample realisation $R_{y_i,a}$ is assumed to be distributed as binomial $B\left(G_{y_i,a}, P_{y_i,a}\right)$. This last binomial is not the same as that described earlier in the main text, $B\left(C_{y_i,a}, \tau_{y_i,a}\right)$, as an approximation to the hypergeometric distribution since $B\left(C_{y_i,a}, \tau_{y_i,a}\right)$ cannot be obtained as the sum across release years of the independent binomials given by $B\left(C_{y_i,a}, \tau'_{y_v,y_i,a}\right)$, where $\tau'_{y_v,y_i,a} = g_{y_v,y_i,a} / N_{y_i,a}$. This is because the $\tau'_{y_v,y_i,a}$ take different values across release years and therefore the convolution of the $r_{y_v,y_i,a}$ is not the binomial $B\left(C_{y_i,a}, \tau_{y_i,a}\right)$. These binomials have the same expected value but slightly different variances with $B\left(G_{y_i,a}, P_{y_i,a}\right)$ having a lower variance since, although $\left(1 - \tau_{y_i,a}\right)$ and $\left(1 - P_{y_i,a}\right)$ are both close to 1, the former is closer. In either case these binomials are approximations to the theoretically ideal distribution, have the same expected value, and similar variance.