Appendix A. (Ocean quahog) KLAMZ Assessment Model – Technical Documentation

The KLAMZ assessment model is based on the Deriso-Schnute delay-difference equation (Deriso 1980; Schnute 1985; Quinn and Deriso 1999). The delay-difference equation is a relatively simple and implicitly age structured approach to counting fish in either numerical or biomass units. It gives the same results as explicitly age-structured models (e.g. Leslie matrix model) if fishery selectivity is "knife-edged", if somatic growth follows the von Bertalanffy equation, and if natural mortality is the same for all age groups in each year. Knife-edge selectivity means that all individuals alive in the model during the same year experience the same fishing mortality rate.[d] Natural and fishing mortality rates, growth parameters and recruitment may change from year to year, but delay-difference calculations assume that all individuals share the same mortality and growth parameters within each year. The KLAMZ model includes simple numerical models (e.g. Conser 1995) as special cases because growth can be turned off so that all calculations are in numerical units (see below).

As in many other simple models, the delay difference equation explicitly distinguishes between two age groups. In KLAMZ, the two age groups are called "new" recruits (R_t in biomass or numerical units at the beginning of year t) and "old" recruits (S_t) that together comprise the whole stock (B_t). New recruits are individuals that recruited at the beginning of the current year (at nominal age k).[e] Old recruits are all older individuals in the stock (nominal ages $k+1$ and older, survivors from the previous year). As described above, KLAMZ assumes that new and old recruits are fully vulnerable to the fishery. The most important differences between the delay-difference and other simple models (e.g. Prager 1994; Conser 1995; Jacobson et al. 1994) are that von Bertalanffy growth is used to calculate biomass dynamics and that the delay-difference model captures transient age structure effects due to variation in recruitment, growth and mortality exactly. Transient effects on population dynamics are captured exactly because, as described above, the delay-difference equation is algebraically equivalent to an explicitly age-structured model with von Bertalanffy growth.

The KLAMZ model incorporates a few extensions to Schnute's (1985) revision of Deriso's (1980) original delay difference model. Most of the extensions facilitate tuning to a wider variety of data that anticipated in Schnute (1985). The KLAMZ model is programmed in both Excel and in C++ using AD Model Builder[f] libraries. The AD Model Builder version is faster, more reliable and probably better for producing "official" stock assessment results. The Excel version is slower and implements fewer features, but the Excel version remains useful in developing prototype assessment models, teaching and for checking calculations.

---

[1]In applications, assumptions about knife-edge selectivity can be relaxed by assuming the model tracks "fishable", rather that total, biomass (NEFSC 2000a; 2000b). An analogous approach assigns pseudo-ages based on recruitment to the fishery so that new recruits in the model are all pseudo-age k. The synthetic cohort of fish pseudo-age k may consist of more than one biological cohort. The first pseudo-age (k) can be the predicted age at first, 50% or full recruitment based a von Bertalanffy curve and size composition data (Butler et al. 2002). The "incomplete recruitment" approach (Deriso 1980) calculates recruitment to the model in each year R_t as the weighted sum of contributions from two or more biological cohorts (year-classes) from spawning during successive years (i.e. $R_t = \sum_{a=1}^{k} r_a \Pi_{t-a}$ where k is the age at full recruitment to the fishery, r_a is the contribution of fish age k-a to the fishable stock, and Π_{t-a} is the number or biomass of fish age k-a during year t).

[2]In some applications, and more generally, new recruits might be defined as individuals recruiting at the beginning or at any time during the current time step (e.g. NEFSC 1996).

[3]Otter Research Ltd., Box 2040, Sydney, BC, Canada V8L 3S3 (otter@otter-rsch.com).

The most significant disadvantage in using the KLAMZ model and other delay-difference approaches, beyond the assumption of knife-edge selectivity, is that age and length composition data are not used in tuning. However, one can argue that age composition data are used indirectly to the extent they are used to estimate growth parameters or if survey survival ratios (e.g. based on the Heinke method) are used in tuning (see below).

Population dynamics

The assumed birth date and first day of the year are assumed the same in derivation of the delay-difference equation. It is therefore natural (but not strictly necessary) to tabulate catch and other data using annual accounting periods that start on the assumed biological birthday of cohorts.

Biomass dynamics

As implemented in the KLAMZ model, Schnute's (1985) delay-difference equation is:

$$B_{t+1} = (1 + \rho)\, \tau_t\, B_t - \rho\, \tau_t\, \tau_{t-1}\, B_{t-1} + R_{t+1} - \rho\, \tau_t\, J_t\, R_t$$

where B_t is total biomass of individuals at the beginning of year t; ρ is Ford's growth coefficient (see below); $\tau_t = exp(-Z_t) = exp[-(F_t + M_t)]$ is the fraction of the stock that survived in year t, Z_t, F_t, and M_t are instantaneous rates for total, fishing and natural mortality; and R_t is the biomass of new recruits (at age k) at the beginning of the year. The natural mortality rate M_t may vary over time. Instantaneous mortality rates in KLAMZ model calculations are biomass-weighted averages if von Bertalanffy growth is turned on in the model. However, biomass-weighted mortality estimates in KLAMZ are the same as rates for numerical estimates under the assumption of knife-edge selectivity because all individuals are fully recruited. The growth parameter $J_t = w_{t-1,k-1} / w_{t,k}$ is the ratio of mean weight one year before recruitment (age k-1 in year t-1) and mean weight at recruitment (age k in year t).

It is not necessary to specify body weights at and prior to recruitment in the KLAMZ model (parameters v_{t-1} and V_t in Schnute 1985) because the ratio J_t and recruitment biomass contain the same information. Schnute's (1985) original delay difference equation is:

$$B_{t+1} = (1 + \rho)\, \tau_t\, B_t - \rho\, \tau_t \tau_{t-1}\, B_{t-1} + w_{t+1,k} N_{t+1} - \rho \tau_t w_{t-1,k-1}\, N_t$$

To derive the equation used in KLAMZ, substitute recruitment biomass R_{t+1} for the product $w_{t+1,k} N_{t+1,k}$ and adjusted recruitment biomass $J_t R_t = (w_{t-1,k-1}/w_{t,k})\, w_{t,k}\, N_{t,k} =$
$w_{t-1,k-1} N_t$ in the last term on the right hand side. The advantage in using the alternate parameterization for biomass dynamic calculations in KLAMZ is that recruitment is estimated directly in units of biomass and the number of growth parameters is reduced. The disadvantage is that numbers of recruits are not estimated directly by the model. When required, numerical recruitments must be calculated externally as the ratio of estimated recruitment biomass and the average body weight for new recruits.

Numerical population dynamics

Growth can be turned on off so that abundance, rather than biomass, is tracked in the KLAMZ model. Set $J_t=1$ and $\rho=0$ in the delay difference equation, and use N_t (for numbers) in place of B_t to get:

$$N_{t+1} = \tau_t\, N_t + R_{t+1}$$

Mathematically, the assumption $J_t=1$ means that no growth occurs the assumption $\rho=0$ means that the von Bertalanffy K parameter is infinitely large (Schnute 1985). All tuning and population dynamics calculations in KLAMZ for biomass dynamics are also valid for numerical dynamics.

Growth
As described in Schnute (1985), biomass calculations in the KLAMZ model are based on Schnute and Fournier's (1980) re-parameterization of the von Bertalanffy growth model:

$$w_a = w_{k-1} + (w_k - w_{k-1})(1 + \rho^{1+a-k})/(1-\rho)$$

where $w_k=V$ and $w_{k-1}=v$. Schnute and Fournier's (1980) growth model is the same as the traditional von Bertalanffy growth model {$W_a= W_{max} [1 - exp(-K(a-t_{zero})]$ where W_{max}, K and t_{zero} are parameters}. The two growth models are the same because $W_{max} = (w_k - \rho w_{k-1})/(1-\rho)$, $K = -ln(\rho)$ and $t_{zero} = ln[(w_k - w_{k-1})/(w_k - \rho w_{k-1})] / ln(\rho)$.

In the KLAMZ model, the growth parameters J_t can vary with time but ρ is constant.   Use of time-variable J_t values with ρ is constant is the same as assuming that the von Bertalanffy parameters W_{max} and t_{zero} change over time. Many growth patterns can be mimicked by changing W_{max} and t_{zero} (Overholtz et al., 2003). K is a parameter in the C++ version and, in principal, estimable. However, in most cases it is necessary to use external estimates of growth parameters as constants in KLAMZ.

Instantaneous growth rates
Instantaneous growth rate (IGR) calculations in the KLAMZ model are an extension to the original Deriso-Schnute delay difference model. IGRs are used extensively in KLAMZ for calculating catch biomass and projecting stock biomass forward to the time at which surveys occur. The IGR for new recruits depends only on growth parameters:

$$G_t^{New} = ln\left(\frac{w_{k+1,t+1}}{w_{k,t}}\right) = ln(1 + \rho - \rho J_t)$$

IGR for old recruits is a biomass-weighted average that depends on the current age structure and growth parameters. It can be calculated easily by projecting biomass of old recruits $S_t=B_t-R_t$ (escapement) forward one year with no mortality:

$$S_t^* = (1+\rho)S_t - \rho\tau_{t-1}B_{t-1}$$

where the asterisk (*) means just prior to the start of the subsequent year $t+1$. By definition, the IGR for old recruits in year t is $G_t^{Old} = ln(S_t^*/S_t)$. Dividing by S_t gives:

$$G_t^{Old} = ln\left[(1+\rho) - \rho\tau_{t-1}\frac{B_{t-1}}{S_t}\right]$$

IGR for the entire stock is the biomass weighted average of the IGR values for new and old recruits:

$$G_t = \frac{R_t G_t^{New} + S_t G_t^{Old}}{B_t}$$

All IGR values are zero if growth is turned off.

Recruitment

In the Excel version of the KLAMZ model, annual recruitments are calculated $R_t = e^{\Omega_t}$ where Ω_t is a log transformed annual recruitment parameter, which is estimated in the model. In the C++ version, recruitments are calculated based on log geometric mean recruitment (μ) and a set of annual log scale deviation parameters (ω_t):

$$\Omega_t = \mu + \omega_t$$

The deviations ω_t are constrained to average zero.[g] With the constraint, estimation of μ and the set of ω_t values (1+ n years parameters) is equivalent to estimation of the smaller set (n years) of Ω_t values.

Natural mortality

Natural mortality rates (M_t) are assumed constant in the Excel version of the KLAMZ model. In the C++ version, natural mortality rates may be estimated as a constant value or as a set of values that vary with time. In the model:

$$M_t = me^{\varpi_t}$$

where $m=exp(\pi)$ is the geometric mean natural mortality rate, π is a model parameter that may be estimated (in principal but not in practical terms), and ϖ_t is the log scale year-specific deviation. Deviations may be zero (turned off) so that M_t is constant, may vary in a random fashion due to autocorrelated or independent process errors, or may based on a covariate.[h] Model scenarios with zero recruitment may be initializing the parameter π to a small value (e.g. 10^{-16}) and not estimating it.

Random natural mortality process errors are effects due to predation, disease, parasitism, ocean conditions or other factors that may vary over time but are not included in the model. Calculations are basically the same as for survey process errors (see below).

Natural mortality rate covariate calculations are similar to survey covariate calculations (see below) except that the *user should standardized covariates to average zero over the time period included in the model*:

$$\kappa_t = K_t - \overline{K}$$

where κ_t is the standardized covariate, K_t is the original value, and $\overline{K}$ is the mean of the original covariate for the years in the model. *Standardization to mean zero is important because otherwise m is not the geometric mean natural mortality rate* (the convention is important in some calculations, see text).

Log scale deviations that represent variability around the geometric mean are calculated:

$$\varpi_t = \sum_{j=1}^{n} p_j \kappa_t$$

where n is the number of covariates and p_j is the parameter for covariate j. These conventions mean that the units for the covariate parameter p_j are 1/units of the original covariate, the parameter p_j measures the

---

[g] The constraint is implemented by adding $L = \lambda \overline{\varpi}^2$ (where $\overline{\varpi}$ is the average deviation) to the objective function, generally with a high weighting factor ($\lambda = 1000$) so that the constraint is binding.

[h] Another approach to using time dependent natural mortality rates is to treat estimates of predator consumption as discarded catch (see "Predator consumption as discard data"). In addition, estimates of predator abundance can be used in fishing effort calculations (see "Predator data as fishing effort").

log scale effect of changing the covariate by one unit, and the parameter m is the log scale geometric mean.

Fishing mortality and catch

Fishing mortality rates (F_t) are calculated so that predicted and observed catch data (landings plus estimated discards in units of weight) "agree" to the extent specified by the user. It is not necessary, however, to assume that catches are measured accurately (see "Observed and predicted catch").

Fishing mortality rate calculations in Schnute (1985) are exact but relating fishing mortality to catch in weight is complicated by continuous somatic growth throughout the year as fishing occurs. The KLAMZ model uses a generalized catch equation that incorporates continuous growth through the fishing season. By the definition of instantaneous rates, the catch equation expresses catch as the product:

$$\hat{C}_t = F_t \overline{B}_t$$

where $\hat{C}_t$ is predicted catch weight (landings plus discard) and $\overline{B}_t$ is average biomass.

Following Chapman (1971) and Zhang and Sullivan (1988), let $X_t = G_t - F_t - M_t$ be the net instantaneous rate of change for biomass.[i] If the rates for growth and mortality are equal, then $X_t = 0$, $\overline{B}_t = B_t$ and $C_t = F_t B_t$. If the growth rate G_t exceeds the combined rates of natural and fishing mortality ($F_t + M_t$), then $X_t > 0$. If mortality exceeds growth, then $X_t < 0$. In either case, with $X_t \neq 0$, average biomass is computed:

$$\overline{B}_t \approx -\frac{\left(1 - e^{X_t}\right) B_t}{X_t}$$

When $X_t \neq 0$, the expression for $\overline{B}_t$ is an approximation because G_t approximates the rate of change in mean body weight due to von Bertalanffy growth. However, the approximation is reasonably accurate and preferable to calculating catch biomass in the delay-difference model with the traditional catch equation that ignores growth during the fishing season.[j] Average biomass can be calculated for new recruits, old recruits or for the whole stock by using either G_t^{New}, G_t^{Old} or G_t.

In the KLAMZ model, the modified catch equation may be solved analytically for F_t given C_t, B_t, G_t and M_t (see the "Calculating F_t" section below). Alternatively, fishing mortality rates can be calculated using a log geometric mean parameter (Φ) and a set of annual log scale deviation parameters (ψ_t):

$$F_t = e^{\Phi + \psi_t}$$

where the deviations ψ_t are constrained to average zero. When the catch equation is solved analytically, catches must be assumed known without error but the analytical option is useful when catch is zero or very near zero, or the range of fishing mortality rates is so large (e.g. minimum F=0.000001 to maximum F=3) that numerical problems occur with the alternative approach. The analytical approach is also useful if the user wants to reduce the number of parameters estimated by nonlinear optimization. In any case, the two methods should give the same results for catches known without error.

---

[i] By convention, the instantaneous rates G_t, F_t and M_t are always expressed as numbers ≥ 0.

[j] The traditional catch equation $C_t = F_t(1 - e^{-Z_t}) B_t / Z_t$ where $Z_t = F_t + M_t$ underestimates catch biomass for a given level of fishing mortality F_t and overestimates F_t for a given level of catch biomass. The errors can be substantial for fast growing fish, particularly if recent recruitments were strong.

Surplus production

Annual surplus production is calculated "exactly" by projecting biomass at the beginning of each year forward with no fishing mortality:

$$B_t^* = (1+\rho)\,e^{-M}\,B_t - \rho\,e^{-M}\,L_{t-1}\,B_{t-1} - \rho\,e^{-M}\,J_t\,R_t$$

By definition, surplus production $P_t = B_t^* - B_t$ (Jacobson et al. 2002).

Per recruit modeling

Per recruit model calculations in the Excel version of the KLAMZ simulate the life of a hypothetical cohort of arbitrary size (e.g. $R=1000$) starting at age k with constant M_t, F (survival) and growth (ρ and J) in a population initially at zero biomass. In the first year:

$$B_1 = R$$

In the second year:

$$B_2 = (1+\rho)\,\tau\,B_1 - \rho\,\tau\,J\,R_1$$

In the third and subsequent years:

$$B_{t+1} = (1+\rho)\,\tau\,B_t - \rho\,\tau^2\,B_{t-1}$$

This iterative calculation is carried out until the sum of lifetime cohort biomass from one iteration to the next changes by less than a small amount (0.0001). Total lifetime biomass, spawning biomass and yield in weight are calculated by summing biomass, spawning biomass and yield over the lifetime of the cohort. Lifetime biomass, spawning biomass and yield per recruit are calculated by dividing totals by initial recruitment (R).

Status determination variables

The user may specify a range of years (e.g. the last three years) to use in calculating recent average fishing mortality $\overline{F}_{Recent}$ and biomass $\overline{B}_{Recent}$ levels. These status determination variables are used in calculation of status ratios such as $\overline{F}_{Recent} / F_{MSY}$ and $\overline{B}_{Recent} / B_{MSY}$.

Goodness of Fit and Parameter Estimation

Parameters estimated in the KLAMZ model are chosen to minimize an objective function based on a sum of weighted negative log likelihood (NLL) components:

$$\Xi = \sum_{v=1}^{N_\Xi} \lambda_v L_v$$

where N_Ξ is the number of NLL components (L_v) and the λ_v are emphasis factors used as weights. The objective function Ξ may be viewed as a NLL or a negative log posterior (NLP) distribution, depending on the nature of the individual L_v components and modeling approach. Except during sensitivity analyses, weighting factors for objective function components (λ_v) are usually set to one. An arbitrarily large weighting factor (e.g. $\lambda_v = 1000$) is used for "hard" constraints that must be satisfied in the model. Arbitrarily small weighting factors (e.g. $\lambda_v = 0.0001$) can be used for "soft" model-based constraints. For example, an internally estimated spawner-recruit curve or surplus production curve might be estimated with a small weighting factor to summarize stock-recruit or surplus production results with minimal

influence on biomass, fishing mortality and other estimates from the model. Use of a small weighting factor for an internally estimated surplus production or stock-recruit curve is equivalent to fitting a curve to model estimates of biomass and recruitment or surplus production in the output file, after the model is fit (Jacobson et al. 2002).

Likelihood component weights vs. observation-specific weights

Likelihood component weights (λ_v) apply to entire NLL components. Entire components are often computed as the sum of a number of individual NLL terms. The NLL for an entire survey, for example, is composed of NLL terms for each of the annual survey observations. In KLAMZ, observation-specific (for data) or instance-specific (for constraints or prior information) weights (usually w_j for observation or instance j) can be specified as well. Observation-specific weights for a survey, for example, might be use to increase or decrease the importance of one or more observations in calculating goodness of fit.

NLL kernels

NLL components in KLAMZ are generally programmed as "concentrated likelihoods" to avoid calculation of values that do not affect derivatives of the objective function.[k] For $x \sim N(\mu, \sigma^2)$, the complete NLL for one observation is:

$$L = \ln(\sigma) + \ln(\sqrt{2\pi}) + 0.5\left(\frac{x - u}{\sigma}\right)^2$$

The constant $\ln(\sqrt{2\pi})$ can always be omitted because it does not affect derivatives. If the standard deviation is known or assumed known, then $\ln(\sigma)$ can be omitted as well because it is a constant that does not affect derivatives. In such cases, the concentrated negative log likelihood is:

$$L = 0.5\left(\frac{x - \mu}{\sigma}\right)^2$$

If there are N observations with possible different variances (known or assumed known) and possibly different expected values:

$$L = 0.5\sum_{i=1}^{N}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2$$

If the standard deviation for a normally distributed quantity is not known and is (in effect) estimated by the model, then one of two equivalent calculations is used. Both approaches assume that all observations have the same variance and standard deviation. The first approach is used when all observations have the same weight in the likelihood:

$$L = 0.5N\ln\left[\sum_{i=1}^{N}(x_i - u)^2\right]$$

where N is the number of observations. The second approach is equivalent but used when the weights for each observation (w_i) may differ:

---

[k] Unfortunately, concentrated likelihood calculations cannot be used with MCMC and other Bayesian approaches to characterizing posterior distributions. Therefore, in the near future, concentrated NLL calculations will be replaced by calculations for the entire NLL. At present, MCMC calculations in KLAMZ are not useful.

$$L = \sum_{i=1}^{N} w_i \left[\ln(\sigma) + 0.5 \left(\frac{x_i - u}{\sigma} \right)^2 \right]$$

In the latter case, the maximum likelihood estimator:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \hat{x})^2}{N}}$$

(where $\hat{x}$ is the average or predicted value from the model) is used for σ. The maximum likelihood estimator is biased by $N/(N-d_f)$ where d_f is degrees of freedom for the model. The bias may be significant for small sample sizes but d_f is usually unknown.

Landings, discards, catch
        Discards are from external estimates (d_t) supplied by the user. If $d_t \geq 0$, then the data are used as the ratio of discard to landed catch so that:

$$D_t = L_t \Delta_t$$

where $\Delta_t = D_t/L_t$ is the discard ratio. If $d_t < 0$ then the data are treated as discard in units of weight:

$$D_t = abs(d_t).$$

In either case, total catch is the sum of discards and landed catch ($C_t = L_t + D_t$). It is possible to use discards in weight $d_t < 0$ for some years and discard as proportions $d_t > 0$ for other years in the same model run. If catches are estimated (see below) so that the estimated catch $\hat{C}_t$ does not necessarily equal observed landings plus discard, then estimated landings are computed:

$$\hat{L}_t = \frac{\hat{C}_t}{1 + \Delta_t}$$

and estimated discards are:

$$\hat{D}_t = \Delta_t \hat{L}_t.$$

Calculating F_t
        As described above, fishing mortality rates may be estimated based on the parameters Φ and ψ_t to satisfy a NLL for observed and predicted catches:

$$L = 0.5 \sum_{t=0}^{N} w_t \left(\frac{\hat{C}_t - C_t}{\kappa_t} \right)^2$$

where the standard error $\kappa_t = CV_{catch} \hat{C}_t$ with CV_{catch} and weights are w_t supplied by the user. The weights can be used, for example, if catch data in some years are less precise than in others. Using observation specific weights, any or every catch in the time series can potentially be estimated.

        The other approach to calculating F_t values is by solving the generalized catch equation (see above) iteratively. Subtracting predicted catch from the generalized catch equation gives:

$$g(F_t) = C_t + \frac{F_t (1 - e^{X_t})}{X_t} \overline{B}_t = 0$$

where $X_t = G_t - M_t - F_t$. If $X_t = 0$, then $\overline{B}_t = B_t$ and $F_t = C_t / B_t$.

If $X_t \neq 0$, then the Newton-Raphson algorithm is used to solve for F_t (Kennedy and Gentle 1980). At each iteration of the algorithm, the current estimate F_t^i is updated using:

$$F_t^{i+1} = F_t^i - \frac{g(F_t^i)}{g'(F_t^i)}$$

where $g'(F_t^i)$ is the derivative F_t^i. Omitting subscripts, the derivative is:

$$g'(F) = -\frac{Be^{-F}\left[(e^F - e^\gamma)\gamma + e^\gamma F\gamma - e^\gamma F^2\right]}{X^2}$$

where $\gamma = G - M_t$. Iterations continue until $g(F_t^i)$ and $abs\left[g(F_t^{i+1}) - g(F_t^{i+1})\right]$ are both less than a small number (e.g. ≤ 0.00001).

Initial values are important in algorithms that solve the catch equation numerically (Sims 1982). If $M_t + F_t > G_t$ so that $X_t < 0$, then the initial value F_t^0 is calculated according to Sims (1982). If $M_t + F_t < G_t$ so that $X_t > 0$, then initial values are calculated based on a generalized version of Pope's cohort analysis (Zhang and Sullivan 1988):

$$F_t^0 = \gamma_t - \ln\left[\frac{(B_t e^{0.5\gamma_t} - C_t)e^{0.5\gamma_t}}{B_t}\right]$$

<u>F for landings versus F for discards</u>

The total fishing mortality rate for each year can be partitioned into a component due to landed catch $^LF_t = \frac{D_t}{C_t}F_t$, and a component due to discard $^DF_t = \frac{L_t}{C_t}F_t$.

<u>Predator consumption as discard data</u>

In modeling population dynamics of prey species, estimates of predator consumption can be treated like discard in the KLAMZ model as a means for introducing time dependent natural mortality. Consider a hypothetical example with consumption data (mt y^{-1}) for three important predators. If the aggregate consumption data are included in the model as "discards", then the fishing mortality rate for discards ^dF_t (see above) would be an estimate of the component of natural mortality due to the three predators. In using this approach, the average level of natural mortality m would normally be reduced (e.g. so that $m_{new} + {}^d\overline{F} = m_{old}$) or estimated to account for the portion of natural mortality attributed to bycatch.

Surplus production calculations are harder to interpret if predator consumption is treated as discard data because surplus production calculations assume that $F_t=0$ (see above) and because surplus production is defined as the change in biomass from one year to the next in the absence of fishing (i.e. no landings or bycatch). However, it may be useful to compare surplus production at a given level of biomass from runs with and without consumption data as a means of estimating maximum changes in potential fishery yield if the selected predators were eliminated (assuming no change in disease, growth rates, predation by other predators, etc.).

Effort calculations

Fishing mortality rates can be tuned to fishing effort data for the "landed" catch (i.e. excluding discards). Years with non-zero fishing effort used in the model must also have landings greater than zero. Assuming that effort data are lognormally distributed, the NLL for fishing effort is:

$$NLL = 0.5 n_{eff} \ln \left[\sum_{y=1}^{n_{eff}} w_y \ln \left(\frac{E_y}{\overline{E}_y} \right)^2 \right]$$

where n_{eff} is the number of effort observations, w_y is an observation-specific weight, E_y and $\overline{E}_y$ are observed and predicted fishing effort data, and the log scale variance is estimated internally. Predicted fishing effort data are calculated:

$$\overline{E}_y = \zeta F_y^{\vartheta}$$

where $\zeta = e^u$, $\vartheta = e^b$, and u and b are parameters estimated by the model. If the parameter b is not estimated, then $=1$ so that the relationship between fishing effort and fishing mortality is linear. If the parameter b is estimated, then 1 and the relationship is a power function.

Predator data as fishing effort

As described under "Predator consumption as discard data", predator consumption data can be treated as discard. If predator abundance data are available as well, and assuming that mortality due predators is a linear function of the predator-prey ratio, then both types of data may be used together to estimate natural mortality. The trick is to: 1) enter the predator abundance data as fishing effort; 2) enter the actual fishery landings as "discard"; 3) enter predator consumption estimates of the prey species as "landings" so that the fishing effort data in the refer to the predator consumption data; 4) use an option in the model to calculate the predator-prey ratio for use in place of the original predator abundance "fishing effort" data; and 5) tune fishing mortality rates for landings (a.k.a. predator consumption) to fishing effort (a.k.a. predator-prey ratio).

Given the predator abundance data κ_y, the model calculates the predator-prey ratio used in place of fishing effort (E_y) as:

$$E_y = \frac{\kappa_y}{B_y}$$

where B_y is the model's current estimate of total (a.k.a "prey") biomass. Subsequent calculations with E_y and the model's estimates of "fishing mortality" (F_y, really a measure of natural mortality) are exactly as described above for effort data. In using this approach, it is probably advisable to reduce m (the estimate of average mortality in the model) to account for the proportion of natural mortality due to predators included in the calculation. Based on experience to date, natural mortality due to consumption by the suite of predators can be estimated but only if m is assumed known.

Initial population age structure

In the KLAMZ model, old and new recruit biomass during the first year (R_1 and $S_1 = B_1 - R_1$) and biomass prior to the first year (B_0) are estimated as log scale parameters. Survival in the year prior to the first year ("year 0") is $\tau_0 = e^{-F_0 - M_1}$ with F_0 chosen to obtain catch C_0 (specified as data) from the estimated biomass B_0. IGRs during year 0 and year 1 are assumed equal ($G_0 = G_1$) in catch calculations.

Biomass in the second year of as series of delay-difference calculations depends on biomass (B_0) and survival (τ_0) in year 0:

$$B_2 = (1 + \rho)\, \tau_1\, B_1 - \rho\, \tau_1\, \tau_0\, B_0 + R_2 - \rho\, \tau_1\, J_1\, R_1$$

There is, however, there is no direct linkage between B_0 and escapement biomass ($S_1 = B_1 - R_1$) at the beginning of the first year.

The missing link between B_0, S_1 and B_1 means that the parameter for B_0 tends to be relatively free and unconstrained by the underlying population dynamics model. In some cases, B_0 can be estimated to give good fit to survey and other data, while implying unreasonable initial age composition and surplus production levels. In other cases, B_0 estimates can be unrealistically high or low implying, for example, unreasonably high or low recruitment in the first year of the model (R_1). Problems arise because many different combinations of values for R_1, S_1 and B_0 give similar results in terms of goodness of fit. This issue is common in stock assessment models that use forward simulation calculations because initial age composition is difficult to estimate. It may be exacerbated in delay-difference models because age composition data are not used.

The KLAMZ model uses two constraints to help estimate initial population biomass and initial age structure.[1] The first constraint links IGRs for escapement (G^{Old}) in the first years to a subsequent value. The purpose of the constraint is to ensure consistency in average growth rates (and implicit age structure) during the first few years. For example, if IGRs for the first n_G years are constrained[m], then the NLL for the penalty is:

$$L_G = 0.5 \sum_{t=1}^{n_G} \left[\frac{\ln\left(G_t^{Old} / G_{n_G+1}^{Old}\right)}{\sigma_G} \right]^2$$

where the standard deviation σ_G is supplied by the user. It is usually possible to use the standard deviation of Q_t^{Old} for later years from a preliminary run to estimate σ_G for the first few years. The constraint on initial IGRs should probably be "soft" and non-binding ($\lambda \approx 1$) because there is substantial natural variation in somatic growth rates due to variation in age composition.

The second constraint links B_0 to S_1 and ensures conservation of mass in population dynamics between years 0 and 1. In other words, the parameter for escapement biomass in year 1 is constrained to match an approximate projection of the biomass in year 0, accounting for growth, and natural and fishing mortality. The constraint is intended to be binding and satisfied exactly (e.g. $\lambda = 1000$) because incompatible values of S_1 and B_0 are biologically impossible. In calculations:

$$S_1^p = B_0 e^{G_1 - F_0 - M_1}$$

where S_1^p is the projected escapement in year 1 and B_0 is the model's estimate of total biomass in year 0. The instantaneous rates for growth and natural mortality from year 1 (G_1 and M_1) are used in place of G_0 and M_0 because the latter are unavailable. The NLL for the constraint:

$$L = \left[\ln\left(\frac{S_1^p}{S_1} \right)^2 \right]^2 + \left(S_1^p - S_1 \right)^2$$

---

[1] Quinn and Deriso (1999) describe another approach attributed to a manuscript by C. Walters.
[m] Normally, n_G 2.

uses a log scale sum of squares and an arithmetic sum of squares. The former is effective when S_I is small while the latter is effective when S_I is large. Constants and details in calculation of NLL for the constraint are not important because the constraint is binding (e.g. $\lambda = 1000$).

Equilibrium pristine biomass

It may be useful to constrain the biomass estimate for the first year in a model run towards an estimate of equilibrium pristine biomass if, for example, stock dynamics tend to be stable and catch data are available for the first years of the fishery, or as an alternative to the approach described above for initializing the age structure of the simulated population in the model. Equilibrium pristine biomass $\widetilde{B}_0$ is calculated based on the model's estimate of average recruitment and with no fishing mortality (calculations are similar to those described under "Per-recruit modeling" except that average recruitment is assumed in each year).[n] The NLL term for the constraint is:

$$L = \ln\left(\frac{\widetilde{B}_0}{B_0}\right)^2$$

Pristine equilibrium biomass is used as a hard constraint with a high emphasis factor (λ) so that the variance and constants normally used in NLL calculations are not important.

Estimating natural mortality

As described above, natural mortality calculations involve a parameter for the geometric mean value (m) and time dependent deviations (ϖ_t, which may or may not be turned on). Constraints on natural mortality process errors and natural mortality covariates can be used to help estimate the time dependent deviations and overall trend. The geometric mean natural mortality rate is usually difficult to estimate and best treated as a known constant. However, in the C++ version of the KLAMZ model, $m = e^{\pi}$ (where π is an estimable parameter in the model) and estimates of m can be conditioned on the constraint:

$$L = 0.5\left[\frac{\ln(w/w_{Target})}{\sigma_{\varpi}}\right]^2$$

where w_{Target} is a user supplied mean or target value and σ_{ϖ} is a log scale standard deviation. The standard deviation is calculated from an arithmetic scale CV supplied by the user. Upper and lower bounds for m may be specified as well.

Goodness of fit for trend data

Assuming lognormal errors[o], the NLL used to measure goodness-of-fit to "survey" data that measure trends in abundance or biomass (or survival, see below) is:

---

[n] Future versions of the KLAMZ model will allow equilibrium initial biomass to be calculated based on other recruitment values and for a user-specified level of F (Butler et al. 2003).

[o] Abundance indices with statistical distributions other than log normal may be used as well, but are not currently programmed in the KLAMZ model. For example, Butler et al. (2003) used abundance indices with binomial distributions in a delay-difference model for cowcod rockfish. The next version of KLAMZ will accommodate presence-absence data with binomial distributions.

$$L = 0.5 \sum_{j=1}^{N_v} \left[\frac{\ln\left(I_{v,j} \Big/ \hat{I}_{v,j} \right)}{\sigma_{v,j}} \right]^2$$

where $I_{v,t}$ is an index datum from survey v, hats "^" denote model estimates, $\sigma_{v,j}$ is a log scale standard error (see below), and N_v is the number of observations. There are two approaches to calculating standard errors for log normal abundance index data in KLAMZ and it is possible to use different approaches for different types of abundance index data in the same model (see below).

Standard errors for goodness of fit

In the first approach, all observations for one type of abundance index share the same standard error, which is calculated based on overall goodness of fit. This approach implicitly estimates the standard error based on goodness of fit, along with the rest of the parameters in the model (see "NLL kernels" above).

In the second approach, each observation has a potentially unique standard error that is calculated based on its CV. The second approach calculates log scale standard errors from arithmetic CVs supplied as data by the user (Jacobson et al. 1994):

$$\sigma_{v,t} = \sqrt{\ln\left(1 + CV_{v,t}^2\right)}$$

Arithmetic CV's are usually available for abundance data. It may be convenient to use $CV_{v,t}=1.31$ to get $\sigma_{v,t}=1$.

There are advantages and disadvantages to both approaches. CV's carry information about the relative precision of abundance index observations. However, CV's usually overstate the precision of data as a measure of fish abundance[p] and may be misleading in comparing the precision of one sort of data to another as a measure of trends in abundance (e.g. in contrasting standardized LPUE that measure fishing success, but not abundance, precisely with survey data that measure trends in fish abundance directly, but not precisely). Standard errors estimated implicitly are often larger and more realistic, but assume that all observations in the same survey are equally reliable.

Predicted values for abundance indices
Predicted values for abundance indices are calculated:

$$\hat{I}_{v,t} = Q_v A_{v,t}$$

where Q_v is a survey scaling parameter (constant here but see below) that converts units of biomass to units of the abundance index. $A_{v,t}$ is available biomass at the time of the survey.

In the simplest case, available biomass is:

---

[p] The relationship between data and fish populations is affected by factors (process errors) that are not accounted for in CV calculations.

$$A_{v,t} = s_{v,New} R_t e^{-X_t^{New}\Delta_{v,t}} + s_{v,Old} S_t e^{-X_t^{Old}\Delta_{v,t}}$$

where $s_{v,New}$ and $s_{v,Old}$ are survey selectivity parameters for new recruits (R_t) and old recruits (S_t); $X_t^{New} = G_t^{New} - F_t - M_t$ and $X_t^{Old} = G_t^{Old} - F_t - M_t$; $j_{v,t}$ is the Julian date at the time of the survey, and $\Delta_{v,t} = j_{v,t}/365$ is the fraction of the year elapsed at the time of the survey.

Survey selectivity parameter values ($s_{v,New}$ and $s_{v,Old}$) are specified by the user and must be set between zero and one. For example, a survey for new recruits would have $s_{v,New}=1$ and $s_{v,Old}=0$. A survey that measured abundance of the entire stock would have $s_{v,New}=1$ and $s_{v,Old}=1$.

Terms involving $\Delta_{v,t}$ are used to project beginning of year biomass forward to the time of the survey, making adjustments for mortality and somatic growth.[q] As described below, available biomass $A_{v,t}$ is adjusted further for nonlinear surveys, surveys with covariates and surveys with time variable $Q_{v,t}$.

Scaling parameters (Q) for log normal abundance data
Scaling parameters for surveys with lognormal statistical errors were computed using the maximum likelihood estimator:

$$Q_v = e^{\dfrac{\sum\limits_{i=1}^{N_v}\left[\ln\left(\dfrac{I_{v,i}}{A_{v,i}}\right)\middle/\sigma_{v,j}^2\right]}{\sum\limits_{j=1}^{N_j}\left(1\middle/\sigma_{v,j}^2\right)}}$$

where N_v is the number of observations with individual weights greater than zero. The closed form maximum likelihood estimator gives the same answer as if scaling parameters are estimated as free parameters in the assessment model assuming lognormal survey measurement errors.

Survey covariates
Survey scaling parameters may vary over time based on covariates in the KLAMZ model. The survey scaling parameter that measures the relationship between available biomass and survey data becomes time dependent:

$$\hat{I}_{v,t} = Q_{v,t} A_{v,t}$$

and

$$Q_{v,t} = Q_v e^{\sum\limits_{r=1}^{n_v} d_{r,t}\theta_r}$$

with n_v covariates for the survey and parameters θ_r estimated in the model. Covariate effects and available biomass are multiplied to compute an adjusted available biomass:

---

[q] It may be important to project biomass forward if an absolute estimate of biomass is available (e.g. from a hydroacoustic or daily egg production survey), if fishing mortality rates or high or if the timing of the survey varies considerably from year to year.

$$A'_{v,t} = A_{v,t} e^{\sum_{r=1}^{n_v} d_{r,t}\theta_r}$$

The adjusted available biomass $A'_{v,t}$ is used instead of the original value $A_{v,t}$ in the closed form maximum likelihood estimator described above.

Covariates might include, for example, a dummy variable that represents changes in survey bottom trawl doors or a continuous variable like average temperature data if environmental factors affect distribution and catchability of fish schools. Dummy variables are usually either 0 or 1, depending on whether the effect is present in a particular year. With dummy variables, Q_v is the value of the survey scaling parameter with no intervention ($d_{r,t}=0$).

For ease in interpretation of parameter estimates for continuous covariates (e.g. temperature data), it is useful to center covariate data around the mean:

$$d_{r,t} = d'_{r,t} - \overline{d'_r}$$

where $d'_{r,t}$ is the original covariate. When covariates are continuous and mean-centered, Q_v is the value of the survey scaling parameter under average conditions ($d_{r,t}=0$) and units for the covariate parameter are easy to interpret (for example, units for the parameter are $1/{}^{\circ}C$ if the covariate is mean centered temperature in ${}^{\circ}C$).

It is possible to use a survey covariate to adjust for differences in relative stock size from year to year due to changes in the timing of a survey. However, this adjustment may be made more precisely by letting the model calculate $\Delta_{v,t}$ as described above, based on the actual timing data for the survey during each year.

Nonlinear abundance indices

With nonlinear abundance indices, and following Methot (1990), the survey scaling parameter is a function of available biomass:

$$Q_{v,t} = Q_v A_{v,t}^{\Gamma}$$

so that:

$$\hat{I}_{v,t} = \left(Q_v A_{v,t}^{\Gamma}\right) A_{v,t}$$

Substituting $e^{\gamma}=\Gamma+1$ gives the equivalent expression:

$$\hat{I}_{v,t} = Q_v A_{v,t}^{e^{\gamma}}$$

where γ is a parameter estimated by the model and the survey scaling parameter is no longer time dependent. In calculations with nonlinear abundance indices, the adjusted available biomass:

$$A'_{v,t} = A_{v,t}^{e^{\gamma}}$$

is computed first and used in the closed form maximum likelihood estimator described above to calculate the survey scaling parameter. In cases where survey covariates are also applied to a nonlinear index, the adjustment for nonlinearity is carried out first.

Survey Q process errors

The C++ version of the KLAMZ model can be used to allow survey scaling parameters to change in a controlled fashion from year to year (NEFSC 2002):

$$Q_{v,t} = Q_v e^{\varepsilon_{v,t}}$$

where the deviations $\varepsilon_{v,t}$ are constrained to average zero. Variation in survey Q values is controlled by the NLL penalty:

$$L = 0.5 \sum_{j=1}^{N_v} \left[\frac{\varepsilon_{v,j}}{\sigma_v} \right]^2$$

where the log scale standard deviation σ_v based on an arithmetic CV supplied by the user (e.g. see NEFSC 2002). In practice, the user increases or decreases the amount of variability in Q by decreasing or increasing the assumed CV.

Survival ratios as surveys

In the C++ version of KLAMZ, it is possible to use time series of survival data as "surveys". For example, an index of survival might be calculated using survey data and the Heinke method (Ricker 1975) as:

$$A_t = \frac{I_{k+1,t+1}}{I_{k,t}}$$

so that the time series of A_t estimates are data that may potentially contain information about scale or trends in survival. Predicted values for an a survival index are calculated:

$$\hat{A}_t = e^{-Z_t}$$

After predicted values are calculated, survival ratio data are treated in the same way as abundance data (in particular, measurement errors are assumed to be lognormal). Selectivity parameters are ignored for survival data but all other features (e.g. covariates, nonlinear scaling relationships and constraints on Q) are available.

Recruitment models

Recruitment parameters in KLAMZ may be freely estimated or estimated around an internal recruitment model, possibly involving spawning biomass. An internally estimated recruitment model can be used to reduce variability in recruitment estimates (often necessary if data are limited), to summarize stock-recruit relationships, or to make use of information about recruitment in similar stocks. There are four types of internally estimated recruitment models in KLAMZ: 1) random variation around a constant mean; 2) random walk around a constant mean (autocorrelated variation); 3) random variation around a Beverton-Holt recruitment model; and 4) random variation around a Ricker recruitment model. The user must specify a type of recruitment model but the model is not active unless the likelihood component for the recruitment model is turned on ($\lambda > 0$).

The first step in recruit modeling is to calculate the expected log recruitment level $E[ln(R_t)]$ given the recruitment model. For random variation around a constant mean, the expected log recruitment level is the log geometric mean recruitment:

$$E[\ln(R_t)] = \sum_{j=1}^{N} \ln(R_j) \Big/ N$$

For a random walk around a constant mean recruitment, the expected log recruitment level is the logarithm of recruitment during the previous year:

$$E[\ln(R_t)] = \ln(R_{t-1})$$

with no constraint on recruitment during the first year R_1.

For the Beverton-Holt recruitment model, the expected log recruitment level is:

$$E[\ln(R_t)] = \ln[e^a T_{t-\ell} / (e^b + T_{t-\ell})]$$

where $a=e^\alpha$ and $b=e^\beta$, the parameters α and β are estimated in the model, T_t is spawning biomass, and   is the lag between spawning and recruitment. Spawner-recruit parameters are estimated as log transformed values (e^α and e^β) to enhance model stability and ensure the correct sign of values used in calculations. Spawning biomass is:

$$T_t = m_{new} R_t + m_{old} S_t$$

where m_{new} and m_{old} are maturity parameters for new and old recruits specified by the user. For the Ricker recruitment model, the expected log recruitment level is:

$$E[\ln(R_t)] = \ln(S_{t-\ell} e^{a - bS_{t-\ell}})$$

where $a=e^\alpha$ and $b=e^\beta$, and the parameters α and β are estimated in the model.

Given the expected log recruitment level, log scale residuals for the recruitment model are calculated:

$$r_t = \ln(R_t) - E[\ln(R_t)]$$

Assuming that residuals are log normal, the NLL for recruitment residuals is:

$$L = \sum_{t=t_{first}}^{N} w_t \left[\ln(\sigma_r) + 0.5 \left(\frac{r_t}{\sigma_r} \right)^2 \right]$$

where λ_t is an instance-specific weight usually set equal one. The additional term in the NLL $[\ln(\sigma_r)]$ is necessary because the variance σ_r^2 is estimated internally, rather than specified by the user.

The log scale variance for residuals is calculated using the maximum likelihood estimator:

$$\sigma_r^2 = \frac{\sum_{j=t_{first}}^{N} r_j}{N}$$

where N is the number of residuals. For the recruitment model with constant variation around a mean value, $t_{first}=1$. For the random walk recruitment model, $t_{first}=2$. For the Beverton-Holt and Ricker models, $t_{first}=\ell+1$ and the recruit model imposes no constraint on variability of recruitment during years 1 to ℓ (see below). The biased maximum likelihood estimate for σ^2 (with N in the divisor instead of the degrees of freedom) is used because actual degrees of freedom are unknown. The variance term σ^2 is calculated explicitly and stored because it is used below.

Constraining the first few recruitments

It may be useful to constrain the first   years of recruitments when using either the Beverton-Holt or Ricker models if the unconstrained estimates for early years are erratic. In the KLAMZ model, this constraint is calculated:

$$NLL = \sum_{t=1}^{t_{first}-1} w_t \left\{ \ln(\sigma_r + 0.5 \left[\frac{\ln(R_t / E(R_{t_{first}}))}{\sigma_r} \right]^2 \right\}$$

where t_{first} is the first year for which expected recruitment $E(R_l)$ can be calculated with the spawner-recruit model. In effect, recruitments that not included in spawner-recruit calculations are constrained towards the first spawner-recruit prediction. The standard deviation is the same as used in calculating the NLL for the recruitment model.

Prior information about abundance index scaling parameters (Q)

A constraint on one or more scaling parameters (Q_v) for abundance or survival indices may be useful if prior information is available (e.g. NEFSC 2000; NEFSC 2001; NEFSC 2002). In the Excel version, it is easy to program these (and other) constraints in an *ad-hoc* fashion as they are needed. In the AD Model Builder version, log normal and beta distributions are preprogrammed for use in specifying prior information about Q_v for any abundance or survival index.

The user must specify which surveys have prior distributions, minimum and maximum legal bounds (q_{min} and q_{max}), the arithmetic mean $(\bar{q})$ and the arithmetic CV for the prior the distribution. Goodness of fit for Q_v values outside the bounds (q_{min}, q_{max}) are calculated:

$$L = \begin{vmatrix} 10000\left(Q_v - q_{max}\right)^2 & if \ Q_v \geq q_{max} \\ 10000\left(q_{min} - Q_v\right)^2 & if \ Q_v \leq q_{min} \end{vmatrix}$$

Goodness of fit for Q_v values inside the legal bounds depend on whether the distribution of potential values is log normal or follows a beta distribution.

Lognormal case

Goodness of fit for lognormal Q_v values within legal bounds is:

$$L = 0.5\left[\frac{\ln(Q_v) - \tau}{\varphi}\right]^2$$

where the log scale standard deviation $\varphi = \sqrt{\ln(1 + CV)}$ and $\tau = \ln(\bar{q}) - \dfrac{\varphi^2}{2}$ is the mean of the corresponding log normal distribution.

Beta distribution case

The first step in calculation goodness of fit for Q_v values with beta distributions is to calculate the mean and variance of the corresponding "standardized" beta distribution:

$$\bar{q}' = \frac{\bar{q} - q_{min}}{D}$$

and

$$Var(q') = \left(\frac{\bar{q} \ CV}{D}\right)^2$$

where the range of the standardized beta distribution is $D = q_{max} - q_{min}$. Equating the mean and variance to the estimators for the mean and variance for the standardized beta distribution (the "method of moments") gives the simultaneous equations:

$$\bar{q}' = \frac{a}{a + b}$$

and

$$Var(q') = \frac{ab}{(a+b)^2(a+b+1)}$$

where a and b are parameters of the standardized beta distribution.[r] Solving the simultaneous equations gives:

$$b = \frac{(\bar{q}'-1)[Var(q')+(\bar{q}'-1)\bar{q}']}{Var(q')}$$

and:

$$a = \frac{b\bar{q}'}{1-\bar{q}'}$$

Goodness of fit for beta Q_v values within legal bounds is calculated with the NLL:

$$L = (a-1)\ln(Q'_v) + (b-1)\ln(1-Q'_v)$$

where $Q'_v = Q_v/(Q_v - q_{min})$ is the standardized value of the survey scaling parameter Q_v.

Surplus production modeling

Surplus production models can be fit internally to biomass and surplus production estimates in the model (Jacobson et al. 2002). Models fit internally can be used to constrain estimates of biomass and recruitment, to summarize results in terms of surplus production, or as a source of information in tuning the model. The NLL for goodness of fit assumes normally distributed process errors in the surplus production process:

$$L = 0.5 \sum_{j=1}^{N_P} \left(\frac{\tilde{P}_j - P_j}{\sigma} \right)^2$$

where N_p is the number of surplus production estimates (number of years less one), $\tilde{P}_t$ is a predicted value from the surplus production curve, P_t is the assessment model estimate, and the standard deviation σ is supplied by the user based, for example, on preliminary variances for surplus production estimates.[s] Either the symmetrical Schaefer (1957) or asymmetric Fox (1970) surplus production curve may be used to calculate $\tilde{P}_t$ (Quinn and Deriso 1999).

It may be important to use a surplus production curve that is compatible with recruitment patterns or assumptions about the underlying spawner-recruit relationship. More research is required, but the asymmetric shape of the Fox surplus production curve appears reasonably compatible with the assumption that recruitment follows a Beverton-Holt spawner-recruit curve (Mohn and Black 1998). In contrast, the symmetric Schaefer surplus production model appears reasonably compatible with the assumption that recruitment follows a Ricker spawner-recruit curve.

---

[r] If x has a standardized beta distribution with parameters a and b, then the probability of x is $P(x) = \frac{x^{a-1}(1-x)^{b-1}}{\Gamma(a,b)}$.

[s] Variances in NLL for surplus production-biomass models are a subject of ongoing research. The advantage in assuming normal errors is that negative production values (which occur in many stocks, e.g. Jacobson et al. 2001) are accommodated. In addition, production models can be fit easily by linear regression of P_t on B_t and B_t^2 with no intercept term. However, variance of production estimate residuals increases with predicted surplus production. Therefore, the current approach to fitting production curves in KLAMZ is not completely satisfactory.

The Schaefer model has two log transformed parameters that are estimated in KLAMZ:

$$\widetilde{P}_t = e^{\alpha} B_t - e^{\beta} B_t^2$$

The Fox model also has two log transformed parameters:

$$\widetilde{P}_t = -e\left(e^{e^{\alpha}}\right)\frac{B_t}{e^{\beta}}\log\left(\frac{B_t}{e^{\beta}}\right)$$

See Quinn and Deriso (1999) for formulas used to calculate reference points (F_{MSY}, B_{MSY}, MSY, and K) for both surplus production models.

Catch/biomass

Forward simulation models like KLAMZ may tend to estimate absurdly high fishing mortality rates, particularly if data are limited. The likelihood constraint used to prevent this potential problem is:

$$L = 0.5\sum_{t=0}^{N} d_t^2$$

where:

$$d_t = \left| \begin{array}{c} (C_t/B - \kappa)\,if\ \ C_t/B > \kappa \\ 0\ otherwise \end{array} \right.$$

with the threshold value κ normally set by the user to about 0.95. Values for κ can be linked to maximum F values using the modified catch equation described above. For example, to use a maximum fishing mortality rate of about F 4 with M=0.2 and G=0.1 (maximum X=4+0.2-0.1=4.1), set $\kappa \approx F/X(1-e^{-X})$=4 / 4.1 (1-e^{-4})=0.96.

Uncertainty

The AD Model Builder version of the KLAMZ model automatically calculates variances for parameters and quantities of interest (e.g. R_t, F_t, B_t, F_{MSY}, B_{MSY}, $\overline{F}_{Recent}$, $\overline{B}_{Recent}$, $\overline{F}_{Recent}/F_{MSY}$, $\overline{B}_{Recent}/B_{MSY}$, etc.) by the delta method using exact derivatives. If the objective function is the log of a proper posterior distribution, then Markov Chain Monte Carlo (MCMC) techniques implemented in AD Model Builder libraries can be used estimate posterior distributions representing uncertainty in the same parameters and quantities.[t]

---

[t] MCMC calculations are not available in the current version because objective function calculations use concentrated likelihood formulas. However, the C++ version of KLAMZ is programmed in other respects to accommodate Bayesian estimation.

Bootstrapping

A FORTRAN program called BootADM can be used to bootstrap survey and survival index data in the KLAMZ model. Based on output files from a "basecase" model run, BootADM extracts standardized residuals:

$$r_{v,j} = \frac{\ln\left(I_{v,j} \middle/ \hat{I}_{v,j}\right)}{\sigma_{v,j}}$$

along with log scale standard deviations ($\sigma_{v,j}$, originally from survey CV's or estimated from goodness of fit), and predicted values $\left(\hat{I}_{v,j}\right)$ for all active abundance and survival observations. The original standardized residuals are pooled and then resampled (with replacement) to form new sets of bootstrapped survey "data":

$$^{x}I_{v,j} = \hat{I}_{v,j}e^{r\sigma_{v,j}}$$

where r is a resampled residual. Residuals for abundance and survival data are combined in bootstrap calculations. BootADM builds new KLAMZ data files and runs the KLAMZ model repetitively, collecting the bootstrapped parameter and other estimates at each iteration and writing them to a comma separated text file that can be processed in Excel to calculate bootstrap variances, confidence intervals, bias estimates, etc. for all parameters and quantities of interest (Efron 1982).

Projections

Stochastic projections can be carried out using another FORTRAN program called SPROJDDF based on bootstrap output from BootADM. Basically, bootstrap estimates of biomass, recruitment, spawning biomass, natural and fishing mortality during the terminal years are used with recruit model parameters from each bootstrap run to start and carryout projections.[u] Given a user-specified level of catch or fishing mortality, the delay-difference equation is used to project stock status for a user-specified number of years. Recruitment during each projected year is based on simulated spawning biomass, log normal random numbers, and spawner-recruit parameters (including the residual variance) estimated in the bootstrap run. This approach is similar to carrying out projections based on parameters and state variables sampled from a posterior distribution for the basecase model fit. It differs from most current approaches because the spawner-recruit parameters vary from projection to projection.

---

[u] At present, only Beverton-Holt recruitment calculations are available in SPROJDDF.

REFERENCES

Butler, J.L., L.D. Jacobson, J.T. Barnes, and H.G. Moser. 2003. Biology and population dynamics of cowcod (*Sebastes levis*) in the southern California Bight. Fish. Bull. 101: 260-280.

Chapman, D.W. 1971. Production. *In*: W.E. Ricker (ed.). Methods for assessment of fish production in fresh waters. IBP Handbook No. 3. Blackwell Scientific Publications, Oxford, UK.

Conser, R.J. 1995. A modified DeLury modeling framework for data limited assessment: bridging the gap between surplus production models and age-structured models. A working paper for the ICES Working Group on Methods of Fish Stock Assessment. Copenhagen, Denmark. Feb. 1995.

Deriso, R, B. 1980. Harvesting strategies and parameter estimation for an age-structured model. Can. J. Fish. Aquat. Sci., 37: 268-282.

Efron, B. 1982. The jackknife, the bootstrap and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia, PA, 92 p.

Fox, W. W. R. Jr. 1970. An exponential yield model for optimizing exploited fish populations. Trans. Am. Fish. Soc. 99: 80-88.

Jacobson, L.D., N.C.H. Lo, and J.T. Barnes. 1994. A biomass based assessment model for northern anchovy *Engraulis mordax*. Fish. Bull. U.S. 92: 711-724.

Jacobson, L.D., J.A.A. De Oliveira, M. Barange, M.A. Cisneros-Mata, R. Felix-Uraga, J.R. Hunter, J.Y. Kim, Y. Matsuura, M. Niquen, C. Porteiro, B. Rothschild, R.P. Sanchez, R. Serra, A. Uriarte, and T. Wada. 2001. Surplus production, variability and climate change in the great sardine and anchovy fisheries. Can. J. Fish. Aquat. Sci. 58: 1891-1903.

Jacobson, L.D., S.X. Cadrin, and J.R. Weinberg. 2002. Tools for estimating surplus production and F_{MSY} in any stock assessment model. N. Am. J. Fish. Mgmt. 22:326–338

Kennedy, W.J. Jr., and J.E. Gentle. 1980. Statistical computing. Marcel Dekker, Inc., New York.

Methot, R.D. 1990. Synthesis model: an adaptable framework for analysis of diverse stock assessment data. Int. N. Pac. Fish. Comm. Bull. 50: 259-277.

Mohn, R., and G. Black. 1998. Illustrations of the precautionary approach using 4TVW haddock, 4VsW cod and 3LNO plaice. Northwest Atlantic Fisheries Organization NAFO SCR Doc. 98/10.

NEFSC. 1996. American lobster. *In:* Report of the 22nd Northeast Regional Stock Assessment Workshop (22nd SAW): Stock Assessment Review Committee (SARC) consensus summary of assessments. Northeast Fish. Sci. Cent. Ref. Doc. 96-13.

NEFSC. 2000. Ocean quahog. *In:* Report of the 31st Northeast Regional Stock Assessment Workshop (31st SAW): Stock Assessment Review Committee (SARC) consensus summary of assessments. Northeast Fish. Sci. Cent. Ref. Doc. 00-15.

NEFSC. 2001. Sea scallops. In: Report of the 32st Northeast Regional Stock Assessment Workshop (32st SAW): Stock Assessment Review Committee (SARC) consensus summary of assessments. Northeast Fish. Sci. Cent. Ref. Doc. 01-18.

NEFSC. 2002. Longfin squid. *In*: Report of the 34th Northeast Regional Stock Assessment Workshop (34th SAW): Stock Assessment Review Committee (SARC) consensus summary of assessments. Northeast Fish. Sci. Cent. Ref. Doc. 02-07.

Overholtz, W.J., L. Jacobson, G.D. Melvin, M. Cieri, M. Power, and K. Clark. 2003. Stock assessment of the Gulf of Maine-Georges Bank herring complex, 1959-2002. Northeast Fish. Sci. Cent. Ref. Doc. 03- xx.

Prager, M.H. 1994. A suite of extensions to a nonequilibrium surplus-production model. Fish. Bull. 92: 374-389.

Quinn II, T.J., and R.B. Deriso. 1999. Quantitative fish dynamics. Oxford University Press, NY.

Ricker, W. E. 1975. Computation and interpretation of biological statistics of fish populations. Bull. Fish. Res. Bd. Can. No. 191. 382p.

Schaefer, M.B. 1957. A study of the dynamics of the fishery for yellowfin tuna in the eastern tropical Pacific Ocean. Inter-Am. Trop. Tuna Comm. Bull. 2: 245-285.

Schnute, J. 1985. A general theory for analysis of catch and effort data. Can. J. Fish. Aquat. Sci. 42: 414-429.

Schnute, and D. Fournier. 1980. A new approach to length-frequency analysis: growth structure. Can. J. Fish. Aquat. Sci., 37: 1337-1351.

Sims, S.E. 1982. Algorithms for solving the catch equation forward and backward in time. Can. J. Fish. Aquat. Sci. 39: 197-202.

Zhang, C. I., and P. J. Sullivan. 1988. Biomass-based cohort analysis that incorporates growth. Trans. Am. Fish. Soc. 117: 180-189.