Estimating Growth and Mortality in Steady-State Fish Stocks from Length-Frequency Data

J.A. WETHERALL
J.J. POLOVINA
S. RALSTON

Southwest Fisheries Center, Honolulu Laboratory
National Marine Fisheries Service, NOAA
2570 Dole St., Honolulu, Hawaii 96822-2396
USA

Wetherall, J.A., J.J. Polovina and S. Ralston. 1987. Estimating growth and mortality in steady-state fish stocks from length-frequency data, p. 53-74. *In* D. Pauly and G.R. Morgan (eds.) Length-based methods in fisheries research. ICLARM Conference Proceedings 13, 468 p. International Center for Living Aquatic Resources Management, Manila, Philippines, and Kuwait Institute for Scientific Research, Safat, Kuwait.

Abstract

Methods of using length-frequency statistics to estimate L_∞ and the ratio Z/K in steady-state fish stocks with von Bertalanffy growth and exponential mortality are studied. Several standard procedures applicable when mortality is constant are reviewed, and new methods are introduced which have superior performance. The various methods are evaluated using Monte Carlo techniques. Problems of systematic bias are discussed, and remedial measures are suggested. Finally, a method is developed (but not yet evaluated) to estimate length-specific mortality rates under the steady-state model.

Introduction

The current revival of interest in length-frequency based stock assessment methods has led to several new techniques which provide fisheries biologists with alternatives to well-established procedures or allow analysis in situations where traditional age-based methods are infeasible. Because age estimation is often difficult and usually expensive, length-frequency based methods may be the most efficient and reliable means for estimating some of the key parameters of fishery models (see Mathews, Part I, this vol.).

Most of the new length-based procedures, like the graphical and manual methods they are designed to replace, are applicable to situations where spawning is periodic and the identity of cohorts is adequately maintained in the population's length-frequency distribution. Under such circumstances, it is possible to estimate jointly parameters of growth, mortality and recruitment by fitting composite models to single length-frequency samples (e.g., Schnute and Fournier 1980), or to sequences of samples taken over time (e.g., Pauly and David 1981). Estimation and hypothesis testing are facilitated by adopting specific structural assumptions on the underlying processes and then maximizing likelihood functions or fitting model expectations to observed length class frequencies by least squares or other criteria. Solutions are found by iterative search techniques which take advantage of, and indeed require, the numerical power of computers.

Less attention has been paid lately to the relatively simple methods applicable in situations where cohorts are not clearly distinguishable in catch length distributions, individual length-frequency samples are too small to be treated separately, first approximations are adequate, or sophisticated computer systems and optimization software are unavailable. Among these simple methods, probably the best known is the formula due to Beverton and Holt (1956):

$$Z = K\left(\frac{L_\infty - \bar{\ell}}{\bar{\ell} - L_c}\right) \qquad \ldots 1)$$

which estimates the total instantaneous mortality coefficient, Z, in a steady-state population with constant exponential mortality and von Bertalanffy growth, from the mean length, $\bar{\ell}$, in a random sample of fish above length L_c. In using (1), it is assumed that L_c and the von Bertalanffy parameters, K and L_∞, are given. If only L_c and L_∞ are known (or estimated from other information), slight rearrangement of (1) yields an estimate of the ratio $\theta = Z/K$, a component of many standard yield models and stock assessment procedures (see Gulland 1983).

The Beverton-Holt formula represents a particular class of methods for computing analytical estimates of growth and mortality parameters on the basis of length-frequency data, procedures requiring rather strong assumptions but whose use is justified by their simplicity and their robustness under variable recruitment. This paper reviews three such methods reported in the literature, and introduces some new ones which require more computation but the same amount, or less, input information. First, the assumptions underlying the steady-state methods are given and a structure for length-frequency samples is described. Next, estimators based on the assumption of constant mortality rate are discussed, followed by a consideration of systematic biases. Next, the assumption of constant mortality is relaxed and methods for estimating length-specific mortality are developed. Finally, some Monte Carlo experiments on the constant-mortality methods are described; the concluding section offers advice on use of the various procedures.

Basic Assumptions

PROBABILITY DENSITY OF FISH LENGTHS

In all the methods treated here, a steady-state population is assumed. Growth is assumed to follow the deterministic von Bertalanffy curve with parameters K and L_∞. The curve's location parameter, t_0, usually included in modelling length at age, is omitted since it does not influence length-frequency distributions. Unless otherwise stated, mortality of fish above a knife-edge selection size, L_c, is assumed to occur at a constant instantaneous rate, Z, and recruitment to this size to take place at a constant instantaneous rate, R.

Under these conditions, the probability density $(g(\ell))$ of fish length in the sampled stock is

$$g(\ell) = \frac{\theta\,(L_\infty - \ell)^{\theta - 1}}{(L_\infty - L_c)^{\theta}} \qquad L_c \leqslant \ell < L_\infty \cdot \qquad \ldots 2)$$

On integrating (2), the cumulative length-distribution function ($G(\ell)$) is seen to be

$$G(\ell) = 1 - \left(\frac{L_\infty - \ell}{L_\infty - L_c}\right)^\theta \qquad \ldots 3)$$

$$L_c \leqslant \ell < L_\infty$$

and the mean and variance of length for fish above length L_c are

$$E(\ell) = L_\infty - \left(\frac{\theta}{\theta + 1}\right)(L_\infty - L_c) \qquad \ldots 4)$$

and

$$V(\ell) = \left(\frac{\theta}{\theta + 2}\right)\left(\frac{L_\infty - L_c}{\theta + 1}\right)^2 \qquad \ldots 5)$$

A convenient ancillary variable may be defined as $X = \ell/L_\infty$, the actual length of a fish in proportion to its maximum possible length. By changing variables in (2) it is seen that the density for X is

$$h(X) = \frac{\theta\,(1 - X)^{\theta - 1}}{(1 - X_c)^\theta} \qquad \ldots 6)$$

$$X_c \leqslant X < 1$$

where $X_c = L_c/L_\infty$, and its distribution function is

$$H(X) = 1 - \left(\frac{1 - X}{1 - X_c}\right)^\theta \qquad \ldots 7)$$

SAMPLING SCHEME

It is assumed that a random sample of n fish is taken from $g(\ell)$ above the minimum catchable size, L_c, with lengths $\ell_1, \ell_2, \ldots, \ell_n$. We assume the lengths are observed without error. In practice this assumption is usually violated, but the consequences may not be serious unless measurement errors are very large. We also assume, more critically, that there is no systematic measurement bias.

Measurement error aside, observed fish lengths are recorded with various degrees of "accuracy." As a result, sample length-frequency distributions range from the very sparse (multiple observations of a given length being uncommon in samples of moderate size) to the highly aggregated (the data being grouped into relatively few length intervals, each containing relatively many observations). To describe each data set we adopt a general scheme wherein the $n.$ sample observations are grouped into $r > 2$ contiguous length intervals, each fish in a particular interval being assigned a length equal to the interval midpoint. Let ℓ_i denote the midpoint of the i-th interval of width Δ_i ($i = 1, 2, \ldots, r$). We define L_{min} as the lower bound of the first length interval and L_{max} as the upper bound of the last interval, i.e., $L_{min} = \ell_1 - \Delta_1/2$, and $L_{max} = \ell_r + \Delta_r/2$. A data set consists of a set of length interval midpoints $\{\ell_1, \ell_2, \ldots, \ell_r\}$ and a corresponding set of frequencies $\{n_1, n_2, \ldots, n_r\}$. (Note that the interval frequencies sum to the total sample size, $n.$.)

Estimators

The procedures used to estimate growth and mortality parameters from length-frequency samples differ in their statistical properties, the assumptions they require, and the amount of information they produce. In this section, assuming constant mortality, we review three established methods and explore several new approaches. We first examine some methods applicable when both L_c and L_∞ are known and only θ is to be estimated. Then we consider the joint estimation of θ and L_∞ when only L_c is given.

BEVERTON-HOLT METHOD

The classical formula for $\hat{\theta}$, given by rearranging (1), is

$$\hat{\theta}_{BH} = \frac{L_\infty - \bar{\ell}}{\bar{\ell} - L_c} \qquad \ldots 8)$$

It was derived in a non-probabilistic way by Beverton and Holt (1956), from an expression for mean length in the catch. However, it is readily seen that $\hat{\theta}_{BH}$ is identical to the moment estimator based formally on the density at (2) and the resulting expected length (4). Further analysis with Taylor series shows that $\hat{\theta}_{BH}$ has a statistical bias approximately equal to

$$\text{Bias}(\hat{\theta}_{BH}) = \frac{\theta(\theta+1)}{n.\,(\theta+2)} \qquad \ldots 9)$$

and a large-sample variance of

$$\text{Var}(\hat{\theta}_{BH}) = \frac{\theta(\theta+1)^2}{n.\,(\theta+2)} \qquad \ldots 10)$$

In constructing $\hat{\theta}_{BH}$, L_∞ and L_c were assumed known. However, this is never true, so in applying $\hat{\theta}_{BH}$ independent estimates of these parameters, or simply educated guesses, must be substituted. The statistical properties of $\hat{\theta}_{BH}$ are consequently altered, so that (9) and (10) no longer apply. This problem of systematic bias will be discussed in a later section.

SSENTONGO-LARKIN METHOD

Ssentongo and Larkin (1973) also assumed L_c and L_∞ were known. They developed an estimator for θ by first deriving a moment estimator for Z, based on the mean age of fish in the sample, and then, given the assumed relationship between length and age, changing variables. The resulting expression, assuming fish lengths are measured and recorded exactly, is

$$\hat{\theta}_{SL} = (\bar{y} - y_c)^{-1} \qquad \ldots 11)$$

where

$$\bar{y} = -\frac{1}{n.} \sum_{i=1}^{n.} \ln\left(1 - \frac{\ell_i}{L_\infty}\right) = -\frac{1}{n.} \sum_{i=1}^{n.} \ln(1 - X_i)$$

and

$$y_c = -\ln(1 - X_c)$$

It may readily be shown that $\hat{\theta}_{SL}$ is also the maximum likelihood estimator based on the density at (6), and is a special case of a more general maximum likelihood procedure discussed later.

As Ssentongo and Larkin report, $\hat{\theta}_{SL}$ has a statistical bias of

$$\text{Bias}(\hat{\theta}_{SL}) = \frac{\theta}{n.} \qquad \ldots 12)$$

and a large-sample variance of

$$\text{Var}(\hat{\theta}_{SL}) = \frac{\theta^2}{n.} \qquad \ldots 13)$$

Thus the bias of $\hat{\theta}_{SL}$ is greater than that of $\hat{\theta}_{BH}$, but its asymptotic variance is smaller, as expected.

As in the Beverton-Holt estimator, the properties of θ_{SL} are altered in practice, since L_∞ and L_c must invariably be estimated. Note further that the estimate of L_∞ must exceed L_{max}; we return to this point later.

POWELL METHOD

Powell (1979) considered the problem of estimating both θ and L_∞ in a more general context, in which the asymptotic lengths of fish in a population are regarded as random variables, λ, with expectation L_∞ and variance σ^2_λ. Of various estimation schemes suggested by Powell, the most interesting is based on use of the first and second moments of the resulting probability distribution of ℓ. Manipulation of Powell's results (Equations (4) and (6) in his paper) reveals that

$$E(\ell) = L_\infty - \left(\frac{\theta}{\theta + 1}\right)(L_\infty - L_c) \qquad \ldots 14)$$

and

$$V(\ell) = \left(\frac{1}{\theta + 2}\right)\left\{\theta\left(\frac{L_\infty - L_c}{\theta + 1}\right)^2 + 2\sigma_\lambda^2\right\} \cdot \qquad \ldots 15)$$

We suppose L_c is given. Then, provided σ_λ^2 is known, or the range of λ is negligible compared with the expected maximum range of ℓ in the data (in which case set $\sigma_\lambda^2 = 0$), $E(\ell)$ and $V(\ell)$ can be equated to the corresponding sample statistics and solved for θ and L_∞. Let $\bar{\ell}$ denote the sample mean length for fish above L_c, and S_ℓ^2 the sample variance. Then the resulting moment estimators of θ and L_∞ (when $\sigma_\lambda^2 = 0$) are

$$\hat{\theta}_P = \frac{2S_\ell^2}{(\bar{\ell} - L_c) - S_\ell^2} \qquad \ldots 16)$$

and

$$\hat{L}_{\infty_P} = \bar{\ell} + \frac{2S_\ell^2(\bar{\ell} - L_c)}{(\bar{\ell} - L_c)^2 - S_\ell^2} \qquad \ldots 17)$$

$$= \bar{\ell} + \hat{\theta}_P(\bar{\ell} - L_c) \quad .$$

Since Powell's method makes use of additional information in the sample to estimate L_∞, $\hat{\theta}_P$ is a considerable improvement over $\hat{\theta}_{BH}$ and $\hat{\theta}_{SL}$ when accurate, independent information on L_∞ is unavailable.

REGRESSION METHOD

Another method for estimating θ and L_∞ jointly may be developed by considering the equation at (4), and exploiting the fact that $E(\ell)$ is a linear function of L_c. Let $\{w_1, w_2, \ldots, w_m\}$ be an increasing sequence of fish lengths in the interval (L_c, L_∞). For a random sample of lengths in this interval, let $\overline{\ell}_j$ denote the average length of the n_j fish in the sample whose length exceeds w_j. It follows from (4) that

$$E(\overline{\ell}_j) = L_\infty \left(\frac{1}{1+\theta}\right) + w_j \left(\frac{\theta}{1+\theta}\right) \qquad \ldots 18)$$

$$j = 1, 2, \ldots, m$$

i.e., $E(\overline{\ell}_j)$ is linear in w_j. The idea of the method is to partition the length-frequency sample using a specified sequence $\{w\}$ and estimate α and β in the linear regression model

$$\overline{\ell}_j = \alpha + \beta w_j + \varepsilon_j \qquad \ldots 19)$$

where the ε_j are random errors with zero mean and covariance matrix A. Minimum variance unbiased estimates of α and β are found by weighted least squares, using the weight matrix A^{-1}. An estimate of A may be computed from the sample statistics, the (i, j)-th element being

$$A_{ij} = \mathrm{Cov}(\overline{\ell}_i, \overline{\ell}_j) = \sigma_j^2/n_i \qquad \ldots 20)$$

where σ_j^2 is the variance among the n_j lengths in the smaller, "included" sample, i.e., the variance among the common elements in the i-th and j-th partitions.

Consistent estimates of θ and L_∞ are then provided by

$$\hat{\theta}_R = \frac{\beta}{1-\hat{\beta}} \qquad \ldots 21)$$

and

$$\hat{L}_{\infty R} = \frac{\hat{\alpha}}{1-\hat{\beta}} \qquad \ldots 22)$$

Although the choice of divisions in the length-frequency distribution is somewhat arbitrary, there is undoubtedly some optimal partitioning. To a degree, precision in parameter estimates is increased by creating more partitions (data points). One possibility is to let the observed length interval midpoints define a set of partitions, resulting in r data points for the regression, namely,

$$\omega_j = \tilde{\ell}_j - \frac{\Delta_j}{2} \qquad \ldots 23)$$

and

$$\bar{\ell}_j = \sum_{i=j}^{r} \tilde{\ell}_i n_i \Big/ \sum_{i=j}^{r} n_i$$

$$j = 1, 2, \ldots, r.$$

The drawback to this approach is that a large weight matrix must be inverted. The task is greatly simplified when fewer data points are created. In particular, if the divisions are not close together the covariances among the partitions are reduced. Acceptable results can then be obtained by ignoring the off-diagonal elements of A, and weighting the mean lengths by the reciprocals of their variances, or simply by the partition sample sizes.

Asymptotic variances and covariances among the estimates of θ and L_∞ are approximately

$$\mathrm{Var}\,(\hat{\theta}_R) = \frac{\sigma^2_{\hat{\beta}}}{(1-\beta)^4} \qquad \ldots 24)$$

$$\mathrm{Var}\,(\hat{L}_{\infty R}) = \frac{1}{(1-\beta)^2} \left\{ \sigma^2_{\hat{\alpha}} + \left(\frac{\alpha}{1-\beta}\right)^2 \sigma^2_{\hat{\beta}} + \left(\frac{2\alpha}{1-\beta}\right) \sigma_{\hat{\alpha},\hat{\beta}} \right\} \qquad \ldots 25)$$

and

$$\mathrm{Cov}\,(\hat{\theta}_R, \hat{L}_{\infty R}) = \frac{\alpha}{(1-\beta)^4}\, \sigma^2_{\hat{\beta}} + \frac{1}{(1-\beta)^3}\, \sigma_{\hat{\alpha},\hat{\beta}} \;. \qquad \ldots 26)$$

Another linear regression approach has been developed by Jones (1981). From (3), observe that the probability of a fish length exceeding w is

$$P\,(w) = 1 - G\,(w) \qquad \ldots 27)$$

$$= \left(\frac{L_\infty - w}{L_\infty - L_c}\right)^{\theta} \;.$$

This may be estimated empirically by p (w), the proportion of fish in the sample whose length exceeds w. Jones, assuming L_c and L_∞ are known, suggests plotting $\ln P\,(w_i)$ vs. $\ln\,(L_\infty - w_i)$, $i = 1, 2, \ldots, m$ and estimating θ as the slope of a regression line fit to these points. We note that if this regression is fitted by least squares, an appropriate weighting scheme should be used to account for the correlated errors in the sequence of $\ln P\,(w)$. There are also unresolved questions related to choice of partitions.

Some other regression methods are discussed in a later section.

MAXIMUM LIKELIHOOD METHOD

If $n_.$ fish are sampled from the density (2), and assigned to length interval midpoints as described above, the joint likelihood of the length-frequency sample is

$$\mathcal{L} = \prod_{i=1}^{r} f_i^{\,n_i} \qquad \ldots 28)$$

where

$$f_i = \frac{[L_\infty - (\tilde{\ell}_i - \frac{\Delta_i}{2})]^{\theta} - [L_\infty - (\tilde{\ell}_i + \frac{\Delta_i}{2})]^{\theta}}{(L_\infty - L_c)^{\theta}} \qquad \ldots 29)$$

is the probability of falling in the i-th length class interval. When the Δ_i are small, the exact interval probabilities may be approximated by

$$f'_i = \frac{\theta\,(L_\infty - \tilde{\ell}_i)^{\theta-1}\,\Delta_i}{(L_\infty - L_c)^{\theta}} \quad . \qquad \ldots 30)$$

For any values of θ and L_∞, $\mathcal{L}$ is maximized by setting L_c as large as possible. Therefore, under the sampling scheme adopted here, the maximum likelihood estimate of L_c is L_{min}. Further, when $\theta = 1$, the maximum likelihood estimate of L_∞ is L_{max}. Except in this singular case, the joint maximum likelihood estimates of θ and L_∞ ($\hat{\theta}_{ML}$ and $\hat{L}_{\infty ML}$) must be found numerically, applying iterative search methods to $\mathcal{L}$ or ln $\mathcal{L}$. If Newton-type methods are used, the asymptotic covariance matrix for the estimates is also easily computed. Note that in the search for $\hat{L}_{\infty ML}$, L_{max} is a lower bound. Observe also that if both L_c and L_∞ are assumed known, the maximum likelihood estimator of θ is identical to the Ssentongo-Larkin estimator, $\hat{\theta}_{SL}$.

Another likelihood model may be constructed by considering the conditional probabilities of the length class frequencies given the total sample. Such frequencies are multinomial with probabilities

$$\mu_i = f_i \Big/ \sum_{i=1}^{r} f_i \quad . \qquad \ldots 31)$$

From (29), we have

$$\mu_i = \frac{[L_\infty - (\tilde{\ell}_i - \frac{\Delta i}{2})]^{\theta} - [L_\infty - (\tilde{\ell}_i + \frac{\Delta i}{2})]^{\theta}}{[L_\infty - L_{min}]^{\theta} - [L_\infty - L_{max}]^{\theta}} \qquad \ldots 32)$$

$$i = 1, 2, \ldots, r \quad .$$

Given $n.$, L_{min}, and L_{max}, the likelihood of the sample is therefore

$$\mathcal{L} = \binom{n.}{n_1 \ n_2 \ \dots \ n_r} \prod_{i=1}^{r} \mu_i^{n_i} \qquad \dots 33)$$

and maximum likelihood estimates of L_∞ and θ and their asymptotic standard errors may be found in the usual way by maximizing $\mathcal{L}$ or ln $\mathcal{L}$ directly, or, for example, by fitting the expected length class frequencies to corresponding observations using an iteratively reweighted Gauss-Newton algorithm. In the latter approach, one finds θ and L_∞ which minimize

$$\mathcal{S} = \sum_{i=1}^{r} \omega_i \ (n_i - n.\mu_i)^2 \qquad \dots 34)$$

with weights ω_i equal to the reciprocals of the expected frequencies. The resulting maximum likelihood estimates will also be minimum chi-square estimates, and the minimized $\mathcal{S}$ will be asymptotically distributed as X^2 with $r - 3$ degrees of freedom.

In fitting the multinomial model, a question arises concerning the optimal choice of the length classes into which the data are grouped. Intervals of equal width are customary in length-frequency analysis, but as a rule of thumb observations are often pooled where necessary to ensure an expected frequency of at least 5 in each interval. This is typically required in the right tail of the distribution. More specific guidance on grouping for a fixed number of intervals, r, may be offered if an optimality criterion is adopted. For example, one sensible choice would be to set the grouping intervals in such a way that the power of the usual multinomial chi-square goodness of fit test (or likelihood ratio test) for the null distribution is maximized with respect to some selected class of alternatives. Suppose we wish to test $g(\ell)$ against the open (unspecified) class of alternative length-frequency distributions. In this situation it has been shown (Cox and Hinkley 1974) that the optimal grouping is that which uniquely maximizes the expected "entropy" of samples drawn from the null distribution, namely,

$$I = - \sum_{i=1}^{r} \mu_i \log \mu_i \ , \qquad \dots 35)$$

and that this is accomplished, in our context, by dividing the length range into intervals of equal probability under $g(\ell)$. Thus the r "optimal" length class intervals should have upper end points equal to the sequence of quantiles of the length distribution, of orders 1/r, 2/r, . . . , 1. For our multinomial model these partitions are at

$$\ell_j^* = L_{max} - (L_{max} - L_{min}) \left[1 - \frac{j}{r}\right]^{\frac{1}{\theta}} \qquad \dots 36)$$

$$j = 1, 2, \dots, r \ .$$

In practice the unknown parameters would be replaced by estimates. We have not evaluated this procedure.

Because we would expect grouping of data to result in loss of information and reduced efficiency of estimators, the actual number of intervals used should in general be as large as possible, subject to the rule that expected frequencies in each length class exceed 4. This recommendation applies provided the assumed steady-state model is correct. In cases where recruitment is actually cyclic but the steady-state model is being used to estimate θ, simulation studies (S. Ralston, unpublished data) suggest that coarser grouping of data will have the effect of smoothing out recruitment fluctuations and increasing the precision of $\hat{\theta}$ (in the same manner that pooling of several samples taken over the course of a spawning cycle will render the steady-state model applicable). In such cases, trial and error will indicate the optimal grouping.

Systematic Biases

We turn now to a consideration of systematic biases in the estimators, which arise when the underlying model assumptions are violated. In some situations serious biases result, in others the estimation procedures are robust and biases negligible.

SUBSTITUTION OF L_{MIN} FOR L_C

Since L_c is a location parameter, and L_∞ a scale parameter, they have no bearing on the shape of the theoretical length distribution. This is determined solely by θ. Estimators of θ based on samples drawn randomly from (2), and assuming L_c and L_∞ are known, therefore have distributions unaffected by L_c and L_∞, and dependent only on θ and $n_.$. However, in practice both L_c and L_∞ are unknown. In most of the procedures studied here an estimate of L_c, or of both L_c and L_∞ must be substituted, and properties of the estimators are affected.

We consider L_c first. In one of the likelihood models, the multinomial, problems related to L_c were circumvented by treating it as a nuisance parameter and conditioning the length interval probabilities on the sample end points, i.e., on L_{min} and L_{max}. In all the other methods specification of L_c is required. In these cases, L_c (or w_1 in the regression method) is typically set to L_{min}, the lower bound of the sample length-frequency distribution. Indeed, in most applications the frequency distribution is formed by truncating a larger sample at a selected L_{min}, setting aside the data below this length, and retaining for analysis only the upper "descending" portion. Setting L_c equal to L_{min} in this manner is clearly reasonable. As pointed out in the section on Maximum Likelihood Method, doing so maximizes the likelihood of the sample under (2). Further, assuming continuous sampling (all $\Delta_i = 0$), analysis of the relevant order statistics shows that

$$E\,(L_{min}) = L_c + \frac{(L_\infty - L_c)}{n\theta + 1}, \qquad \ldots 37)$$

so that for θ greater than about 2, L_{min} is essentially unbiased as an estimator for L_c, even for fairly small samples. In virtually all cases, then, systematic biases due to estimating L_c by L_{min} will be negligible.

SUBSTITUTION OF L_{MAX} FOR L_∞ IN $\hat{\theta}_{BH}$ AND $\hat{\theta}_{SL}$

In the earlier sections, we noted that both $\hat{\theta}_{BH}$ and $\hat{\theta}_{SL}$ were susceptible to systematic bias resulting from the replacement of L_∞ by an independent estimate or guess. To avoid the bias, one would need to apply one of the methods which estimate θ and L_∞ jointly. However, heretofore only Powell's method was available for this, and we are unaware of any situations in which it has been used. In the majority of cases, one of the simpler methods has been adopted instead. Except when reliable, independent estimates of L_∞ have been available from tagging or analysis of hard parts, the usual practice has been to set $L_\infty = L_{max}$, the length of the longest fish measured, or the upper bound of the length distribution. When growth is deterministic with L_∞ the upper bound to length, as assumed in (2), this substitution biases $\hat{\theta}_{BH}$ and $\hat{\theta}_{SL}$ downward, the degree of negative bias depending on θ, X_c, and $n_{\cdot}$.

If θ is large and X_c small, very large samples are required for this sytematic bias to be negligible. Specifically, under (2) it may be shown that the sample size required to ensure that L_{max} is within $100\delta\%$ of L_∞ with probability ϕ is the smallest integer exceeding

$$n(\delta, \theta) = \frac{\ln(1-\phi)}{\ln\left[1-\left(\frac{\delta}{1-X_c}\right)^{\theta}\right]} \quad \cdot \qquad \ldots 38)$$

For example, if $X_c = 0.3$ and $\phi = 0.95$, we have the following results for $n(\delta, \theta)$:

δ	$\theta = 3.0$	$\theta = 4.0$
0.01	1.0×10^6	7.2×10^7
0.05	8.2×10^3	1.2×10^5
0.10	1.0×10^3	7.2×10^3
0.20	1.3×10^2	4.5×10^2

Thus for L_{max} to be a reasonably accurate estimate of L_∞, it must represent the longest fish in a sample of several thousand. This sample need not be the same one used to estimate θ. Indeed, if the deterministic von Bertalanffy model is correct and samples are drawn under identical conditions, L_{max} should represent the longest specimen seen in all the samples taken.

To correct for this sytematic bias, it is often recommended that L_{max} be adjusted upward by some arbitrary amount, typically about 5% (e.g., Pauly 1983, who suggests dividing L_{max} by 0.95). Depending on the true value of θ and the size of the sample from which L_{max} is derived, a 5% adjustment may be a fortuitous choice, but clearly some better procedure is needed. The obvious choice, as suggested above, would be to make fuller use of the sample information to estimate θ and L_∞ jointly, i.e., use Powell's method, the regression method, or a maximum likelihood procedure. In the same vein, an interesting alternative approach which accounts for the systematic bias explicitly is based on the following pair of equations in θ and L_∞, derived from (2):

$$E(\bar{\ell}) = L_\infty - (L_\infty - L_c)\left(\frac{\theta}{\theta+1}\right) \qquad \ldots 39)$$

$$E(L_{max}) = L_\infty - (L_\infty - L_c)\prod_{i=1}^{n}\left(\frac{i\theta}{i\theta+1}\right) \qquad \ldots 40)$$

Substituting sample values of $\bar{\ell}$ and L_{max} for their expectations and solving by iteration, one can compute joint moment-type estimates of L_∞ and θ. Sampling variances and other properties of the parameter estimates can be derived empirically using bootstrap techniques.

Numerical evaluation of the bias term in (40) shows that a simple 5% adjustment of L_{max} (or division by 0.95) is rarely appropriate (Table 1). In most fishery applications, where X_c is 0.5 or greater, the correct adjustment is apt to be anywhere between about 0.5% and 25%, depending on θ and n_*. The consequences of using an incorrect adjustment may be judged by various criteria, one being the so-called "D-measure" described by Majkowski (1982). For example, a D-measure analysis of $\hat{\theta}_{BH}$ shows that the estimator is quite sensitive to errors in the estimate or guess of L_∞ (Table 2). Since $\hat{\theta}_{BH}$ is linear in L_∞, the relevant D-measure is directly proportional to the errors, and symmetrical. Combining the information in Tables 1 and 2, one can compute the expected bias in $\hat{\theta}_{BH}$ when a simple 5% upward adjustment is applied to L_{max} (Table 3). Obviously, too large an adjustment simply replaces the negative bias in $\hat{\theta}_{BH}$ with an unknown positive bias.

Table 1. Percentage negative bias in L_{max} as estimator of L_∞, as a function of X_c, θ, and n_*.

		θ		
X_c	n_*	2.0	3.0	4.0
0.0	200	6.2	15.2	24.1
	500	4.0	11.2	19.2
	1,000	2.8	8.9	16.1
0.5	200	3.1	7.6	12.0
	500	2.0	5.6	9.6
	1,000	1.4	4.5	8.0
0.75	200	1.6	3.8	6.0
	500	1.0	2.8	4.7
	1,000	0.7	2.2	4.0

Table 2. Percentage carat, i.e., $\hat{L}_\infty$ D-measure biases in $\hat{\theta}_{BH}$ due to systematic errors in $\hat{L}_\infty$ of ± δ%. (Biases in $\hat{\theta}_{BH}$ have same sign as errors in $\hat{L}_\infty$.)

		θ		
X_c	δ	2.0	3.0	4.0
0.0	1	1.5	1.3	1.3
	2	3.0	2.7	2.5
	4	6.0	5.3	5.0
	8	12.0	10.7	10.0
	16	24.0	21.3	20.0
	32	48.0	42.7	40.0
0.5	1	3.0	2.7	2.5
	2	6.0	5.3	5.0
	4	12.0	10.7	10.0
	8	24.0	21.3	20.0
	16	48.0	42.7	40.0
	32	96.0	85.3	80.0
0.75	1	6.0	5.3	5.0
	2	12.0	10.7	10.0
	4	24.0	21.3	20.0
	8	48.0	42.7	40.0
	16	96.0	85.3	80.0
	32	192.0	170.7	160.0

Table 3. Percentage bias in $\hat{\theta}_{BH}$ when L_∞ is set at $L_{max}/0.95$, under various combinations of X_c, θ, and n_*.

		θ		
X_c	n_*	2.0	3.0	4.0
0.0	200	−1.9	−14.3	−25.1
	500	+1.6	−8.7	−18.7
	1,000	+3.5	−5.5	−14.6
	∞	+7.9	+7.0	+6.6
0.5	200	+6.0	−7.3	−18.4
	500	+9.5	−1.7	−12.1
	1,000	+11.4	+1.4	−7.9
	∞	+15.8	+14.0	+13.2
0.75	200	+21.5	+6.7	−5.3
	500	+25.3	+12.3	+1.6
	1,000	+27.2	+15.7	+5.3
	∞	+31.6	+28.1	+26.3

VARIABILITY IN ASYMPTOTIC LENGTH

Another problem of systematic bias can arise if L_{max} is substituted for L_∞ in the Beverton-Holt or Ssentongo-Larkin formulas but the deterministic von Bertalanffy model is incorrect. Specifically, if the growth trajectories of individual fish converge on different asymptotic lengths, i.e., if $\sigma_\lambda^2 > 0$, L_{max} can easily exceed the expected asymptotic length, L_∞, when derived from large samples. As Powell's model shows, the positive bias in L_{max} as an estimator of L_∞ can be significant when σ_λ^2 is relatively large and θ is not. Subsequent estimates of θ will also be inflated.

Under such conditions, the maximum likelihood methods based on (2) will also give biased results, since L_{max} is regarded as a lower bound on $\hat{L}_{\infty ML}$. Partial remedies might be to base estimation on the more realistic likelihood arising from the stochastic model of Powell, or to estimate L_∞ by some statistic reflecting the average size of the largest fish seen in the catches. Using Powell's likelihood would entail estimation of σ_λ^2 as well as θ and L_∞, and would require numerical approximation of the theoretical length frequencies.

We note that because the regression method at (19) is based on fitting sample mean lengths to their expected values, which are unaffected by σ_λ^2, the estimates of L_∞ it produces do not suffer from these types of bias. In contrast, Powell's estimators of θ and L_∞ are derived assuming $\sigma_\lambda^2 = 0$, so are presumably biased when this assumption is violated.

VARIABLE RECRUITMENT

Variability in recruitment violates a basic assumption of the methods, but is not particularly troublesome as long as the process is stationary, i.e., there has been no systematic trend in R. When spawning is periodic and reasonably regular, and survival of pre-recruits is stable, the effects of variable recruitment can be smoothed out by sampling the population on several occasions during a complete spawning cycle and pooling the individual length-frequency distributions. Distributions should be weighted by an index of population size, such as average catch per unit effort. When recruitment is trending up or down the smoothing procedure is ineffective, and serious bias in estimates of θ can result from use of any of the estimators, the bias being in the same direction as the trend in R.

VARIABLE MORTALITY

If the simple, constant mortality models are applied to length-frequency data when in fact mortality is variable, subsequent population assessments and yield computations based on the estimates of θ and L_∞ are likely to be biased. But since the robustness of the methods in such circumstances has not been studied, the specific ramifications are unclear.

Nevertheless, several forms of variability may be identified, and some are more easily handled than others. One kind is temporal variability in mortality, caused, for example, by a systematic change in fishing effort or stock vulnerability. Accommodating this kind of variability would be relatively difficult, because the simplifying steady state conditions no longer exist. A second type of variation is that due to size (or age) dependence of mortality. Assuming such size-specificity is time invariant, construction of steady-state population models and parameter estimation procedures is straightforward. We develop such procedures in the following section.

Estimating Length-Specific Fishing Mortality Rates

When aging of the catch is feasible, the analysis of cohort catch histories is one of the chief methods of estimating virtual populations and age-specific mortality rates. When it is not, but the catch is known by length class interval, one may estimate population size and length-specific mortality using a length-based analog of the cohort methods (e.g., Jones 1981; Pauly 1984). These are applicable under more restrictive assumptions than cohort analysis. Recruitment must be approximately constant, or smoothed out by pooling catch-at-length data, a growth model must be assumed, and mortality rates must be constant in time.

The method developed by Jones and extended by Pauly is based on the deterministic von Bertalanffy growth model and a constant instantaneous natural mortality rate, M. Assuming M, the fishing mortality rate in the largest length class, F_r, and the von Bertalanffy parameters K and L_∞ are known, the length-based VPA proceeds recursively in the same manner as the popular age-based methods, beginning with the catch in the largest length class and working backward to estimate the sequence of length-specific population sizes and fishing mortality rates. Alternatively, if only L_∞ and the "exploitation rate" F_t/Z_t in the terminal length class are specified (where $Z_t = F_t + M$), estimates of M/K and the remaining F_i/Z_i may be found, there being in either case one parameter estimated for every data point, i.e., every length-interval catch.

In this section we adopt the same steady-state assumptions, and describe an alternative procedure for estimating length-specific mortality rates, growth parameters and population size jointly from a regression analysis of catch by length class. This is simply an extension of the age-based catch curve models developed by Ricker (1948, 1975) and frequently applied to cohort analysis.

Historically, in fact, the age-based catch curves were preceded by length-based methods due to Edser (1908), Heincke (1913), and Baranov (1918).

We assume a steady state, with constant recruitment, R, constant natural mortality rate M, and knife-edge selection at age t_c. Beyond t_c, the fishing mortality rate F is described as a step function of age t, namely

$$F(t) = F_i \qquad \ldots 41)$$

$$t_i < t < t_i + \delta_i$$

$$i = 1, 2, \ldots, r$$

where $t_i = t_c + \sum_{j=1}^{i} \delta_{j-1}$ and the δs are successive age increments ($\delta_0 = 0$). Under these conditions, the number of fish caught in the i-th age class, $(t_i, t_i + \delta_i)$, during a time period of duration τ, is

$$C(t_i, t_i + \delta_i) = R\tau \left(\frac{F_i}{F_i + M} \right) \left(1 - e^{-(F_i + M)\delta_i} \right) e^{-\sum_{j=1}^{i} (F_{j-1} + M)\delta_{j-1}} \qquad \ldots 42)$$

$$= C_i, \text{ say.}$$

Let the δ_i correspond, by virtue of a bijective growth model, to a specific sequence of length intervals into which the catch is conveniently grouped, say the sequence $(w_i, w_i + \Delta_i)$, $i = 1, 2, \ldots, r$, where $w_i = L_c + \sum_{j=1}^{i} \Delta_{j-1} = \tilde{\ell}_i - \Delta_i/2$ and $\Delta_0 = 0$. If the von Bertalanffy model is used, the correspondence is established by the relation

$$\delta_i = -\frac{1}{K} \ln \left\{ \frac{L_\infty - (w_i + \Delta_i)}{L_\infty - w_i} \right\} \cdot \qquad \ldots 43)$$

Making the subsitution of length for age, we have

$$C_i = R\tau\left(\frac{F_i}{Z_i}\right)\left[1-\left(\frac{L_\infty-(w_i+\Delta_i)}{L_\infty-w_i}\right)^{\theta_i}\right]\prod_{j=1}^{i}\left(\frac{L_\infty-(w_{j-1}+\Delta_{j-1})}{L_\infty-w_{j-1}}\right)^{\theta_{j-1}} \qquad \ldots 44)$$

where $Z_i = F_i + M$, $\theta_i = Z_i/K$, and $w_o = \Delta_o = Z_o = 0$.

If the fishing mortality rate is constant after recruitment, (44) reduces to

$$C_i = R\tau\left(\frac{F}{Z}\right)\left[1-\left(\frac{L_\infty-(w_i+\Delta_i)}{L_\infty-w_i}\right)^{\theta}\right]\left(\frac{L_\infty-w_i}{L_\infty-L_c}\right)^{\theta} \qquad \ldots 45)$$

$$= C_. \left[\frac{(L_\infty-w_i)^{\theta} - (L_\infty-(w_i+\Delta_i))^{\theta}}{(L_\infty-L_c)^{\theta}}\right]$$

where $C_./\tau = R\ F/Z$ is the constant total catch in a unit time interval. Alternatively, (45) may be derived from the length density at (2), i.e.,

$$C_i = \tau F\int_{w_i}^{w_i+\Delta_i} N(\ell)\,d\ell \qquad \ldots 46)$$

$$= \tau F\,\frac{R}{Z}\int_{w_i}^{w_i+\Delta_i} g(\ell)\,d\ell$$

$$= C_.\ [\,G(w_i+\Delta_i) - G(w_i)\,]$$

where $N(\ell)$ is the number of fish of length ℓ in the steady-state population. Recalling the earlier discussion on maximum likelihood methods, $C_i/C_.$ is seen to be the probability that a fish taken at random from the population will be in the i-th length class interval, a function of L_c, L_∞, and Z/K, as in (29).

As formulated in (44), the length-based catch model is a set of r equations in r + 4 unknown parameters (R, M, K, L_∞, and the F_i, $i = 1, 2, \ldots, r$), so additional restrictions must be imposed to permit estimation. As usual, this may be done by making further structural assumptions about fishing mortality. One possibility is to assume that fishing mortality rates are constant and distinct within $m < r - 4$ segments of the catchable length range. Another is to relate fishing mortality rate to length class by a particular parametric model, such as the quadratic $F(i) = a + bi + ci^2$. In the latter event the problem is reduced to estimating seven parameters from the r observed length class catches.

With proper structuring, so that adequate degrees of freedom are available, the q-model parameters can be estimated by iteratively reweighted least squares. When weights are set equal to the expected catches, this amounts to minimizing the familiar Pearson chi-square statistic. For large samples, this statistic will be distributed as χ^2 with $r - q - 1$ degrees of freedom. The resulting parameter estimates will also be maximum likelihood estimates, and if a Gauss-Newton algorithm is used in the optimization an estimate of their asymptotic covariance matrix will be readily available.

For some purposes individual estimates of M, K and the fishing mortality rates may not be necessary, and in this event the catch models can be parameterized in terms of R, L_∞, and the ratios M/K and F_i/Z_i, $i = 1, 2, \ldots, r$. (In the structured mortality model, let $F_i/Z_i = a' + b'i + c'i^2$). In (44), for example, this is accomplished by replacing θ_i by the product

$$\theta_i = \left(\frac{M}{K}\right)\left(1 - \frac{F_i}{Z_i}\right)^{-1} \quad . \qquad \ldots 47)$$

In the models just described, it is assumed that the entire catch is taken by a single gear, so that estimating the length-specific fishing mortality rates is equivalent to estimating the selectivity curve for that gear with respect to the species under study. If more than one gear is significantly involved, then the catch data from all of them must be combined (properly weighted), and the analysis applied to the aggregate information. As in age-based VPA methods, gear-specific fishing mortality rates can be computed by allocating the fishing mortality estimate for each length class in proportion to the associated catch by each gear.

We have not yet applied these procedures, nor investigated their robustness and other properties. Nevertheless, we would speculate that length-specific fishing mortality rates will not be identifiable, or estimable with much precision, unless they vary fairly dramatically with length. Further, separation of natural mortality and the fishing mortality rates and useful precision in the mortality rate estimates will likely require that a substantial fraction of the total mortality be due to fishing. Thus, the model will probably not estimate fishing mortality rates well when applied to lightly exploited stocks.

However, when adequate length-frequency data are available from virgin populations or very lightly exploited stocks, a regression model may be developed to estimate L_∞, K and a set of length-specific natural mortality rates. Taking C_i to be a small sample proportional to τ and the steady state population size in the i-th length class, we will have, corresponding to (44), the model

$$C_i = \delta\left(\frac{R\tau}{M_i}\right)\left[1 - \left(\frac{L_\infty - (w_i + \Delta_i)}{L_\infty - w_i}\right)^{\frac{M_i}{K}}\right] \prod_{j=1}^{i} \left(\frac{L_\infty - (w_{j-1} + \Delta_{j-1})}{L_\infty - w_{j-1}}\right)^{\frac{M_{j-1}}{K}} \qquad \ldots 48)$$

where δ is the proportionality constant and M_i is the instantaneous natural mortality rate in the i-th length interval. Again, further restrictions on the parameters are necessary to allow estimation. Structural relationships among the M_i could be introduced, as above. If the segmented mortality model is adopted, such independent information as average size at maturity could be used to define points of change in the model. Note that the steady-state recruitment level, R, is no longer estimable, and difficulty in estimating the mortality parameters can be expected unless the length-specific changes in mortality rate are striking. When variation in natural mortality rates is negligible, only L_∞ and $\theta = M/K$ are estimable from the length-frequency information.

Evaluation of Methods

Four of the constant-mortality methods described above were compared by Monte Carlo procedures in situations where both θ and L_∞ were to be estimated from length-frequency data and L_c was assumed known. Specifically, performance of the following procedures was examined:

(i) L_∞ was set equal to L_{max}, and the classical Beverton-Holt estimator then applied to estimate θ.

(ii) Powell's method of moments was applied to estimated L_∞ and θ, given the sample mean and variance of ℓ.

(iii) The weighted regression method, based on the linear relationship between $\bar{\ell}$ and L_c, was used to estimate L_∞ and θ.

(iv) The continuous maximum likelihood method was applied to estimate L_∞ and θ.

For specified values of X_c and θ samples of lengths were drawn from (6) by the inverse transformation method, a well-known Monte Carlo procedure. Let u be a random variate uniformly distributed on (0, 1). Then a corresponding random variate from (6), x, is uniquely determined by

$$x = 1 - (1 - X_c)(1 - u)^{\frac{1}{\theta}} \qquad \ldots 49)$$

A uniform random number generator was used to produce sets of u_i and the associated random lengths, x_i, $i = 1, 2, \ldots, n_{\cdot}$. Except when the level of aggregation was being studied, experiments were conducted using the raw generated lengths. Thus a continuous distribution of the sample variates was assumed (all $\Delta_i = 0$). Recall that use of the auxiliary variable, X, is equivalent to scaling the length axis so that $L_\infty = 1.0$.

Performance of the estimators was measured by the coefficient of variation (CV), the relative bias (B), and the coefficient of error (CE). For a particular parameter π, we define these as

$$CV(\hat{\pi}) = \frac{\sqrt{var(\hat{\pi})}}{\pi}, \qquad \ldots 50)$$

$$B(\hat{\pi}) = \frac{Bias(\hat{\pi})}{\pi}, \qquad \ldots 51)$$

and

$$CE(\hat{\pi}) = \sqrt{CV(\hat{\pi})^2 + B(\hat{\pi})^2} = \frac{\sqrt{MSE(\hat{\pi})}}{\pi} \qquad \ldots 52)$$

where MSE ($\hat{\pi}$) is the mean squared error of $\hat{\pi}$. For each set of experimental circumstances, these measures were computed with respect to $\hat{L}_\infty$ and $\hat{\theta}$ on the basis of 200 replicate samples. Trial simulations showed that this number of replicates was sufficient to describe the sampling distributions of the estimators accurately.

Experimental variables of primary interest were the level of θ and the sample size, $n_{\cdot}$. Of lesser interest were $1 - X_c$, the proportion of the total possible length range selected in the sample, the level of aggregation in the data, and, in the regression method, the number of partitions and type of weighting used. Below we describe each particular simulation experiment, and the chief results.

Experiment 1

The performance of the four estimation procedures was compared at three levels of θ (2.0, 3.0, 4.0) and three levels of n_* (200, 500, 1,000), with $X_c = 0$.

Results: The Beverton-Holt method produces estimates with the lowest CV, but these estimates have a substantial negative bias, especially with small n_* and large θ (Tables 4 and 5). The Powell method, on the other hand, computes estimates with the highest CV of any method examined, but generally with the smallest B (positive) as well. Estimates derived by the regression

Table 4. Performance measures for various estimators of L_∞, as a function of θ and n_*.

						θ				
			2.0			3.0			4.0	
n_*	Method	CV	B	CE	CV	B	CE	CV	B	CE
200	BH	3.3	−5.9	6.8	6.1	−15.3	16.5	7.1	−23.2	24.3
	P	10.7	2.1	10.9	15.9	1.8	16.0	39.5	12.4	41.4
	R	5.8	−1.0	5.9	10.3	−5.3	11.6	19.2	−7.2	20.5
	ML	4.7	−1.7	5.0	10.5	−3.2	11.0	25.2	3.7	25.4
500	BH	2.2	−3.7	4.3	4.0	−11.4	12.1	5.8	−19.2	20.1
	P	6.6	0.8	6.6	11.2	1.7	11.3	13.0	1.3	13.0
	R	3.4	−0.4	3.4	6.6	−1.9	6.9	11.1	−4.2	11.9
	ML	2.6	−0.7	2.7	6.0	−1.1	6.1	9.6	−1.5	9.7
1,000	BH	1.7	−2.7	3.2	3.5	−8.6	9.3	4.6	−16.7	17.3
	P	4.6	−0.1	4.6	7.3	1.2	7.4	11.6	0.8	11.7
	R	2.4	−0.4	2.4	5.0	−0.8	5.1	7.9	−3.8	8.8
	ML	1.8	−0.6	1.9	4.5	−0.2	4.5	7.5	−1.5	7.7

Table 5. Performance measures for various estimators of θ, as a function of θ and n_*.

						θ				
			2.0			3.0			4.0	
n_*	Method	CV	B	CE	CV	B	CE	CV	B	CE
200	BH	8.7	−9.0	12.5	8.8	−20.3	22.2	9.7	−28.5	30.1
	P	20.0	3.5	20.3	24.0	3.0	24.2	54.8	17.5	57.5
	R	12.2	−3.0	12.6	16.7	−8.0	18.5	28.6	−9.0	30.0
	ML	12.2	−3.0	12.6	17.3	−4.3	17.9	36.3	5.8	36.8
500	BH	5.0	−5.0	7.1	5.8	−15.0	16.1	7.5	−23.8	24.9
	P	12.2	1.5	12.3	17.0	3.0	17.3	18.0	2.0	18.1
	R	7.1	−0.5	7.1	10.5	−2.7	10.9	16.4	−4.8	17.1
	ML	7.1	−1.0	7.1	10.5	−1.3	10.6	13.9	−1.5	14.0
1,000	BH	5.0	−4.0	6.4	5.8	−10.7	12.1	6.1	−20.7	21.6
	P	8.7	−0.5	8.7	11.1	2.3	11.3	15.4	1.3	15.5
	R	5.0	−1.0	5.1	8.2	−0.3	8.2	10.6	−5.0	11.7
	ML	5.0	−1.0	5.1	7.5	0.3	7.5	10.3	−2.0	10.5

method, with optimal weighting and 10 partitions, appear to have slightly greater bias than the maximum likelihood estimates (negative in both cases) and the Powell method, but are far superior to the Beverton-Holt estimates in this regard. In terms of precision, the regression method is almost as good as the maximum likelihood method.

As expected, the maximum likelihood method, among the estimators examined, generally has the best overall performance in terms of mean squared error and CE, followed closely by the regression method. The maximum likelihood estimates of L_∞ almost always had the smallest CE. At $\theta = 2$, the Beverton-Holt procedure performed relatively well in terms of CE, virtually on a par with the maximum likelihood and regression methods, but it did comparatively poorly at higher values of θ. At $\theta = 2$, the Powell estimates generally had the largest CE of any method; otherwise, Beverton-Holt estimates tended to have the largest CE.

In all methods, the effect of increasing θ is to increase B, CV, and CE for both parameters estimated. As expected, increasing $n_.$ uniformly reduces B, CV, and CE.

Experiment 2

The effects on Beverton-Holt estimates of using a bias-corrected estimate of L_∞ were studied. In each generated sample, L_{max} was increased by the theoretical bias term on the right side of (40), and this corrected estimate of L_∞ was used as a basis for estimating θ.

Results: Adjustment of L_{max} for bias resulted in a significant reduction in B ($\hat{L}_{\infty BH}$), as anticipated, and also in B ($\hat{\theta}_{BH}$), and increases in CV for both parameter estimates. The CE's were changed very little by the adjustment. In practice, the bias correction procedure would have to be implemented by iterative solution of (39) and (40), i.e., the bias term itself would have to be estimated, and the properties of the estimators would be affected.

Experiment 3

The effects of X_c on the performance of all methods was studied, with $n_.$ fixed at 200, and θ at 3.0. Samples were drawn from (6) with X_c set at three levels (0.0, 0.5, 0.75) and estimators were applied with X_c assumed known.

Results: As expected, increasing the value of X_c (with $n_.$ unchanged) reduced B, CV, and CE for $\hat{L}_\infty$ in all methods, but had no effect on estimates of θ. With respect to estimation of L_∞, increases in X_c have the same effect as increasing sample size with X_c fixed. On the other hand, as pointed out earlier, distributions of $\hat{\theta}$ are unaffected by the magnitude of X_c when X_c is known. The amount of information about θ conveyed by the sample depends only on θ and the sample size. (Note, for example, that in the regression model if L_c is known, θ is completely determined by the slope of the regression line, the expected value of which is independent of X_c.)

Experiment 4

The effects of weighting schemes and the number of partitions on estimates produced by the regression method were studied. Weights were determined in three ways: (1) the complete covariance matrix was inverted, (2) the diagonal elements only were inverted, and off diagonal elements neglected and (3) uniform weights were assigned. Under the diagonal weighting scheme, regression estimates were computed with both 10 and 20 partitions.

Results: Use of the full covariance weighting produced the most precise estimates, as expected, but these estimates had greater bias than those produced with either diagonal or uniform weighting. However, the fully weighted estimates had the lowest CEs. When 10 partitions are used, diagonal weighting, or even equal weighting, appears to be satisfactory.

Similarly, increasing the number of partitions (points) in the regression reduced the CVs at the expense of greater bias and resulting estimates had smaller CEs.

Experiment 5

Effects of aggregation were assessed by comparing estimates derived using continuous data (i.e., exact lengths) with those computed from data grouped by intervals of width $L_\infty/100$.

Results: In most cases, aggregation at this level slightly increased the CVs and reduced relative biases (except for increases in bias of Powell estimates). The CEs were increased slightly in all methods except the Beverton-Holt, in which case they were reduced. (Note that among the estimators studied, those of the Beverton-Holt procedure are distinguished by having their CEs composed mostly of bias. In the other methods, the variance dominates.)

Conclusions and Recommendations

Each of the methods outlined here has attributes which make it the method of choice in particular circumstances.

The classical estimator of θ due to Beverton and Holt, for example, is the only one applicable when the available information is limited to the mean length in a sample of fish above L_c and an independent estimate of L_∞. The difficulty with the method, of course, is that considerable bias in $\hat{\theta}_{BH}$ will result unless L_∞ is known accurately. As discussed earlier, estimating L_∞ by L_{max} is advisable only if a very large sample has been measured. Alternatively, the estimate of L_∞ (and K) can be derived independently by fitting the von Bertalanffy model to tag recapture data or to readings of hard parts, and then substituted into the Beverton-Holt estimator. Here, too, L_∞ is rarely determined with sufficient accuracy unless fish are aged, or tagged and recovered, over the entire length range and an asymptote is clearly evident. These conditions are difficult to achieve when there is a low probability that a tagged fish will survive long enough to approach the asymptotic length and be recaptured, or when otoliths and other hard structures are difficult to read and interpret in old fish. The result is often serious bias in estimates of L_∞ and K. However, since these biases are inversely related, some compensation occurs when the pair of parameter estimates is inserted into the Beverton-Holt formula to compute Z.

If the variance among lengths in the sample is known in addition to the mean, a considerable improvement is afforded by Powell's method of moments, which provides joint estimates of θ and L_∞ without reliance on independent information. If tag recapture statistics or otolith data are available, they may then be used to estimate K with L_∞ fixed at $\hat{L}_{\infty_p}$. An estimate of Z follows immediately.

The weighted regression method is attractive for several reasons, and has proven to be the most dependable of the constant-mortality methods in practice. In statistical terms, it is very efficient, giving parameter estimates with almost minimum mean squared error. It also is insensitive to assumptions on L_∞, which may be regarded either as an absolute upper bound to length, or as the expected asymptotic length in a stochastic von Bertalanffy growth process. Among the simple methods studied, only the Beverton-Holt procedure shares this robustness with respect to L_∞.

Another reason for recommending the regression method is that it provides a simple graphical means of judging the appropriateness of underlying model assumptions; serious violations can be detected directly by comparing observed mean lengths above the specified cut-off points with those predicted by the model.

Perhaps the only drawback to the regression method is the need to compute the covariance matrix A to achieve full efficiency. However, little is sacrificed by weighting simply by the partition sample sizes.

From a theoretical standpoint, the maximum likelihood procedure is highly attractive, yielding estimators with greatest asymptotic efficiency and other desirable properties. However, it is not robust to violations of the assumption on L_∞, and this has apparently been a source of difficulty in applications. In a number of cases, searches for L_∞ have converged on L_{max}, suggesting overestimation of L_∞ and θ. Use of Powell's stochastic model (or a suitable variant) in a maximum likelihood context might be more appropriate, but we have not tried it.

In the methods outlined here, estimators are based on length-frequency data entirely, or on combinations of length-frequency data and other information, applied sequentially. Where consistent information on growth and mortality is available from several independent sources, e.g., from tagging, otolith analysis or length-frequency samples, it may also be used jointly to estimate the parameters of a common steady-state model. A maximum likelihood approach is most appropriate here, provided one can specify the forms of the various error distributions. An ancillary benefit of joint estimation is that the relative contribution of each type of information toward the mean squared errors of the parameter estimates can be determined, providing useful guidance for research planning and allocation of resources. Examples of joint estimation of growth parameters from length-increment data and age-length data have been reported by Laurs et al. (1983), who combined tag recapture results with counts of daily increments on sagittae of North Pacific albacore, by Kirkwood (1983), who pooled tag recapture data with information on "aged" length-frequency modes in southern bluefin tuna and by Morgan (Part I, this vol.). Extension of such techniques to joint estimation of growth, mortality and recruitment parameters in more general contexts is straightforward.

References

Baranov, F.I. 1918. On the question of the biological basis of fisheries. Nauchn. Issled. Ikhtiol. Inst. Izv. 1: 81-128. (In Russian.)

Beverton, R.J.H. and S.J. Holt. 1956. A review of methods for estimating mortality rates in fish populations, with special reference to sources of bias in catch sampling. Rapp. P.-v. Réun. Cons. Int. Explor. Mer 140: 67-83.

Cox, D.R. and D.V. Hinkley. 1974. Theoretical statistics. Chapman and Hall, London.

Edser, T. 1908. Note on the number of plaice at each length, in certain samples from the southern part of the North Sea, 1906. J.R. Stat. Soc. 71: 686-690.

Gulland, J.A. 1983. Fish stock assessment: a manual of basic methods. FAO/Wiley Series on Food and Agriculture.

Heinke, F. 1913. Investigations on the plaice. General report. 1. Plaice fishery and protective regulations. Part I. Rapp. P.-v. Réun. Cons. Int. Explor. Mer 17A: 1-153.

Jones, R. 1981. The use of length composition data in fish stock assessments (with notes on VPA and cohort analysis). FAO Fish. Circ. 734. 60 p.

Kirkwood, G.P. 1983. Estimation of von Bertalanffy growth curve parameters using both length-increment and age-length data. Can. J. Fish. Aquat. Sci. 40: 1405-1411.

Laurs, R.M., R. Nishimoto, J.H. Uchiyama and J.A. Wetherall. 1983. Age and growth of North Pacific albacore, *Thunnus alalunga*, determined from tag returns and daily increments on otoliths. Eighth North Pacific Albacore Workshop, Document NPALB/83/1. National Marine Fisheries Service, Southwest Fisheries Center Admin. Report LJ-84-04.

Majkowski, J. 1982. Usefulness and applicability of sensitivity analysis in a multispecies approach to fisheries management, p. 149-165. *In* Pauly, D and G.I. Murphy (eds.) Theory and management of tropical fisheries. ICLARM Conference Proceedings 9, 360 p. International Center for Living Aquatic Resources Management, Manila, Philippines and Division of Fisheries Research, Commonwealth Scientific and Industrial Research Organisation, Cronulla, Australia.

Pauly, D. 1983. Some simple methods for the assessment of tropical fish stocks. FAO Fish. Tech. Pap. 234, Rome, 52 p.

Pauly, D. 1984. Fish population dynamics in tropical waters: a manual for use with programmable calculators. ICLARM Studies and Reviews 8, 325 p. International Center for Living Aquatic Resources Management, Manila, Philippines.

Pauly, D. and N. David. 1981. ELEFAN I, a BASIC program for the objective extraction of growth parameters from length-frequency data. Meeresforsch. 28: 205-211.

Powell, D.G. 1979. Estimation of mortality and growth parameters from the length frequency of a catch. Rapp. P.-v. Réun. Cons. Int. Explor. Mer 175: 167-169.

Ricker, W.E. 1948. Methods of estimating vital statistics of fish populations. Indiana Univ. Publ. Sci., Ser. 15, 101 p.

Ricker, W.E. 1975. Computation and interpretation of biological statistics of fish populations. Fish. Res. Board Can. Bull., 191, 382 p.

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

Ssentongo, G.W. and P.A. Larkin. 1973. Some simple methods of estimating mortality rates of exploited fish populations. J. Fish. Res. Board Can. 30: 698-698.