STOCHASTIC AGE-FREQUENCY ESTIMATION USING THE VON BERTALANFFY GROWTH EQUATION

NORMAN W. BARTOO[1] AND KEITH R. PARKER[2]

ABSTRACT

The method of estimating age frequency from length frequency via the von Bertalanffy growth equation is deterministic and yields biased results. Most of the bias can be removed by incorporating a stochastic element in the von Bertalanffy relationship. The stochastic element is based on estimated probabilities of lengths by intervals at age, the probabilities being estimated from variances in lengths-at-age. Based on age-length samples from the Pacific bonito fishery the stochastic method gives improved age-frequency estimates over those obtained by the deterministic method. The stochastic application may be generalized to all growth models including discontinuous growth such as in crustaceans.

Complex population dynamics techniques rely heavily on age-structure information. Frequently, appropriate assessment techniques for a stock require an estimate of the age frequency of that stock. For example yield-per-recruit analysis (Ricker 1958) is computed on the dynamic relationship between growth and mortality: Mortality rates when computed via cohort analysis (Murphy 1965) are based on estimated age frequency.

For some species accurate aging methods are not available. When feasible, determining the age of fish and consequently computing an age frequency are most accurately accomplished by visual inspection of scales, otoliths, or other structures (Ricker 1958). Such visual inspection is time consuming and often expensive. To reduce the cost and time of estimating the age structure of a fisheries catch, age frequency is usually estimated from sampled length frequency, the age-length relationship being described by either an age-length key or a growth curve such as the von Bertalanffy growth curve (Ricker 1958). The growth curve method is used when there are insufficient data to construct an age-length key.

Age-length keys work on the principle that age can be estimated from length using information contained in a previously or concurrently aged sample from the population. As long as the proportion of length-at-age remains the same for all ages, then the age-length key will yield unbiased

estimates of age for any sampled lengths from that population. However, since the estimated parameters of an age-length key—proportions of age-at-length—are dependent on the sampled population used to construct the key, the application of the key to the population with altered age structures can yield inaccurate results. Kimura (1977) and later Westrheim and Ricker (1978) demonstrated that under conditions of varying year-class strength and substantial overlap of lengths between ages, age-length keys can yield nearly useless estimates of numbers-at-age.

Clark (1981) effectively removes age-length key bias by first proportioning numbers in length intervals at age over time and then using the matrix of these proportions standardized over time to compute least-squares estimates of age frequency from the vector of length frequency. Effective applications of many stock assessment growth and mortality based methods require that ages are expressed in fractions of years (Ricker 1958; Lenarz et al. 1974). The large number of aged fish required to construct a sufficient key for a large number of ages is difficult and expensive to attain. Even with Clark's bias correction procedure, the construction of a sufficient key can present difficulties due to data needs.

In this paper we deal specifically with the von Bertalanffy growth equation and the application of stochastic methods to reduce or eliminate biases. However, it should be noted that the method presented here may be applied to any growth equation as well as to cases where no growth equation has been fitted or where growth is discontinuous as in crustaceans.

The von Bertalanffy growth equation mathe-

[1]Southwest Fisheries Center La Jolla Laboratory, National Marine Fisheries Service, NOAA, 8604 La Jolla Shores Drive, La Jolla, CA 92038.
[2]1837 Puterbaugh Street, San Diego, CA 92103.

matically models the relationship between age and length, length being the dependent variable (see Equation (1)). As suggested by Gulland (1973), age is estimated from length by algebraically rearranging the growth equation so that age is the dependent variable (see Equation (2)). Regardless of whether length or age is the dependent variable, the von Bertalanffy relationship is deterministic: There is a one-to-one correspondence between age and length.

For the von Bertalanffy growth equation, age frequency is estimated from a length sample as follows:

1) For each length compute the corresponding age.
2) For each age interval, usually the interval between midpoint ages of adjacent ages, sum the number of aged fish falling within the interval.
3) The age frequency is then the total number of aged fish falling within each age interval.

Use of the von Bertalanffy growth equation for age-frequency estimation results in several types of biases, different from those inherent in age-length keys. This paper documents these biases and proposes a method for their resolution.

BIASES

When growth is modeled according to the von Bertalanffy age-length relationship (Brody 1945; Ricker 1958),

$$L_t = L_\infty (1 - \exp[-k(t - t_0)]), \qquad (1)$$

then age, t, can be converted to length:

$$t = t_0 + \ln (1 - L_t/L_\infty)/(-k) \qquad (2)$$

where L_t = length at age t
   L_∞ = the asymptotic length
   k = the rate at which length reaches L
   t_0 = hypothetical age at which fish would have zero length.

When computing numbers-at-age from Equation (2), estimation bias occurs. One bias is due to L_∞ being a fitted parameter. Thus, all numbers-at-length greater than L_∞ must either be eliminated or arbitrarily distributed to older ages. Bias also results when lengths approach L_∞ and are mathematically allocated to ages above those

attainable by fish within the stock. As lengths (L) approach L_∞, Equation (2) will yield unreasonably old ages.

Additional bias results from the deterministic nature of the von Bertalanffy equation: Back calculations of length to age, Equation (2), are on a one-to-one basis. Thus, for any length there is a determined age. In reality, there can be a number of possible ages for any given length, the most probable age-at-length being that with the highest relative contribution of numbers-at-length. Since these back calculations are without probabilistic arguments, the determined age is not necessarily the most probable for the given length.

Back calculations of length to age also result in a mathematical estimation bias due to the switching of independent and dependent variables in going from Equation (1) to Equation (2). The degree of bias is likely to be a function of the amount of residual error in fitting Equation (1). The bias will probably not be consistent between cases and the degree of bias will have to be considered separately for each case. Consequently this bias is not specifically dealt with in this paper.

A computer model can readily demonstrate this bias. For von Bertalanffy parameters: $L_\infty = 90.0$ units, $t_0 = 0.0$ units, and $k = 0.30$, predetermined numbers-at-age are arbitrarily distributed normally with a standard deviation equal to 3 units about the von Bertalanffy length-at-age, Equation (1), for ages I through X. A length-frequency vector is then generated by 1) multiplying the number-at-age times the probability of age occurring within each 0.5 unit length interval, thus for each age generating a vector of number-at-length for length intervals between 0 and 100 units, and 2) accumulating numbers-at-length for each length interval over all ages. The numbers-at-age are then deterministically estimated from Equation (2) by accumulating numbers-at-length over the length intervals at age.

The bias from this model is illustrated in Table 1. Input and back-calculated numbers-at-age and their differences are listed in columns 2, 3, and 4, respectively. The input numbers-at-age represent a sample age distribution where either catchability, recruitment, mortality, or some combination thereof, are age-class variant. Differences, column 4, indicate a strong bias which increases with overlap of length distributions at age. One hundred and eleven fish were aged to be

92

TABLE 1.—Input and estimated numbers-at-age for both the deterministic (col. 3) and stochastic (col. 5) models, with the input numbers-at-age in column 2. The differences between the input numbers-at-age and the deterministic estimates are given in column 4.

Age (1)	Numbers-at-age			
	Input (2)	Deterministic (3)	Difference (4)	Stochastic (5)
I	200	199	1	200
II	400	399	1	400
III	800	760	40	800
IV	200	267	-67	200
V	600	441	159	600
VI	300	378	- 78	300
VII	400	320	80	400
VIII	300	258	42	300
IX	100	164	-64	100
X	100	68	32	100
>X	—	111	-111	—
Inf.	—	35	- 35	—

greater than the maximum age, 10. Thirty-five had lengths greater than L_∞ and, consequently, were not classifiable.

BIAS RESOLUTION

With estimated variance of length-at-age a stochastic model can be built from the von Bertalanffy relationship: For any age the probability of a specific length interval is the probability of that interval taken over all length intervals containing that age. Thus for all ages a probability matrix ("P"-matrix) of dimension r by c can be computed, where r = the number of rows, or length intervals, and c = the number of columns, or ages, then $P(1,1)$ = (max. length, min. age). If the number-at-age vector is "a" (a_1 = (min. age)) and the number-at-length vector is L (L_1 = (max. length)), then

$$P\,a = L. \qquad (3)$$

And as long as $r > c$ then the numbers-at-age vector can be uniquely solved via least-squares:

$$a = (P'P)^{-1}P'L. \qquad (4)$$

Applying this stochastic method (Equation (4)) to the previous example, the numbers-at-age generated from the number-at-length vector are given in column 5 of Table 1. Since the probabilities of the P-matrix are the same as those used to generate the number-at-length vector, it is not surprising that the solution yields unbiased results. This computed example illustrates that the stochastic method yields unbiased estimates of age frequency.

PACIFIC BONITO

For the Pacific bonito, *Sarda chiliensis*, of the eastern tropical Pacific, Campbell and Collins (1975), using ages determined from otoliths, estimated the von Bertalanffy growth parameters to be $L_\infty = 76.87$ cm, $t_0 = -0.785$ yr, and $k = 0.6215$. Numbers-at-length for 1 cm intervals for ages I through V are shown in Figure 1 with the corresponding length-frequency plot in Figure 2. These numbers represent the 1973 catch from California waters and are a subset of the data used to estimate the von Bertalanffy parameters. This example serves to demonstrate bias and illustrate application of the stochastic method. If desired a variance-covariance matrix can be generated (Draper and Smith 1981) to estimate precision in the resulting age structure.
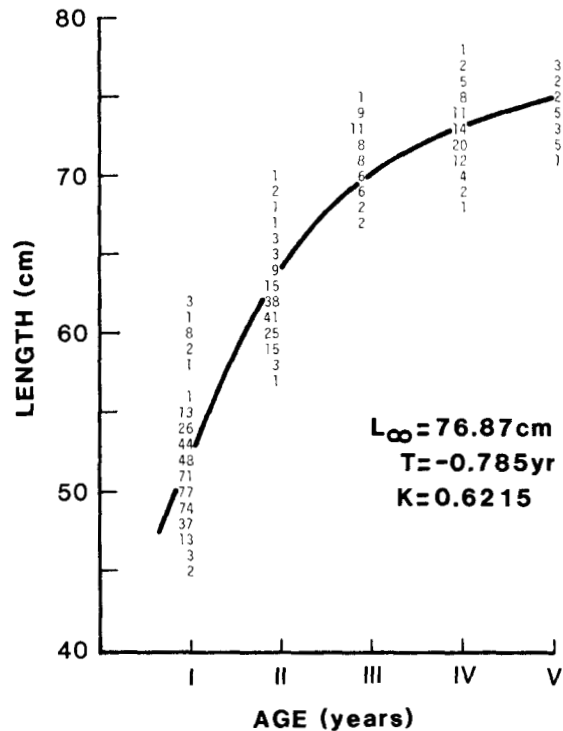

FIGURE 1.—Numbers-at-age by length in 1 cm intervals for the Pacific bonito, *Sarda chiliensis*. Data from 1973 California landings (Campbell and Collins 1975).

The length-frequency information and von Bertalanffy parameters are used to generate both deterministic and stochastic estimates of numbers-at-age. The estimated length-at-age
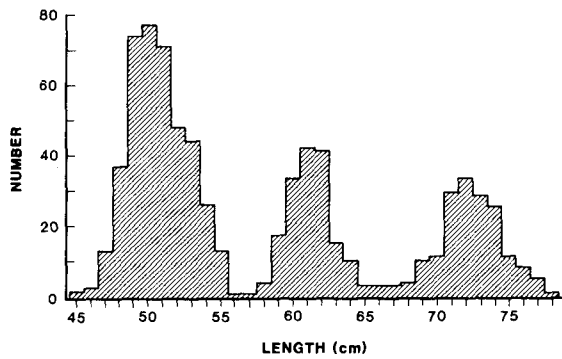
93

FIGURE 2.—Length frequency for the Pacific bonito, *Sarda chiliensis*, from 1973 California landings (Campbell and Collins 1975).

and sample standard deviations are

Age	Length	Standard deviation
I	51.5 cm	2.7 cm
II	63.3 cm	2.1 cm
III	69.5 cm	2.3 cm
IV	72.9 cm	2.0 cm
V	74.8 cm	1.9 cm

Deterministic estimates were made on lengths rounded to the nearest 0.1 cm. From Equation (2) the deterministic numbers-at-age are shown in column 3 of Table 2 with the difference between the true and estimated numbers in column 4. While the estimates are reasonably close over the first two ages, they become increasingly disparate for older ages. Thirteen fish had lengths greater than those at the maximum age. Seven had lengths greater than L and consequently were unclassifiable.

Quarter centimeter intervals were used to compute the stochastic estimates of age frequency. The results are shown in column 6 of Table 2,

with the difference between true and estimated numbers in column 7. For all ages, especially the older ages, the stochastic estimates are closer and less biased than those of the deterministic method (column 3).

Some insight into the improvement of the stochastic age estimates over the deterministic age estimates can be gained by inspection of Figure 1. Lengths of age I fish overlap those of age II fish and vice versa. Since the deterministic cutoff point for age I fish is 58.7 cm (1.5 yr), all overlap is lost in the deterministic model. In contrast, for the stochastic model, overlaps in lengths-at-age are shared between ages, the degree of sharing being relative to the probabilities of length intervals at the respective ages.

With increasing age, the extent of relative overlap and, consequently, misaging increases for the deterministic model; allocation of lengths to ages becomes more sensitive. Only if the degree of overlap between adjacent ages is equal do accurate estimates of numbers-at-age result from the deterministic model. In the present example varying year-class strength and random variability in lengths-at-age offset this sensitive compensatory mechanism needed for accurate estimation with the deterministic model.

Fish lengths above 75.3 cm, the length at age 5.5 yr, are misclassified either as older or of infinite ages for the deterministic model. Since for the stochastic model probabilities of length intervals at age exist for all ages and lengths, even for lengths above L_∞, fish at lengths above the 75.3 cm cutoff point are distributed to all ages relative to their respective probabilities for length intervals.

DISCUSSION

Calculations of age from length via the von Bertalanffy growth equation result in several

TABLE 2.—Deterministic (col. 3) and stochastic (col. 6) estimates of numbers-at-age with their respective differences from the true numbers-at-age in columns 4 and 7 for the Pacific bonito, *Sarda chiliensis*, from 1973 California landings (Campbell and Collins 1975).

Age (1)	True (2)	Deterministic (3)	Difference (4)	% Difference (5)	Stochastic (6)	Difference (7)	% Difference (8)
I	424	411	13	3.1	415	9	2.1
II	158	167	−9	−5.7	162	−4	−2.5
III	54	39	15	27.8	49	5	9.3
IV	80	71	9	11.3	85	−5	−6.3
V	21	29	−8	−38.1	26	−5	−23.8
>V	—	13	−13	Inf.	—	—	0.0
Inf.	—	7	−7	Inf.	—	—	0.0

types of bias. The degree of bias is proportional to overlap in lengths-at-age and changes with weak or strong year classes. When overlap increases with age, age-frequency estimates will generally be more biased for older than younger ages. When overlap occurs, biases will always result, since the numbers-at-length will be allocated to unreasonably old ages. Any numbers-at-length for lengths greater than L_∞ will be undetermined in age estimation, resulting in downward biases for those ages contributing such lengths.

Age estimation biases can be effectively removed by creating a stochastic model based on a matrix of length interval probabilities at age. The probability matrix (P-matrix) is independent of year-class strength and will effectively remove all sources of estimation bias except that due to random variation in length-frequency estimation. As long as the von Bertalanffy growth parameters remain the same over time, the stochastic method based on accurate estimates of variance in length-at-age will always yield unbiased results.

A probability model of the distribution of length-at-age with estimated parameters is necessary for estimating probabilities of length intervals at age for the P-matrix. If age information is unavailable then variances can be estimated from visually separable length-frequency modes. In the case where modes are separable for the first few ages only, there will be a problem in estimating variances for older ages: A model relating the variance in length-at-age with age can be used in estimating variances for these older ages. Ricker (1969) proposed that while distributions in lengths-at-age remain normal, variances increase during the first few years, stabilize, and then decrease over the final years. The trend in variances with age for a similar species might also be substituted in cases where variances are unavailable.

The principal strengths of the stochastic method are that few fish are required to be aged to estimate the P-matrix and that existing von Bertalanffy growth relations can be used. Accurate estimates of variance in length-at-age can probably be achieved with as few as 20 to 30 fish/age, which is likely to be a much smaller number of fish than needed to estimate accurate proportions of age-at-length necessary to construct an age-length key.

Von Bertalanffy growth parameters have been estimated for many species. Since most stocks have variant year-class strength, overlaps in lengths-at-age, and lengths exceeding the upper bound for the last age attainable, conversion to a stochastic model may be necessary, if unbiased estimates of age frequency are desired. Reexamination of age-length data used to estimate the von Bertalanffy parameters may be useful in estimating variances in lengths-at-age for the P-matrix. Taking additional age-length samples may be a cost-effective way of improving age-frequency estimation.

In fishery management, the overestimation of maximum age by the deterministic von Bertalanffy equation may produce underestimates of mortality rates which may result in overestimates of population size and recruitment. Further, the deterministic method tends to "fill in" weak year classes which results in underestimates of year-class variability and overestimates of recruitment stability. In general, all of these affect accuracy of a stock assessment and contribute to improper advice for fishery management.

Application of the stochastic method shown here to cover other growth equations and situations, such as discontinuous growth, is handled by simply estimating appropriate elements in the P-matrix for each case.

ACKNOWLEDGMENTS

LITERATURE CITED

BRODY, S.
 1945. Bioenergetics and growth. Reinhold Publ. N.Y., 1023 p.
CAMPBELL, G., AND R. A. COLLINS.
 1975. The age and growth of the Pacific bonito, *Sarda chiliensis*, in the eastern North Pacific. Calif. Fish Game 61:181-200.
CLARK, W. G.
 1981. Restricted-least squares estimate of age composition from length composition. Can. J. Fish. Aquat. Sci. 38:297-307.
DRAPER, N. R., AND H. SMITH.
 1981. Applied regression analysis. 2d ed. John Wiley and Sons, Inc., N.Y., 709 p.
GULLAND, J. A.
 1973. Manual of methods for fish stock assessment. Part

1. Fish population analysis. FAO Man. Fish. Sci. 4, 154 p.

KIMURA, D. K.
1977. Statistical assessment of the age-length key. J. Fish. Res. Board Can. 34:317-324.

LENARZ, W. H., W. W. FOX, JR., G. T. SAKAGAWA, AND B. J. ROTHSCHILD.
1974. An examination of the yield per recruit basis for a minimum size regulation for the Atlantic yellowfin tuna, *Thunnus albacares*. Fish. Bull., U.S. 72:37-61.

MURPHY, G. I.
1965. A solution of the catch equation. J. Fish. Res. Board Can. 22:191-202.

RICKER, W. E.
1958. Handbook of computations for biological statistics of fish populations. Fish. Res. Board Can. Bull. 119, 300 p.
1969. Effect of size-selective mortality and sampling bias on estimates of growth, mortality, production, and yield. J. Fish. Res. Board Can. 26:479-541.

WESTRHEIM, S. J., AND W. E. RICKER.
1978. Bias in using an age-length key to estimate age-frequency distributions. J. Fish. Res. Board Can. 36:184-189.