An objective method for determining fish growth from length-frequency data

DANIEL PAULY, ICLARM, and NOEL DAVID, ICLARM Affiliate Scientist

To the fishery biologist working on stock assessment, knowledge of how the fish in a given population grow is essential. Usually, the available growth information is reduced to and expressed by means of a simple equation, such as the von Bertalanffy Growth Formula (VBGF), which has, in its simplest form, two major parameters (L_∞, K) and one minor parameter (t_0).

The biological data from which growth parameters can be estimated are of three general types:

- periodic markings (annual or daily) on skeletal parts—scales, otoliths or other bones
- tagging-recapture data
- size-frequency data, most commonly length-frequency data, such as shown in Figs. 1A and 2.

Despite frequent criticism, methods using analysis of length-frequency data have found wider application than skeletal and tagging studies - at least in the case of those relatively small, short-lived fishes which so greatly contribute to both demersal and pelagic tropical fisheries. The reason for this is that length-frequency analyses are easier and cheaper, since less equipment is required.

The methods presently in use for the analysis of length-frequency data find their origin in the work of Petersen who in 1892 pioneered the two commonly applied "paper-and-pencil methods."

The first, named the "Petersen Method", essentially consists of attributing approximate ages to the various "peaks" of a single length-frequency sample (Fig. 1A). The second, generally called "Modal Class Progression Analysis," consists of following the progression, along the length axis, of the peaks in a series of length-frequency samples sequentially arranged in time (see Fig. 2). These two methods have greatly contributed to our understanding of the growth of tropical fishes.

It is, however, also in the tropics that these methods have often been found to generate questionable results. The reasons are obvious: the spawning seasons of tropical fishes are often quite long, and/or spawning may occur in several batches, each later resulting in a peak in the length-frequency distribution of the population. (This phenomenon was briefly discussed in a letter by Goldman in the April 1980 ICLARM Newsletter). Interpretation and interconnection of peaks thus become fraught with uncertainty.

Actually, this is not even the worst shortcoming of these methods. An inherent feature of both is that results obtained by different authors from the same data set generally differ, often to a large extent, because the methods are essentially based on *subjective* interpretation.

That subjectivity, more than anything else, has rendered these methods somewhat suspect.

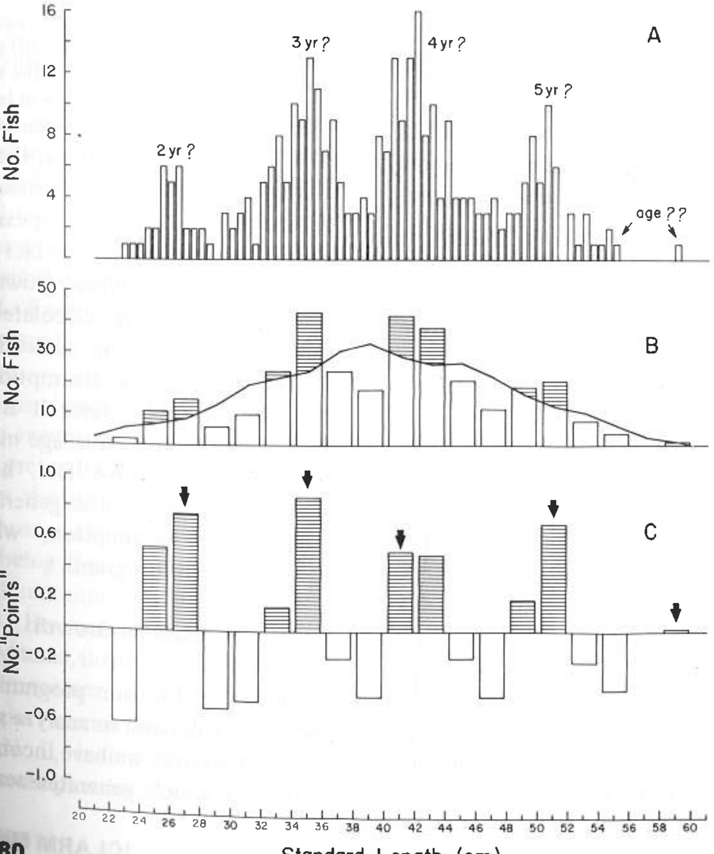

Fig. 1. Length-frequency data on coral trout *(Plectropomus leopardus)* caught near Heron Island (Great Barrier Reef, Australia) in October 1971. From Goeden, G.B. 1978. Queensland Fisheries Service, Research Bulletin 1. 42 p.

A. Original data; the ages are from Goeden, with questions marks added. N=319. Note small class interval (5 mm).

B. Same data, replotted in 2 cm class intervals to smooth out small irregularities, showing running average frequencies (over 5 length classes) to emphasize peaks (striped bars above running averages) and intervening troughs.

C. Same data as in B, after division of each frequency value by the corresponding running average frequency, substraction of 1 from each of the resulting quotients and subsequent minor adjustments to remove potential sources of bias. Note that "peaks" have been alloted similar numbers of "points," irrespective of the number of fish they represent. Arrows show the "points" used in the computation of USP (see text).

The New Approach

The task we undertook was to devise a computing procedure that would "trace," through a series of length-frequency samples sequentially arranged in time, a multitude of growth curves, and select the single curve which, by passing through a maximum of peaks, would best "explain" these peaks. The method would have to be wholly *objective* in the sense that any researcher using the program should arrive, for each data set, at exactly the same results. Thus, the solution would have to be based exclusively on the length-frequency data themselves, and require no additional (necessarily subjective) inputs, such as the assumed number of age groups represented in each sample (as required, e.g., by the NORMSEP program, still often used for dissecting length-frequency samples).

We have succeeded in this, the result being a computer program called ELEFAN I (Electronic LEngth Frequency ANalysis). It is written in BASIC and can be run on most of the cheap microcomputers now available (we used a Radio-Shack TSR-80, Level II, 16K).

Put anthromorphically, for any (set of) length-frequency sample(s), the program:

- restructures the sample(s) that have been entered, such that even small but clearly identifiable peaks are attributed a number of "points" similar to those allocated to peaks based on larger numbers of fishes (see Figs. 1B and 1C)
- calculates the sum of points "available" (see Fig. 1C). This sum is termed "unexplained sum of peaks" (USP) for reasons which should become obvious below
- "traces" a series of growth curves started from the length value corresponding to the base of each peak, for any arbitrary "seed" input of L_∞ and K, projected backward and forward in time to meet all other samples of a chronologically ordered sample set (Fig. 2), and/or the same sample repeated again and again (Fig. 3)
- summates the "points" obtained by each growth curve when passing through peaks (positive points) or troughs separating peaks (negative points) (see Figs. 1C and 3)
- selects the curve which, by passing through most peaks and avoiding most troughs, best "explains" the peaks by scoring the most points. This sum is called "explained sum of peaks" (ESP)
- decrements or increments the "seeded" values of L_∞ and K until the ratio ESP/USP reaches a maximum and outputs the corresponding growth parameters.

The validity of this procedure rests on the following assumptions:

1. That the sample(s) used represent the population investigated.
2. That the growth patterns in the population are the same from year to year.
3. That the VBGF describes the average growth of the investigated population.
4. That all fishes in the (set of) sample(s) have the same length at the same age, such that all differences in length can be attributed to differences in age.

Of these 4 assumptions, the first is trivial and need not be discussed here. Assumptions 2 and 3 appear to be realistic, and they are in fact made—explicitly or not—when growth parameters of fish are calculated from annual markings on skeletal parts.

The last of these assumptions does not strictly apply, since it is known that fishes of the same age may have different length. We feel, however, that no strong bias is generated by making this assumption, which is essential to our program.

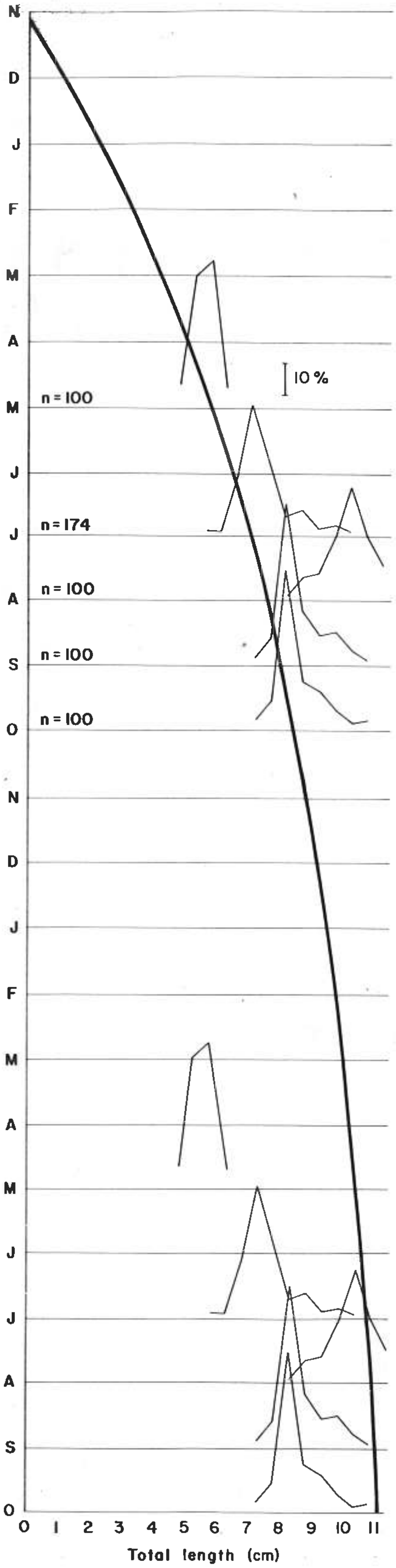

Fig. 2. A set of length-frequency samples arranged sequentially in time, with growth curve fitted by ELEFAN I. Note that the distance between the bases of the samples and the time period between the sampling dates are proportional, and that the set of samples is "repeated" one year later, to allow for the forward projection of the growth curve. The curve has the parameters L_∞ = 12.2 cm and K = 1.3, with ESP/USP = 0.804. It must be emphasized that the curve was *not* fitted by eye, and that *no* inputs were made as to expected ages of the various peaks, which of the peaks should be interconnected, etc., (see text). The data, which pertain to slipmouths ***(Leiognathus bindus)*** caught off Calicut, India, in 1958, were originally published by Balan, V. 1967. Ind. J. Fish. 10(1): 118-134.

Seasonal Growth

In order for our program to be more versatile, and to analyze seasonal growth patterns, we have incorporated a routine which generates seasonally

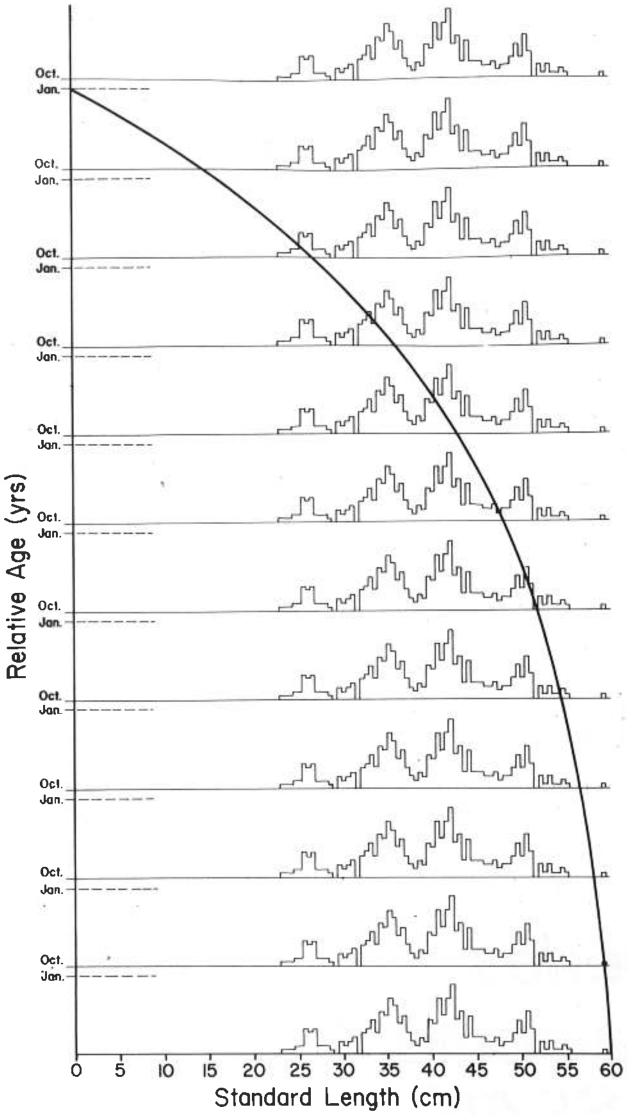

Fig. 3. Length-frequency data on coral trout, fitted with a growth curve by means of ELEFAN I. Note that it is the original sample of Fig. 1A which is shown here, but that the optimization performed by ELEFAN I was based on the "restructured" sample of Fig. 1C. The growth curve has the parameters L_∞ = 62.4 cm and K = 0.31, with ESP/USP = 0.942. It is again emphasized that the curve was traced without any input except for the length-frequency data themselves. The curve provides an interpretation of the age structure of the sample different from that originally presented by Goeden. Particularly, what was identified as age group 5 (see Fig. 1A) appears to be age group 6, while the longevity of the fish appears quite higher than originally assumed.

oscillating growth curves (not shown here). Two additional parameters are included for expressing the timing and intensity of the growth oscillation. The first is called Winter Point (WP) and refers to the *time* of the year when growth is slowest; the second is a constant (C) which expresses the *intensity* of the growth oscillation and which can take values ranging from zero (in tropical fishes) to unity (in temperate fishes).

Incorporation of seasonal growth in our program thus results in an optimization procedure involving not only the parameters L_∞ and K, but also WP and C.

While searching for the optimal combination of the two parameters, L_∞ and K, is a relatively straightforward job, searching for the optimal combination of four parameters is quite another matter. In fact, the amount of computation involved with larger sample sets can become elephantine.

This is compounded by the fact that the execution of programs written in interpreter BASIC is relatively slow, and that the optimization procedure is partly human-aided, the result of these things being that running ELEFAN I on a microcomputer can become quite tedious. However, with time-sharing, larger systems, the time-problem should be less important.

A report containing more details on ELEFAN I, including several computed examples and a full program listing will be made available soon. We hope that this program will eventually become widely used, both to determine growth parameters from newly sampled or already published length-frequency data, and to reassess the validity of earlier growth estimates using paper-and-pencil methods.

Future Work

The interpretation of length-frequency samples performed by ELEFAN I, should, when used in conjunction with catch information, lead to data amenable to subsequent "cohort analysis."

Cohort analysis is the very powerful method in which numbers of fish caught are used to obtain rather reliable estimates of fishing mortality and of the sizes of (past) populations.

Following an invitation by Dr. D. Cushing and Mr. J. Pope, one of us (Pauly) will spend 3 weeks in October at the Fisheries Laboratory in Lowestoft, England, to investigate, with J. Pope, the possibility of making ELEFAN I the basis of a program package for the overall analysis, including cohort analysis, of length-frequency data in tropical and other fishes.

Further Reading

Pauly, D. and G. Gaschütz. 1979. A simple method for fitting oscillating length growth data, with a program for pocket calculators. ICES. C.M. 1979/G:24. Demersal Fish Cttee., 26 p.

Pope, J.G. 1972. An investigation into the accuracy of virtual population analysis using cohort analysis. Int. Comm. Northwest. Atl. Fish. Res. Bull. 9:65-74.