

3. ESTIMATION OF GROWTH PARAMETERS

The study of growth means basically the determination of the body size as a function of age. Therefore all stock assessment methods work essentially with age composition data. In temperate waters such data can usually be obtained through the counting of year rings on hard parts such as scales and otoliths. These rings are formed due to strong fluctuations in environmental conditions from summer to winter and vice versa. In tropical areas such drastic changes do not occur and it is therefore very difficult, if not impossible to use this kind of seasonal rings for age determination.

Only recently methods have been developed to use much finer structures, so-called daily rings, to count the age of the fish in number of days. These methods, however, require special expensive equipment and a lot of manpower, and it is therefore not likely that they will be applied on a routine basis in many places.

Fortunately several numerical methods have been developed which allow the conversion of length-frequency data into age composition. Although these methods do not require the reading of rings on hard parts, the final interpretation of the results becomes much more reliable if at least some direct age readings are available. The best compromise for stock assessment of tropical species is therefore an analysis of a large number of length-frequency data combined with a small number of age readings on the basis of daily rings. This manual does not deal with the techniques of age reading but references to special publications are given (see Section 3.2.1).

3.1 THE VON BERTALANFFY GROWTH EQUATION

Pütter (1920) developed a growth model which can be considered the base of most other models on growth including the one developed as a mathematical model for individual growth by von Bertalanffy (1934), and which has been shown to conform to the observed growth of most fish species. The theory behind various growth models is reviewed by for example Beverton and Holt (1957), Ursin (1968), Ricker (1975), Gulland (1983), Pauly (1984) and Pauly and Morgan (1987), but we shall deal here only with the von Bertalanffy growth model of body length as a function of age. It has become one of the cornerstones in fishery biology because it is used as a sub-model in more complex models describing the dynamics of fish populations. Fig. 3.1.0.1 illustrates the model in graphical as well as in mathematical form.

The mathematical model, B, expresses the length, L , as a function of the age of the fish, t :

$$L(t) = L_{\infty} * [1 - \exp(-K*(t-t_0))] \quad (3.1.0.1)$$

The right hand side of the equation contains the age, t , and some parameters. They are: " L_{∞} " (read "*L-infinity*"), " K " and " t_0 " (read "*t-zero*"). Different growth curves will be created for each different set of parameters, therefore it is possible to use the same basic model to describe the growth of different species simply by using a special set of parameters for each species.

To illustrate the use of the model, assume that the three parameters have been estimated for some particular fish stock and that the values are:

$$L_{\infty} = 50 \text{ cm, } K = 0.5 \text{ per year and } t_0 = -0.2 \text{ year}$$

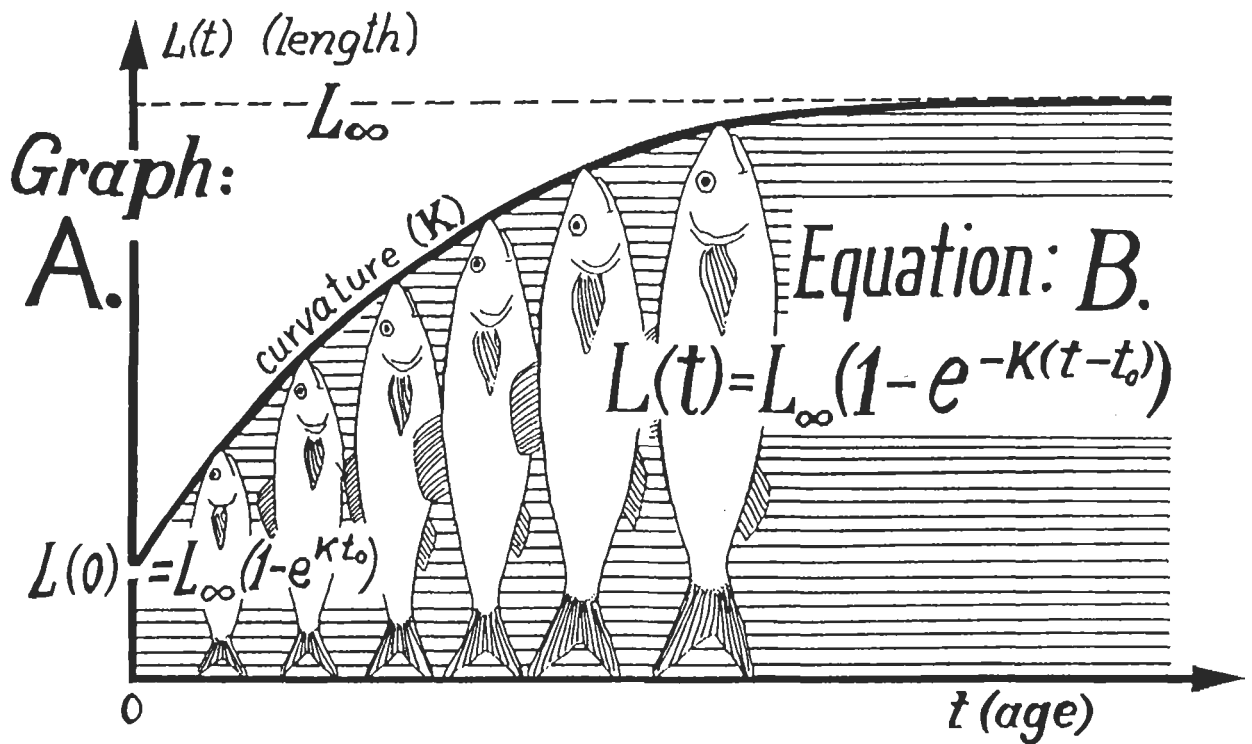


Fig. 3.1.0.1 The von Bertalanffy growth equation

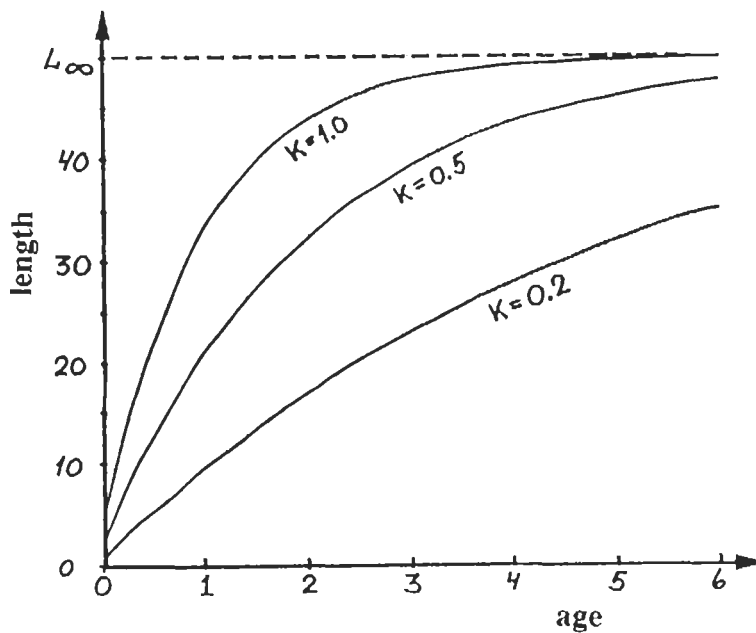


Fig. 3.1.0.2 A family of growth curves with different curvature parameters, different K values

Then we insert these parameter values into the von Bertalanffy growth equation (Eq. 3.1.0.1):

$$L(t) = 50 * [1 - \exp(-0.5 * (t + 0.2))]]$$

The length in cm at a given age of an average fish of the stock in question can now be calculated by inserting a value for t, the age, e.g. t = 2 years:

$$L(2) = 50 * [1 - \exp(-0.5 * (2 + 0.2))]] = 33.4 \text{ cm}$$

Thus, knowing the parameters we can calculate the length at any age of the fish in the stock in question:

age of fish (year)	body length of fish (cm)
0.5	14.8
1.0	22.6
1.5	28.6
2.0	33.4
3.0	39.9
5.0	46.3
..... etc.	

From such a table a graph ("growth curve") can be produced for this set of parameters, as in Fig. 3.1.0.1.

The parameters can to some extent be interpreted biologically. L_{∞} is interpreted as "the mean length of very old (strictly: infinitely old) fish", it is also called the "asymptotic length" (see Fig. 3.1.0.1). K is a "curvature parameter" which determines how fast the fish approaches its L_{∞} . Some species, most of them short-lived, almost reach their L_{∞} in a year or two and have a high value of K . Other species have a flat growth curve with a low K -value and need many years to reach anything like their L_{∞} . This is illustrated in Fig. 3.1.0.2. The third parameter, t_0 , sometimes called "the initial condition parameter", determines the point in time when the fish has zero length. Biologically, this has no meaning, because the growth begins at hatching when the larva already has a certain length, which may be called $L(0)$ when we put $t = 0$ at the day of birth. It is easily identified by inserting $t = 0$ into Eq. 3.1.0.1:

$$L(0) = L_{\infty} * (1 - \exp(K * t_0))$$

However, $L(0)$ may not be a realistic estimate of the length at birth because fish larvae do not always grow according to the von Bertalanffy model. The important point is that fish old enough to be exploited usually do. Let us therefore turn the attention to the description of the growth of larger (exploited) fish.

Fish increase in length as they grow older, but their "growth rate", that is the increment in length per unit time, decreases when they get older, approaching zero when they become very old. The growth rate can be defined by:

$$\frac{\Delta L}{\Delta t} = \frac{L(t+\Delta t) - L(t)}{\Delta t} \quad (3.1.0.2)$$

Time (or age), t , is usually expressed in units of years. If the growth rate is measured per month then

$$\Delta t = 1/12 \text{ years} = 0.0833 \text{ years}$$

or per day then,

$$\Delta t = 1/365 \text{ years} = 0.00272 \text{ years.}$$

In Table 3.1.0.1 the ages (in years) and the lengths at the beginning of each year (in cm) corresponding to the example given in Fig. 3.1.0.1 are given in columns A and B, respectively. The growth rate is given in column C. It is evident that the growth rate

decreases as the fish get older. The mathematical relationship between the length of a fish and the growth rate at a given time is a linear function:

$$\frac{\Delta L}{\Delta t} = a + b \cdot L(t) \quad (3.1.0.3)$$

This linear relationship can be derived from the von Bertalanffy growth equation, as follows:

$$\frac{\Delta L}{\Delta t} = K \cdot (L_{\infty} - L(t)) \quad \text{cm/year} \quad (3.1.0.4)$$

where $K = -b$ and $L_{\infty} = -a/b$

We shall not concern ourselves here with the mathematical proof. This linear relationship will be used in subsequent sections to determine the growth parameters K and L_{∞} . An example is already given in Fig. 3.1.0.3 where the growth rate $\Delta L/\Delta t$, as dependent variable, is plotted against the mean length, $L(t)$, over the corresponding year, as the independent variable (see column D of Table 3.1.0.1):

$$\bar{L}(t) = \frac{L(t+\Delta t) + L(t)}{2}$$

From Eq. 3.1.0.4 it follows that if $\bar{L}(t) = L_{\infty}$ then $\Delta L/\Delta t = K \cdot (L_{\infty} - L_{\infty}) = 0$, that is to say when the fish reaches length L_{∞} the growth rate becomes zero and L_{∞} is thus the maximum average length of a fish. This is also illustrated in Fig. 3.1.0.3. Where the regression line reaches the x-axis, $\Delta L/\Delta t = 0$ and the corresponding $L(t)$ at the axis = L_{∞} . Further, K , can be derived from the slope (see Section 3.3.1).

Table 3.1.0.1 Growth rate as a function of age corresponding to the growth curve in Fig. 3.1.0.1. See also Fig. 3.1.0.3

A	B	C	D
age	length	growth rate	mean length
t	L(t)	$\frac{L(t+\Delta t) - L(t)}{1} = \frac{\Delta L}{\Delta t}$	$\frac{L(t+\Delta t) + L(t)}{2} = \bar{L}(t)$
years	cm	cm/year (y)	cm (x)
1	25.7	10.3	30.9
2	36.0	6.9	39.5
3	42.9	4.6	45.2
4	47.5	3.2	49.1
5	50.7	2.1	51.8
6	52.8	1.4	53.5
7	54.2		

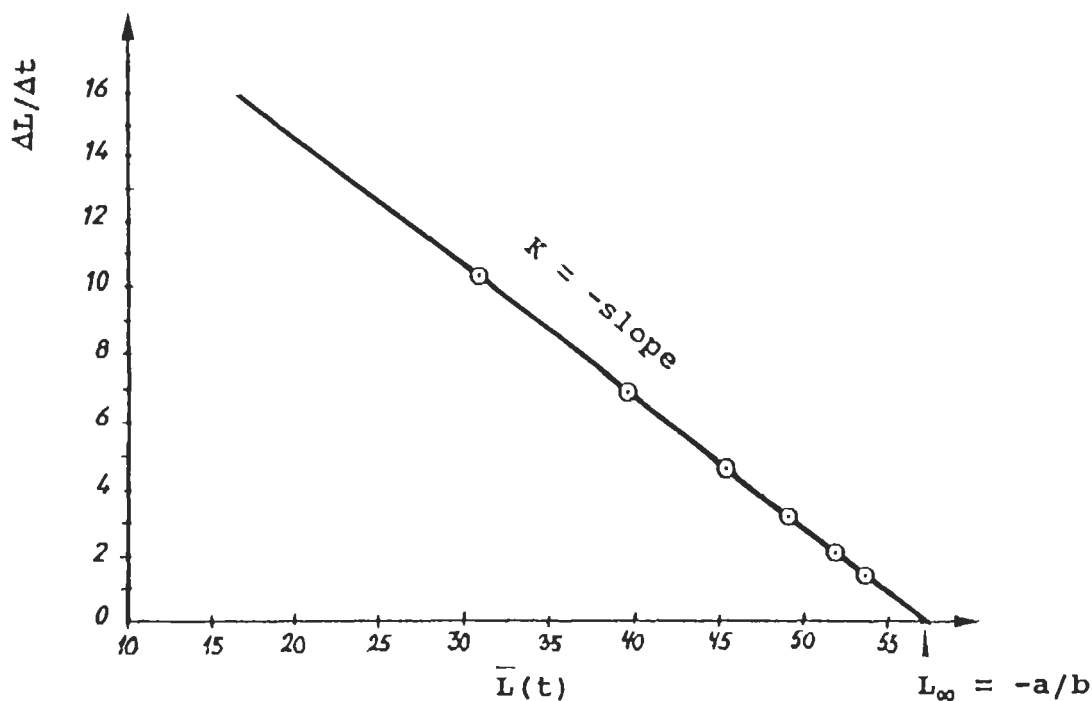


Fig. 3.1.0.3 Plot of growth rate against mean length. From columns C and D of Table 3.1.0.1

3.1.1 Variability and applicability of growth parameters

Growth parameters, of course, differ from species to species, but they may also vary from stock to stock within the same species, i.e. growth parameters of a particular species may take different values in different parts of its range. Also successive cohorts may grow differently depending on environmental conditions. Further growth parameters often take different values for the two sexes. If there are pronounced differences between the sexes in their growth parameters, the input data should be separated by sex and values of K , L_{∞} and t_0 should be estimated for each sex separately.

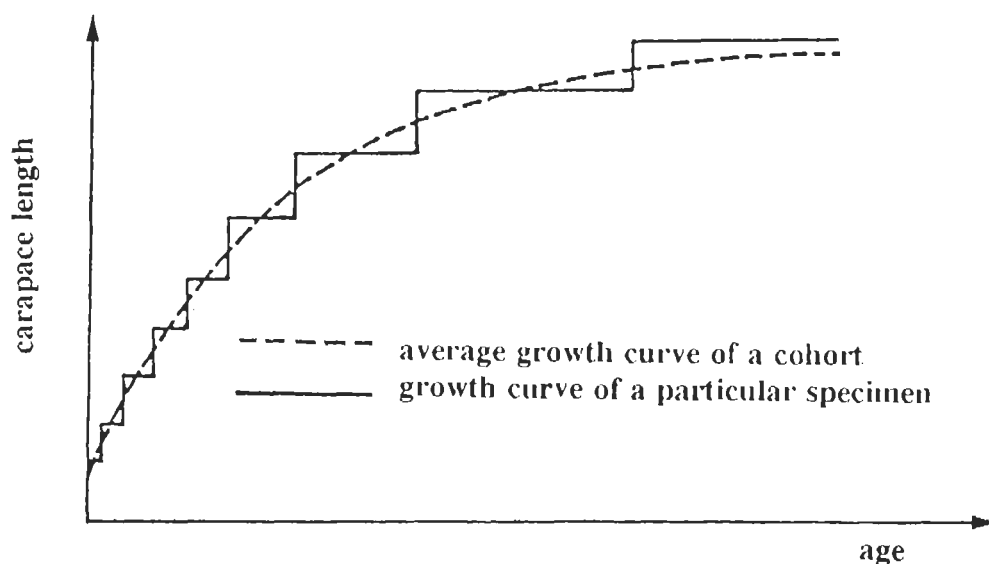


Fig. 3.1.1.1 Individual growth curve and average cohort growth curve of crustaceans

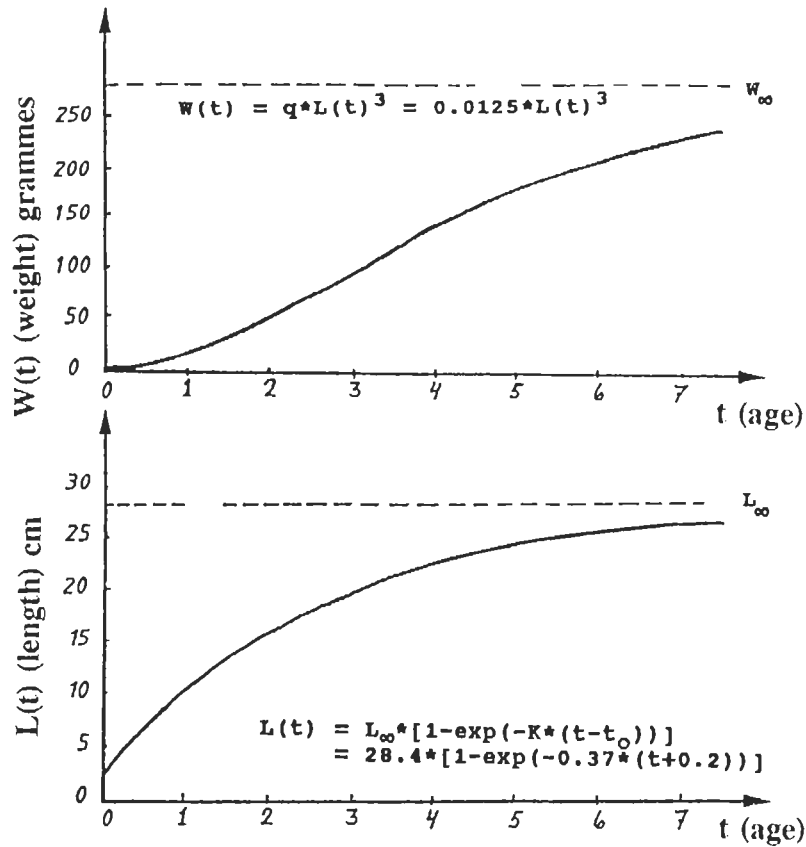


Fig. 3.1.2.1 A length-based growth curve and corresponding weight-based growth curve

Although the physiology of crustaceans is very different from that of fishes, their average body growth appears also to conform to the von Bertalanffy growth model (see Garcia and Le Reste, 1981). An individual crustacean (a shrimp or lobster) does not conform to the von Bertalanffy model, but to some "stepwise curve", with each step accounting for a moult (as illustrated in Fig. 3.1.1.1). However, members of a cohort moult at different times, and therefore the average growth curve of a cohort of crustaceans becomes a smooth curve (dotted line). For further discussion on the modelling of population dynamics of crustaceans see, for example, Jamieson and Bourne (1986) and Caddy (1987).

3.1.2 The weight-based von Bertalanffy growth equation

Combining the von Bertalanffy growth equation (Eq. 3.1.0.1)

$$L(t) = L_\infty \cdot [1 - \exp(-K \cdot (t - t_0))]$$

with the length/weight relationship (Eq. 2.6.1):

$$W(t) = q \cdot L^3(t)$$

gives the weight of a fish as a function of age:

$$W(t) = q \cdot L_\infty^3 \cdot [1 - \exp(-K \cdot (t - t_0))]^3$$

The "asymptotic weight", W_∞ , corresponding to the asymptotic length is (according to Eq. 2.6.1):

$$W_\infty = q \cdot L_\infty^3$$

The parameter, q , is called the "*condition factor*". (Note that the letter q is also used in this manual to designate the catchability coefficient, Section 4.3.) Thus, "*the weight-based von Bertalanffy equation*" can be written:

$$w(t) = w_{\infty} * [1 - \exp(-K*(t-t_0))]^3 \quad (3.1.2.1)$$

Fig. 3.1.2.1, for example, shows the weight-based growth curve for the von Bertalanffy parameters: $L_{\infty} = 28.4$ cm, $K = 0.37$ per year, $t_0 = -0.2$ years and the condition factor $q = 0.0125$ g per cubic cm, of the threadfin bream, *Nemipterus marginatus* in North Borneo waters (Pauly, 1983).

(See Exercise(s) in Part 2.)

3.2 INPUT DATA FOR THE VON BERTALANFFY GROWTH EQUATION

There are several ways of obtaining input data for the methods used to derive the growth parameters L_{∞} , K and t_0 . The methods may be categorized roughly into three groups:

- 1) Age reading and length measurements combined
 - a) data from resource surveys with a research vessel
 - b) data from samples taken from commercial catches
- 2) Length measurements only
 - a) data from resource surveys with a research vessel
 - b) data from samples taken from commercial catches
- 3) Mark-recapture (tagging) experiments, where two (or more) length measurements are obtained, *viz.* at the time of marking (usually on a research vessel) and the time of recovery (usually by the commercial fishery). This method is excellent from a theoretical point of view but difficult and costly to implement. We shall not go further into this method, except for Exercise 3.3.1. (See also Jones, 1977).

Below we shall first consider 1a) in Section 3.2.1 and then 2b) in Section 3.2.2.

3.2.1 Data from age readings and length measurements

As has been stated in the introduction to this chapter, age reading is a relatively simple technique in the case of species from temperate waters, because their otoliths or scales show seasonal rings, one for the summer and one for the winter, which together form an annual ring. Sometimes such rings can be observed with the naked eye. In other cases simple techniques, such as burning can make them visible. The annual rings give sufficient information for most stock assessment purposes.

Unfortunately, tropical fish species seldom show clear annual rings in their otoliths or scales, because the strong seasonality which characterizes temperate zones is lacking. Recent discoveries, however, have created opportunities to also read ages of tropical fish, albeit

within limited ranges and at a high cost in terms of manpower and initial investment. By going deeper into the formation of the rings in otoliths and scales it has been discovered that daily increments (or even increments caused by a certain food intake) can be detected by means of a strong microscope. The latest findings indicate that sometimes the daily rings are so thin that they defeat the ordinary microscope whose detection power is limited by the wavelength of the light. Such rings can be read only by a scanning electron microscope (Morales-Nin, 1991).

A large amount of literature has been produced on this subject in recent years, for example: Panella (1971), Bagenal (1974), Brothers (1980), Beamish and Mc Farlane (1983), Gjøsæter *et al.* (1983), Dayaratne and Gjøsæter (1986) and Williams (1986).

In a manual on tropical fish stock assessment it is necessary to concentrate on length measurements and consequently place less emphasis on age data. However, we are dealing with age here for two reasons. In the first place it may sometimes be possible to carry out a small number of age readings, which can be used to calibrate the findings obtained from length measurements alone. Secondly, it is easier to explain the concepts and theory on the basis of age and length data than on the basis of length data only. Also we use data from research vessels to avoid further complications at this stage. The first example deals with data from a single survey, while the second example deals with data from a time series of surveys.

Example 3: Age/length composition data from a single survey

Suppose we have a random sample of fish from the stock of species A. This sample was taken on a survey with a duration of, say, a fortnight, in which trawl hauls were made over the entire distribution area of the stock in such a manner that the pooled data from all hauls made up a random sample (see Section 7.1). Suppose that the survey took place in October 1983 and that pooled length-frequency data were obtained as presented in the last column of Table 3.2.1.1 (and also in Fig. 3.2.2.1). Suppose also that we have observed two annual peak recruitment seasons and therefore have decided to define two cohorts per year:

Spring cohort: Fish recruited from January to June
Autumn cohort: Fish recruited from July to December

A cohort was defined earlier as "a group of fish all of the same age belonging to the same stock" (see Section 1.3.1).

Now, also suppose that we are able to read the age of each fish, so that we can determine the day on which it was born. After having read the ages of all 439 fish of species A caught in the October 1983 survey, we can assign each fish to a specific cohort. It is then possible to make a length-frequency distribution for each cohort. Theoretically, these length distributions are normal distributions for which we can determine the mean length and the standard deviation.

The complex length-frequency table obtained after the survey has thus been split into six length-frequency tables for different cohorts, of which we also know the average age. The type of information contained in the first seven columns of Table 3.2.1.1 forms a so-called "age/length key" (this concept will be further discussed in Example 7). The main data for each cohort have been summarized in Table 3.2.1.2.

If we further assume that all the six cohorts have the same growth parameters, we can use the data contained in Table 3.2.1.2 to estimate the common growth parameters. In other

words, we can determine the growth parameters which produce the growth curve that gives the best fit to the pairs of data of mean length and corresponding mean age. Exactly how this can be done is explained in subsequent sections.

Table 3.2.1.1 Age/length composition (hypothetical example). Basic data for Table 3.2.1.2. The graph for the total length-frequency is shown in Fig. 3.2.2.1 ("-" stands for a zero observation)

length interval cm	recruitment season						survey October 1983 total of all hauls
	spring 1983	autumn 1982	spring 1982	autumn 1981	spring 1981	autumn 1980	
12-13	1	-	-	-	-	-	1
13-14	4	-	-	-	-	-	4
14-15	11	-	-	-	-	-	11
15-16	24	-	-	-	-	-	24
16-17	38	-	-	-	-	-	38
17-18	42	-	-	-	-	-	42
18-19	33	-	-	-	-	-	33
19-20	20	-	-	-	-	-	20
20-21	7	-	-	-	-	-	7
21-22	2	1	-	-	-	-	3
22-23	-	3	-	-	-	-	3
23-24	-	5	-	-	-	-	5
24-25	-	8	-	-	-	-	8
25-26	-	11	-	-	-	-	11
26-27	-	14	-	-	-	-	14
27-28	-	16	1	-	-	-	17
28-29	-	15	1	-	-	-	16
29-30	-	13	2	-	-	-	15
30-31	-	11	3	-	-	-	14
31-32	-	7	4	-	-	-	11
32-33	-	4	6	1	-	-	11
33-34	-	2	7	1	-	-	10
34-35	-	1	7	1	-	-	9
35-36	-	-	8	2	-	-	10
36-37	-	-	7	3	1	-	11
37-38	-	-	6	3	1	-	10
38-39	-	-	5	4	1	-	10
39-40	-	-	4	4	2	1	11
40-41	-	-	3	5	2	1	11
41-42	-	-	2	4	2	1	9
42-43	-	-	1	3	2	1	7
43-44	-	-	-	3	3	1	7
44-45	-	-	-	2	2	1	5
45-46	-	-	-	2	2	2	6
46-47	-	-	-	1	2	2	5
47-48	-	-	-	1	1	1	3
48-49	-	-	-	-	1	1	2
49-50	-	-	-	-	1	1	2
50-51	-	-	-	-	1	1	2
51-52	-	-	-	-	-	1	1
total	182	111	67	40	24	15	439
mean length	17.3	27.9	35.3	40.2	43.3	45.5	
std. dev.	1.7	2.7	3.4	3.6	3.8	3.6	
mean age (y)	0.64	1.16	1.65	2.10	2.64	3.21	

Please note that the data presented in Table 3.2.1.1 are "hypothetical" or "faked" data. They were actually computed from a set of growth parameters (which determine the mean lengths for each cohort) and a set of standard deviations for the length distribution of each cohort. The mean age of the youngest cohort is 0.64 year or 234 days, this means that the birth date of this cohort was 234 days before 15 October 1983, or 23 February 1983 (northern spring). The other two spring cohorts were born one and two years earlier respectively, while the three autumn cohorts were born 6 months after each spring cohort. Due to random variations the birth dates vary slightly from year to year. The advantage of using such hypothetical data in the context of this manual is that the true parameters are known, which is not the case with data taken from a real stock. That puts us in a position to compare the results of various methods of parameter estimation with the real values. The data presented in Tables 3.2.1.1 and 3.2.1.2 will also be used as examples in Sections 3.4.1 and 3.4.2.

Table 3.2.1.2 Hypothetical example of age and length composition data of species A from one research survey in October 1983 (derived from the "raw data" in Table 3.2.1.1)

COHORT recruitment		number observed	mean age (year)	mean length (cm)
year	season			
1983	spring	182	0.64	17.3
1982	autumn	111	1.16	27.9
1982	spring	67	1.65	35.3
1981	autumn	40	2.10	40.2
1981	spring	24	2.64	43.3
1980	autumn	15	3.21	45.5
total:		439		

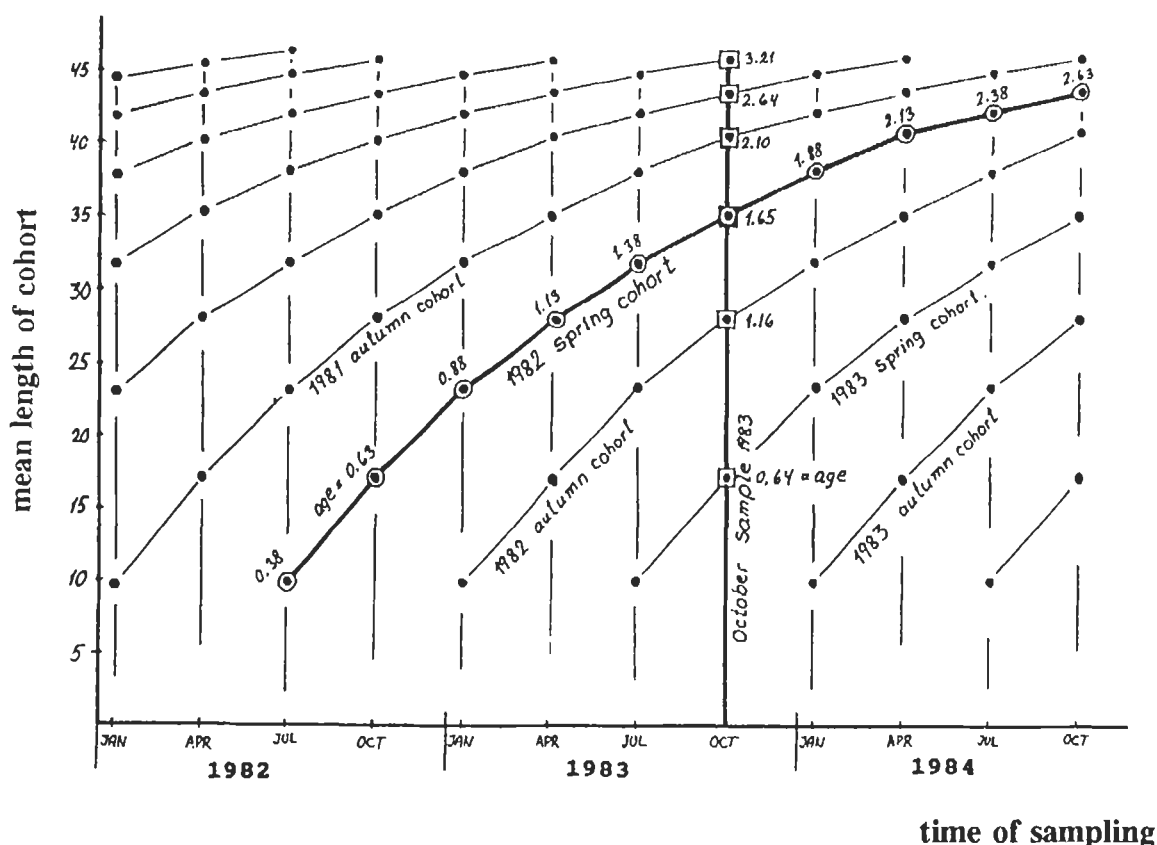


Fig. 3.2.1.1 Illustration of the data on age/length collected during a time series of surveys

Example 4: Age/length composition data from multiple surveys

If we now assume that the survey of Example 3 was only one of a series of 12 surveys carried out during the years 1982-1984, in the months of January, April, July and October each year, then such a survey programme would yield 12 tables like Table 3.2.1.1. By sampling the various cohorts regularly over a length of time, in this case three years, changes in the mean lengths can be determined by plotting them against the time of sampling as shown in Fig. 3.2.1.1. With this data set we are able to estimate the growth parameters for some of the cohorts individually. For the spring cohort in the recruitment year of 1982 for example, there are 10 pairs of age and length data which can be used to estimate the parameters for that particular cohort.

The difference between following one particular cohort in time as shown here, and the determination of mean lengths of different cohorts at a certain moment, as presented in Example 3, is illustrated in Fig. 3.2.1.1, where the two different data types are indicated by heavy lines. The curve starting July 1982 and running to October 1984 shows the "real" growth of a cohort. The vertical line "October sample 1983" shows a "cross section" of the stock at that date.

In case of a short-lived species (with a life span of, say, one to two years) we would have to follow a cohort in time as described in Example 4. The method based on a single sample would not be applicable because it would contain only one or two cohorts. Although there may be differences in the growth of different cohorts this difference is usually so small that it can be ignored. Data like those presented in Fig. 3.2.1.1 therefore could all be pooled into one data set and used in a way similar to the data of the October sample in 1983 (Table 3.2.1.1).

It is likely that bias will be less if sampling is done all the year round. Thus, although we can sometimes manage with a single sample, it is safer to use a time series of samples.

(See Exercise(s) in Part 2.)

Example 5: The use of age/length keys

An age/length key is a table showing, for each length class of fish of a particular stock, the percentage or fractional age-frequency distribution, see Table 3.2.1.4. Once such a key is available, samples of fish which were only measured for length can be distributed over age groups according to the key.

The age/length key of Table 3.2.1.4 could be based on 182 randomly drawn fish with the following length distribution:

Length group (cm)	5-10	10-15	15-20	20-25	Total
Frequency	110	40	22	10	182

The next step is to age the fish in each length group. Let us assume that the results are as shown in Table 3.2.1.3. Table 3.2.1.4 is then derived from Table 3.2.1.3 simply by dividing each row entry by the row total for each length group.

Table 3.2.1.4 can then be used to assign ages to a much larger length-frequency sample of the same stock (for which the age composition is unknown), for example the length-frequency sample of 21041 fish given below:

Length group (cm)	5-10	10-15	15-20	20-25	Total
Frequency	12088	7035	1788	130	21041

By distributing the numbers in each length group over the age groups according to the proportions given in Table 3.2.1.4, we get the results presented in Table 3.2.1.5. Length group 10-15 cm, for example, is estimated to consist of $7035 \times 0.25 = 1759$ 0-group fish and $7035 \times 0.75 = 5276$ 1-group fish. By summing the column entries we finally arrive at the age composition given in the bottom row of Table 3.2.1.5.

Table 3.2.1.3 Input data for estimation of an age/length key (hypothetical example)

length group cm	age group 0	age group 1	age group 2	total
5-10	110	0	0	110
10-15	10	30	0	40
15-20	0	11	11	22
20-25	0	1	9	10
total	120	42	20	182

Table 3.2.1.4 Hypothetical age/length key

length group cm	age group 0	age group 1	age group 2
5-10	1.0	0	0
10-15	0.25	0.75	0
15-20	0	0.5	0.5
20-25	0	0.1	0.9

Table 3.2.1.5 Age composition of a large length-frequency sample estimated by use of the age/length key in Table 3.2.1.4

length group cm	age group 0	age group 1	age group 2	total
5-10	12088	0	0	12088
10-15	1759	5276	0	7035
15-20	0	894	894	1788
20-25	0	13	117	130
total	13847	6183	1011	21041

Thus in order to estimate the age composition of the catch from a particular stock we only need to determine an age/length key based on a small sample of age readings and then we can restrict further sampling to the collection of length-frequency data. These lengths are converted to ages by means of the key. The same key may be used in consecutive years as long as there is no suspicion of major changes in the age composition of the stock. In a

Table 3.2.1.6 Age/length key for *Scomberomorus brasiliensis* based on otolith readings, in percentages per 3-cm length groups. From Sturm, 1974

length group cm	age group (year)									
	0	1	2	3	4	5	6	7	8	9
13-16	100	-	-	-	-	-	-	-	-	-
16-19	-	100	-	-	-	-	-	-	-	-
19-22	-	100	-	-	-	-	-	-	-	-
22-25	-	100	-	-	-	-	-	-	-	-
25-28	-	96	4	-	-	-	-	-	-	-
28-31	-	55	45	-	-	-	-	-	-	-
31-34	-	5	95	-	-	-	-	-	-	-
34-37	-	-	91	9	-	-	-	-	-	-
37-40	-	-	73	27	-	-	-	-	-	-
40-43	-	-	33	63	2	-	-	-	-	-
43-46	-	-	15	77	8	-	-	-	-	-
46-49	-	-	5	65	29	-	-	-	-	-
49-52	-	-	1	47	50	2	-	-	-	-
52-55	-	-	-	38	51	11	-	-	-	-
55-58	-	-	-	10	62	21	7	-	-	-
58-61	-	-	-	3	50	25	22	-	-	-
61-64	-	-	-	-	19	44	31	6	-	-
64-67	-	-	-	-	-	66	17	17	-	-
67-70	-	-	-	-	-	-	75	25	-	-
70-73	-	-	-	-	-	-	-	33	33	33
> 73	-	-	-	-	-	-	-	-	50	50

period of for instance, markedly increased effort the old fish may disappear from the catches and then a new age/length key will have to be prepared.

Table 3.2.1.6 shows an age/length key for a long-lived tropical fish, the Spanish mackerel (*Scomberomorus brasiliensis*). To illustrate the limitations of an age/length key, consider the percentage age distribution of fish of 61-64 cm long, which are 4-7 years old. Now, if the fishing mortality (effort) increases markedly, most fish older than 5 years may be exterminated and the few fish of 61-64 cm still caught will be fast-growing 4-year-olds and a few remaining 5-year-olds, while the 6 and 7 years old fish have disappeared. Using the old key on the new length-frequency distributions would give the impression that the 61-64 cm fish are still 4-7 years old, with the 5-group dominating, when in fact they are only 4 years old.

When collecting samples for an age/length key it is important to include in the sample some very small, and some very large specimens. Otherwise, when large numbers of length measurements are to be distributed over age groups it will be found that some size classes represented in such length samples are not in the key at all. When small and large fish are deliberately over-represented in the key it is important to remember that the key data alone cannot be used for estimation of growth or mortality parameters.

The methodology of fish stock assessment can in fact be entirely based on age/length compositions alone. The application of mathematical growth models is not necessary. To a certain degree this is the case with the assessments made by the International Council for Exploration of the Sea (ICES) in the North Atlantic. However, since reliable age/length keys are not likely to become available for most tropical species in the near future, as well as for a number of other reasons which will be outlined in the following chapters, the highest priority has been given in this manual to the mathematical growth models.

3.2.2 Length composition data (without age compositions)

Assume that we have a data set consisting of length-frequencies of a certain species, but without age readings. The basic data set for a given sampling date would then look like the "total" (last) column of Table 3.2.1.1 or as drawn in Fig. 3.2.2.1. Is it possible to obtain a separation of the various cohorts which have contributed to this sample without using age reading techniques? The answer is that under certain conditions it is possible, except for parts where the ranges of length-frequencies of different cohorts overlap each other too much.

The hypothetical data set presented in Table 3.2.1.1 was created from a number of normally distributed components, representing cohorts, as shown in Fig. 3.2.2.2.

In Fig. 3.2.2.1 the youngest cohort, the spring cohort of 1983, can easily be distinguished from the rest of the sample. The next cohort, further to the right, is somewhat more difficult to see, while the remaining four cohorts may only be distinguished by using more sophisticated methods than visual inspection, or it may not be possible to separate them at all.

In Section 3.4 methods will be introduced which can be used to split length-frequency samples into normally distributed components, which are assumed to represent cohorts. It will be demonstrated, on the basis of the same data set, that it is not feasible in practice to separate more than three or four cohorts from the total data set. The overlap in the length composition of the older cohorts, the largest fish, clearly limits the analysis. Therefore, the conclusions that may be drawn from such a data set, compared to cases where the age of the fish can be determined, are also limited.

3.2.3 Data from commercial catches

Data for estimation of growth parameters may also be obtained from sampling the commercial catches. The basic principles in analysing samples from commercial landings are the same as for research survey data. The major difference lies in the bias problems. Commercial boats never attempt to collect a random sample of the stock, because they always go for the marketable sizes and try to find the areas with the highest concentrations of fish. However, keeping in mind sources of bias, and trying to stratify the sampling to minimize the bias, data from commercial fisheries can also be used for the estimation of growth parameters.

The major advantage of sampling commercial catches is that such samples are much cheaper to collect and thus sampling can be much more frequent than is possible with a single research vessel. In Chapter 7 problems relating to sampling of commercial catches are further elaborated.

3.3 METHODS FOR ESTIMATION OF GROWTH PARAMETERS FROM LENGTH-AT-AGE DATA

In this section we assume that pairs of observations of age and length are available. They may either be derived from readings of ring structures in hard parts or from length-frequency analysis (Sections 3.4 and 3.5). Input data are either in the detailed form of an age/length composition as in Table 3.2.1.1 or in the processed form shown in Table 3.2.1.2. They may or may not be derived from a time series of samples (cf. Fig. 3.2.1.1). In the following we use for simplicity the input data format illustrated by Table 3.2.1.2.

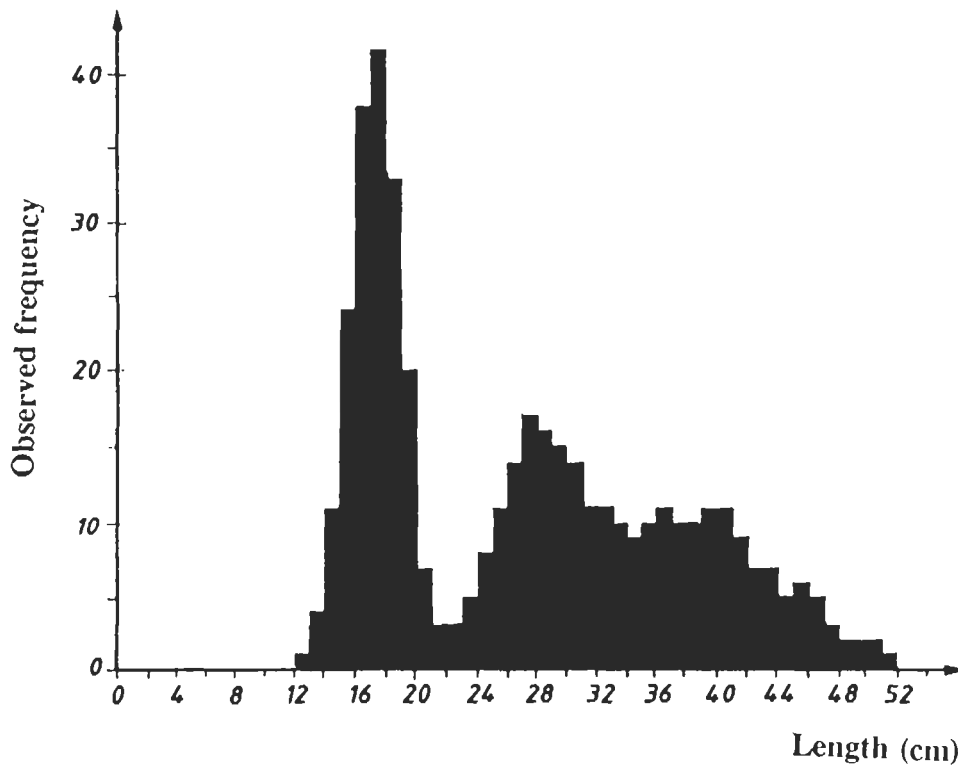


Fig. 3.2.2.1 The length-frequency sample, the only basic data in cases where age reading from hard parts is not possible. (Frequencies from the "total column" of Table 3.2.1.1)

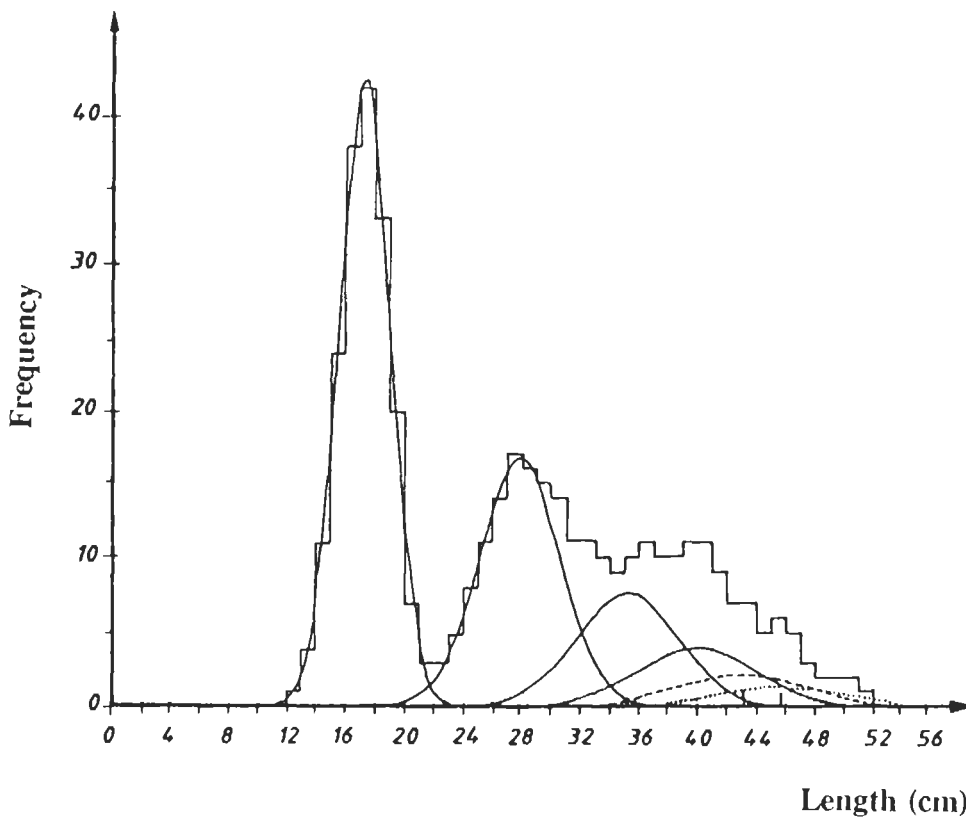


Fig. 3.2.2.2 The length-frequency sample of Fig. 3.2.2.1, separated into normally distributed components. (Frequencies from the "total" column of Table 3.2.1.1). This example is also used to illustrate the "Bhattacharya method" described in Section 3.4.1 and the "maximum likelihood method" discussed in Section 3.5.3

Growth parameters can be derived from such data by graphical methods or plots, which are always based on a conversion to a linear equation, as discussed in Chapter 2. These plots are named after the authors of the papers wherein they were first described, viz. Gulland and Holt (1959), Chapman (1961), Ford-Walford (1933 and 1946 respectively) and von Bertalanffy (1934). An other method that will be discussed is the "least squares method".

3.3.1 The Gulland and Holt plot

The Gulland and Holt (1959) plot was introduced in Section 3.1 by Eq. 3.1.0.4, which can also be written as:

$$\Delta L/\Delta t = K \cdot L_{\infty} - K \cdot \bar{L}(t) \quad (3.3.1.1)$$

The length "L(t)" in Eq. 3.1.0.4 represents the length range from L(t) at age t to L(t+Δt) at age t+Δt. Thus, the natural quantity to enter into Eq. 3.3.1.1 is the mean length (cf. the example in Table 3.1.0.1):

$$\bar{L} = \frac{L(t+\Delta t) + L(t)}{2}$$

Only if Δt is small $\bar{L}(t)$ may be a reasonable approximation to the mean length. However, Δt does not need to be a constant, which is an important advantage over other methods.

Using $\bar{L}(t)$ as the independent variable and ΔL/Δt as the dependent variable Eq. 3.3.1.1 becomes a linear regression:

$$\Delta L/\Delta t = a + b \cdot \bar{L}(t)$$

The growth parameters K and L_{∞} are obtained from:

$$K = -b \quad \text{and} \quad L_{\infty} = -a/b$$

Table 3.1.0.1. contains an example of the input data (columns C and D) and Fig. 3.1.0.3 shows the corresponding plot. The length increment per year or growth rate is plotted against the mean length during the corresponding year. The regression analysis gives:

$$a = 22.40 \quad \text{and} \quad b = -0.3923 \quad \text{from which we get}$$

$$K = -b = 0.39 \quad \text{say } 0.4 \text{ per year, and } L_{\infty} = -a/b = 57.1 \text{ cm}$$

Example 6: Estimating K and L_{∞} with the Gulland and Holt plot

Another example of the Gulland and Holt plot can be derived from Table 3.2.1.2 as shown in Table 3.3.1.1. From the estimates of intercept and slope we get:

$$K = -b = 0.77 \quad \text{say } 0.8 \text{ per year,}$$

$$L_{\infty} = -a/b = -38.52/-0.7670 = 50.2 \text{ cm}$$

Table 3.3.1.1 Input data for the Gulland and Holt plot and regression analysis (data derived from Table 3.2.1.1)

t	Δt	L(t)	ΔL(t)	$\frac{\Delta L(t)}{\Delta t}$ (y)	$\frac{L(t+\Delta t)+L(t)}{2} = \bar{L}(t)$ (x)
0.64		17.3			
1.16	0.52	27.9	10.6	20.4	22.6
1.65	0.49	35.3	7.4	15.1	31.6
2.10	0.45	40.2	4.9	10.9	37.7
2.64	0.54	43.3	3.1	5.7	41.8
3.21	0.57	45.5	2.2	3.9	44.4

b (slope) = -0.7670, a (intercept) = 38.52, $n = 5$, $\bar{x} = 35.62$
 $sb^2 = \frac{1}{n-2} * [(sy/sx)^2 - b^2] = \frac{1}{3} * [(6.7727/8.7362)^2 - 0.7670^2] = 0.004216$
 $sb = 0.06493$, $t_{n-2} = 3.18$, $sb * t_{n-2} = 0.2065$
 95 % confidence limits for b : [-0.974 , -0.561] (cf. Section 2.4)
 $K = -b = 0.77 \pm 0.21$
 $sa^2 = sb^2 * \left[\frac{n-1}{n} * sx^2 + \bar{x}^2 \right] = 0.004216 * \left[\frac{4}{5} * 8.7362^2 + 35.62^2 \right] = 5.607$
 $sa = 2.368$ $sa * t_{n-2} = 7.53$
 95 % confidence limits for a : [31.0 , 46.0]
 $L_\infty = -a/b = -38.52 / -0.7670 = 50.2$ cm

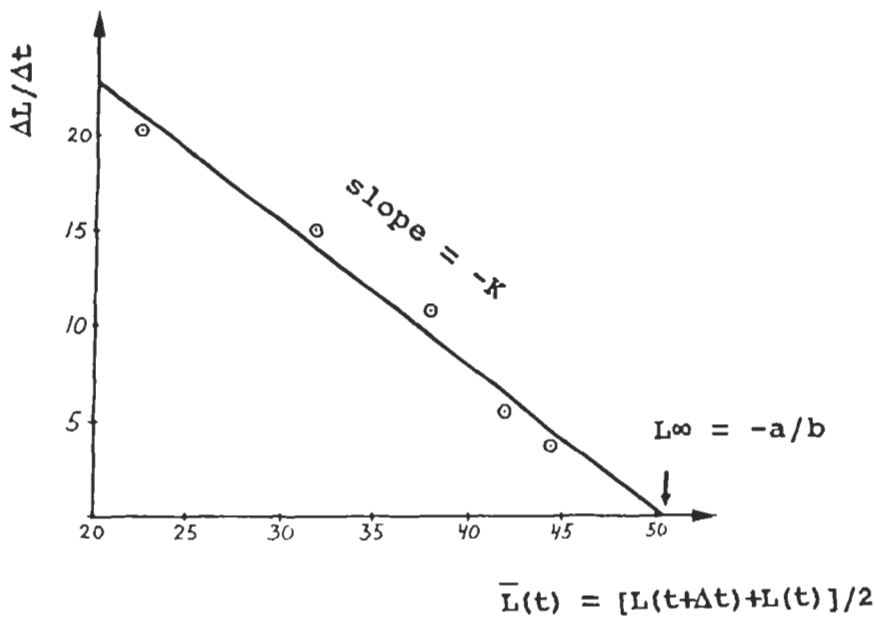


Fig. 3.3.1.1 Gulland and Holt plot corresponding to Table 3.3.1.1 (hypothetical example). The intersection point between the regression line and the x-axis gives L_∞

In Section 3.1 it was stated that it can be proved mathematically that Eq. 3.1.0.4: $\Delta L/\Delta t = K*(L_{\infty}-L(t))$ is equivalent to the von Bertalanffy growth equation (Eq. 3.1.0.1):

$$L(t) = L_{\infty} * [1 - \exp(-K*(t-t_0))]$$

This, however, is correct only if the time interval, Δt , is infinitesimal. Thus, the Gulland and Holt plot, which is based on Eq. 3.1.0.4, is an approximation which is reasonable only for small values of Δt .

(See Exercise(s) in Part 2.)

3.3.2 The Ford-Walford plot and Chapman's method

The method introduced by Ford (1933) and Walford (1946) has gained wide application because the plot could be used to obtain a quick estimate of L_{∞} , without calculations. Nowadays it is not used very much and it has been incorporated here mainly because it will often be found in older papers.

From the von Bertalanffy growth equation (Eq. 3.1.0.1) it follows from a series of algebraic manipulations that:

$$L(t+\Delta t) = a + b*L(t) \tag{3.3.2.1}$$

where

$$a = L_{\infty}*(1-b) \quad \text{and} \quad b = \exp(-K*\Delta t)$$

Since K and L_{∞} are constants, a and also b become constants **if Δt is a constant**. The growth parameters K and L_{∞} are derived from:

$$K = -\frac{1}{\Delta t} * \ln b \quad \text{and} \quad L_{\infty} = \frac{a}{1-b}$$

To illustrate the use of Eq. 3.3.2.1 consider Table 3.3.2.1, where the figures in column A represent lengths, $L(t)$, for a series of ages with a constant time interval of one year, while column B contains the lengths, $L(t+\Delta t)$, the length one year later.

Carrying out the regression analysis we get

$$a = 18.70 \quad \text{and} \quad b = 0.6725$$

from which we derive

$$K = -(1/1) * \ln 0.6725 = 0.3968, \text{ say } 0.4 \text{ per year and}$$

$$L_{\infty} = 18.70 / (1-0.6725) = 57.1 \text{ cm}$$

The actual Ford-Walford plot corresponding to these data is shown in Fig. 3.3.2.1. L_{∞} can be estimated graphically from the intersection point of the 45° diagonal, where $L(t) = L(t+\Delta t)$ and the regression line, because for very old fish, which have stopped growing $L_{\infty} = L(t) = L(t+\Delta t)$.

Table 3.3.2.1 Pairs of consecutive lengths, with $\Delta t = 1$ year, derived from Table 3.1.0.1.
A and B: Input data for the Ford-Walford plot (see Fig. 3.3.2.1)
A and C: Input data for Chapman's method (see Fig. 3.3.2.2)

t	A L(t) (x)	B L(t+ Δt) (y)	C L(t+ Δt)-L(t) (y)
1	25.7	36.0	10.3
2	36.0	42.9	6.9
3	42.9	47.5	4.6
4	47.5	50.7	3.2
5	50.7	52.8	2.1
6	52.8	54.2	1.4

Also the method described by Chapman (1961) and later by Gulland (1969) is based on a constant time interval Δt , that is to say the method is applicable if we have pairs of observations:

$$(t, L(t)), (t+\Delta t, L(t+\Delta t)), (t+2\Delta, L(t+2\Delta t)), \text{ etc.}$$

It can be shown that the von Bertalanffy growth equation (Eq. 3.1.0.1) implies that:

$$L(t+\Delta t) - L(t) = c \cdot L_{\infty} - c \cdot L(t) \quad (3.3.2.2)$$

where

$$c = 1 - \exp(-K \cdot \Delta t)$$

Thus, since K and L_{∞} are constants, and if Δt remains constant, c will remain constant and consequently Eq. 3.3.2.2 becomes a linear regression

$$y = a + bx$$

where

$$y = L(t+\Delta t) - L(t), \quad a = c \cdot L_{\infty}, \quad b = -c \quad \text{and} \quad x = L(t)$$

Note that the slope is negative and also that on the abscissa (x-axis) the smaller of the two lengths is used, instead of the mean value (cf. Section 3.3.1).

The growth parameters are derived from

$$K = -(1/\Delta t) \cdot \ln(1+b) \quad \text{and} \quad L_{\infty} = -a/b \quad \text{or} \quad a/c$$

To illustrate the use of Eq. 3.3.2.2 consider once again Table 3.3.2.1, where $L(t) = x$ in column A and $L(t+1)-L(t) = y$, in column C. A regression analysis gives

$$\begin{aligned} a &= 18.70 \quad \text{and} \quad b = -0.3275 \quad \text{and hence} \quad c = 0.3275 \\ K &= -(1/1) \cdot \ln(1-0.3275) = 0.3968, \text{ say } 0.4 \text{ per year} \\ L_{\infty} &= 18.70/0.3275 = 57.1 \text{ cm} \end{aligned}$$

The plot is given in Fig. 3.3.2.2.

The three methods described in Sections 3.3.1 and 3.3.2 give nearly the same results when applied to the data in Table 3.1.0.1. This is caused by the fact that the data conform exactly to the von Bertalanffy equation because they were obtained from the equation through back

calculation. With real data one should expect to find some differences in the results. These methods can be used to estimate K and L_{∞} . A fourth method, the von Bertalanffy plot can be used to obtain an estimate of K and t_0 . However, this method requires an estimate of L_{∞} as input (see Section 3.3.3).

L_{∞} can also be estimated by the "Powell-Wetherall method". Because this method can also be used to obtain an estimate of the total mortality coefficient, Z , it is presented in the next chapter in Section 4.5.4.

(See Exercise(s) in Part 2.)

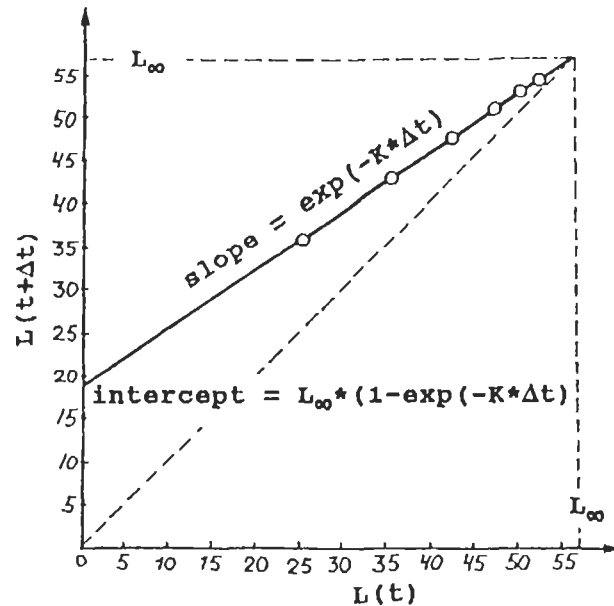


Fig. 3.3.2.1 Ford-Walford plot. Data from columns A and B of Table 3.3.2.1

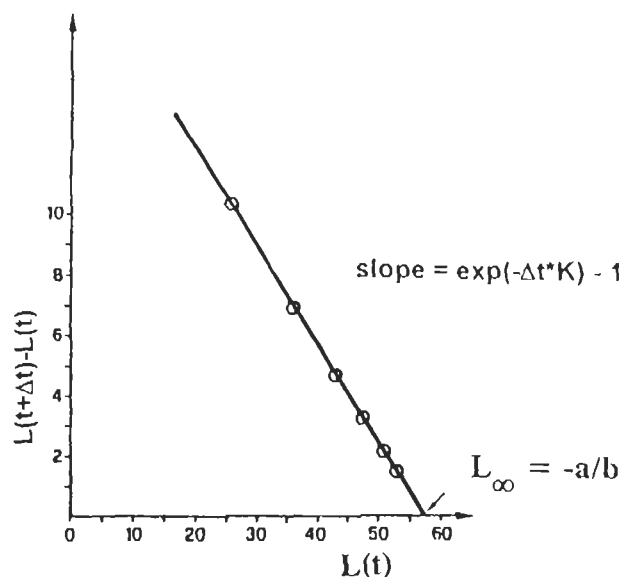


Fig. 3.3.2.2 Chapman's plot. Data from columns A and C of Table 3.3.2.1

3.3.3 The von Bertalanffy plot

The first method for estimating the von Bertalanffy growth parameters was suggested by von Bertalanffy (1934). It can be used to estimate K and t_0 from age/length data, while it requires an estimate of L_{∞} as input.

The von Bertalanffy growth equation (Eq. 3.1.0.1) can be rewritten:

$$-\ln(1 - L(t)/L_{\infty}) = -K*t_0 + K*t \quad (3.3.3.1)$$

With the age, t , as the independent variable (x) and the left-hand side as the dependent variable (y) the equation defines a linear regression, where the slope $b = K$ and the intercept $a = -K*t_0$.

Example 7: Estimating K and t_0 with the von Bertalanffy plot

Table 3.3.3.1 shows how to calculate the input data for the von Bertalanffy plot, based on data from Table 3.3.1.1 with $L_{\infty} = 50$ cm. The plot is shown in Fig. 3.3.3.1. Compare the K value (0.78 per year) with the estimate ($K = 0.77 \pm 0.21$) obtained by the Gulland and Holt plot with the same data.

The von Bertalanffy plot is a more robust method than the Gulland and Holt plot (and the Ford-Walford plot) in the sense that it nearly always gives a reasonable estimate of K , given that a reasonable estimate of L_{∞} is used in the computations (as illustrated in Exercise 3.3.3). One must ascertain, however, that the plot (Fig. 3.3.3.1) "looks" linear. On the other

Table 3.3.3.1 Input data and regression for the von Bertalanffy plot (data derived from Table 3.3.1.1, $L_{\infty} = 50$ cm)

t (x)	$L(t)$	$-\ln(1-L(t)/L_{\infty})$ (y)
0.64	17.3	0.425
1.16	27.9	0.816
1.65	35.3	1.224
2.10	40.2	1.630
2.64	43.3	2.010
3.21	45.5	2.408

$a = -0.0680$ $b = 0.7825$ $K = b = 0.78$ per year $t_0 = -a/b = 0.087$ year
--

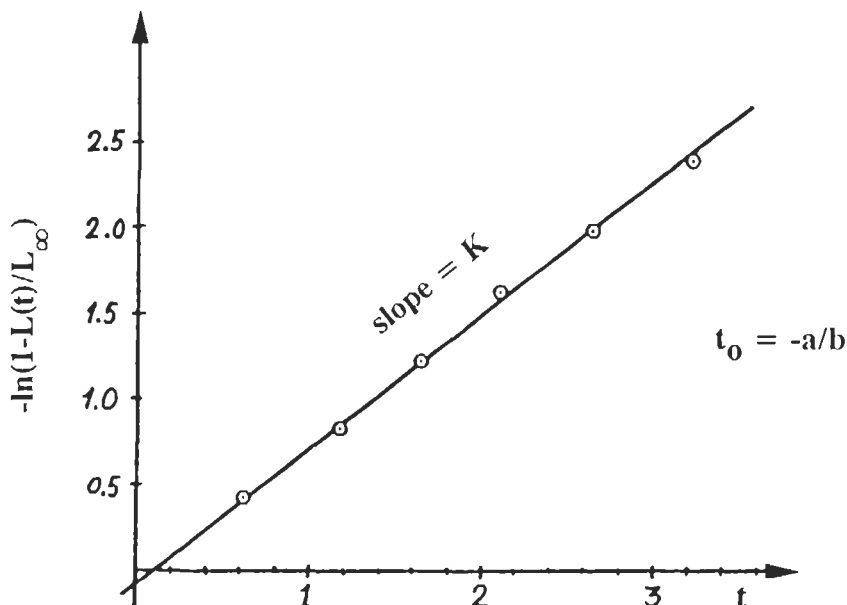


Fig. 3.3.3.1 Von Bertalanffy plot for the data in Table 3.3.3.1

hand, one can say that the Gulland and Holt plot is stronger in the sense that it is better in bringing out cases where the observations are in conflict with the von Bertalanffy model.

Recalling the interpretation of L_{∞} as the average length of a very old fish, there are various short cut methods for estimating L_{∞} for use in the von Bertalanffy plot:

- 1) In small samples you may simply use the largest fish.
- 2) In a very large sample you may take the average of the lengths of, say, the ten largest fish.
- 3) Perhaps the best way of estimating L_{∞} is the Powell-Wetherall method described in Section 4.5.4.

It may not matter as much as one might think which estimate of L_{∞} is used. If you over-estimate L_{∞} the K will be under-estimated, and together they will balance out, so that the resulting growth curve remains nearly the same for the range of ages represented in the data set. (This aspect will be further discussed in Section 3.4.)

There is, however, a problem in using the von Bertalanffy plot in connection with the definition of L_{∞} . The argument of the logarithm in Eq. 3.3.3.1, i.e. $(1 - L(t)/L_{\infty})$ must be positive as the logarithm would otherwise not be defined. Thus the von Bertalanffy plot cannot accept a length greater than L_{∞} , whereas with the definition of L_{∞} as given in Section 3.1.4 it may well happen that for the very old fish, $L(t) > L_{\infty}$ because the observations $(t, L(t))$ fluctuate at random about the line. The von Bertalanffy plot actually uses the "*inverse von Bertalanffy growth equation*":

$t(L) = t_0 - \frac{1}{K} \ln(1 - L/L_{\infty}) \quad (3.3.3.2)$
--

which is Eq. 3.1.0.1 solved for t . It may be necessary to omit the oldest fish to obtain that $1 - L/L_{\infty} > 0$.

The L_{∞} concept as applied in the von Bertalanffy plot is different from that applied in the Gulland and Holt plot, for the same reasons as the parameters in the "inverse linear regression" differ from those of the "original linear regression".

(See Exercise(s) in Part 2.)

3.3.4 The least squares method

This method is assumed to be superior to the methods introduced in the foregoing subsections from an estimation theory point of view. It is the non-linear parallel to the linear regression analysis introduced in Section 2.4. However, the computational work involved is considerable and in practice you need a computer to do the calculations.

Assume a series of pairs of observations (length, age) to be available. These may have been obtained by age reading (cf. Section 3.2.1) or they may have been derived from modal progression analysis (to be discussed in Section 3.4.2). Let there be n pairs of observations:

$$(L(i), t(i)) = (\text{length of fish no. } i, \text{ age of fish no. } i)$$

where

$$i = 1, 2, \dots, n.$$

The method estimates the growth parameters in such a way that the sum of the squares of the deviations between the model and the observations is minimized, i.e. it minimizes the sum with respect to the parameters L_{∞} , K and t_0 :

$$\sum_{i=1}^n [L(i) - L_{\infty} * [1 - \exp(-K * (t(i) - t_0))]]^2 \quad (3.3.4.1)$$

Computer programs

The LFSA package of microcomputer programs for fish stock assessment (Sparre, 1987) contains the program "VONBER" which can do the least square estimation of growth parameters. The method used by this program is rather complicated and a full explanation falls outside the scope of this manual. However, conceptually, non-linear regression analysis is not more complex than linear regression, just as the square root of 3 is conceptually not more complex than the square root of 4, but the latter is much easier to calculate. FiSAT too contains such a program. Many other similar computer programs are available (see Chapter 15).

3.4 ESTIMATION OF AGE COMPOSITION FROM LENGTH-FREQUENCIES

In Section 3.3 we have dealt with methods for the estimation of the growth parameters of the von Bertalanffy growth equation. All these methods require input data on length and age. As has been stated earlier, it is difficult to determine the ages of tropical fish so, in most cases, only length-frequency data will be available. This section deals with the analysis of length-frequency data. The aim of the methods described below is to assign ages to certain length groups. In other words, the aim is to separate a complex length-frequency distribution into cohorts and to assign an arbitrary age to each of those cohorts. Since the mean length of each cohort can also be determined, we have then obtained the combination of length and age data which is necessary to determine the growth parameters using the methods described in Section 3.3. Before going into specific methods, the difficulties involved in this kind of analysis will be illustrated on the basis of an example from the tropics, after a short introduction of the first known application of these methods in Denmark.

Example 8: Estimating the age of species from temperate waters

The basic idea behind the techniques to be described in this section dates back to one of the first works on fishery biology, a paper on the eel-pout (*Zoarces viviparus*) by Petersen (1892). The length measurements of 156 fish are represented by dots in Fig. 3.4.0.1. Petersen divided the 156 fish into juveniles, males and females and he further divided the adult fish into two size groups:

- Medium sized: from 5 to 8 inches
- Large sized : from 9 inches and upwards

From earlier observations Petersen knew that in the winter juveniles of about 1.5" could be caught, while in summer the juveniles would all be from 3" to 5" long. Because certain length groups, depending on the season, appeared to be absent, he concluded that the three length groups in the July sample should be interpreted as follows:

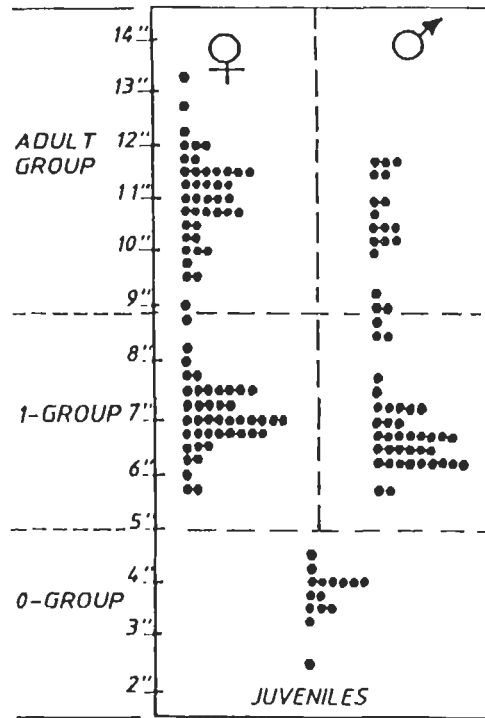


Fig. 3.4.0.1 Length-frequency sample of 156 eel-pout (*Zoarces viviparus*) in units of Danish inches. Collected in Holbaek Fjord (Denmark) 10-11 July 1890. (Redrawn from Petersen, 1892)

less than 5": 0-group, born in winter 1889/90
 from 5" to 8": 1-group, born in winter 1888/89
 from 9" and upwards: 2+-group, born in winter 1887/88 or earlier

(The symbol "2+" stands for the "2-group plus older groups". We call it the "2 plus group".)

Petersen's findings indicated that *Zoarces viviparus* give birth once per year during a restricted period. For most temperate species propagation takes place during 2-4 months in winter or spring. Such a breeding pattern makes it relatively easy to define a cohort. In temperate waters a cohort is simply a year-class of fish. Because all fish grow at approximately the same rate, a cohort can be followed during the first part of its life by tracing the peaks in the length-frequency samples. But when they approach their maximum size this is no longer possible, because by then fish of different ages have reached almost the same length.

Example 9: Estimating the age of coral trout, a tropical species

We shall now discuss a similar analysis with a species from a tropical area. Fig. 3.4.0.2a shows a length-frequency sample of the coral trout (*Plectropomus leopardus*) obtained by Goeden (1978), from Heron Island, Australia. This example appears easy to handle. There are four distinct peaks (A,B,C and D) and it is tempting to interpret these as age groups 1,2,3 and 4, as was done by Goeden. However, a closer examination shows that this interpretation does not conform to the von Bertalanffy model. The mean lengths of the peaks B,C and D are approximately 35 cm, 42 cm and 50 cm respectively. When we interpret these peaks as belonging to successive yearly age groups the growth rates become:

between peaks B and C: $(42-35)/1 = 7$ cm/year
 between peaks C and D: $(50-42)/1 = 8$ cm/year

This does not conform to the von Bertalanffy growth curve, since the growth rate between peaks C and D is expected to be smaller than that between peaks B and C. Thus to give an interpretation conforming to the von Bertalanffy model, peak D must be assigned an age two years older than peak C, and an additional age group must be assumed to exist between peaks C and D. A likely explanation is that peaks C and D represent strong year classes (large number of cohort members), whereas the cohort represented by length groups between peaks C and D stems from a poor year class.

The small solid bars shown on the length axis of Fig. 3.4.0.2a are the lengths at age 1,2,...,7 corresponding to a von Bertalanffy growth curve with the parameters:

$$L_{\infty} = 57 \text{ cm, } K = 0.4 \text{ per year and } t_0 = -0.5 \text{ years}$$

Table 3.4.0.1 Length-at-age for alternative choices of growth parameters. Plots with observed frequencies are shown in Fig. 3.4.0.2 for columns a, b and c

	a *)	b	c		d
age	L_{∞} 57.0 K 0.40 t_0 -0.50	59.9 0.40 -0.50	59.9 0.34 -0.65		70.0 0.21 -1.15
0	10.3	10.8	11.8		15.1
1	25.7	26.8	25.5	-----	25.5
2	36.0	37.6	35.3		33.9
3	42.9	44.8	42.3		40.8
4	47.5	49.7	47.3		46.3
5	50.7	52.9	50.8	-----	50.8
6	52.8	55.1	53.3		54.5
7	54.2	56.5	55.1		57.4
*) see Table 3.1.0.1					

The length corresponding to these bars were also used in Table 3.1.0.1, they have been repeated in column a of Table 3.4.0.1. These parameters interpret the peaks A,B,C and D as the 1-, 2-, 3- and 5-group respectively and place the 4-group between peaks C and D. This particular choice of growth parameters is not based on any fitting techniques or any other rational method. They were derived by a short series of trials with different parameters until a curve was obtained which placed the mean lengths of the various cohorts close to where the peaks are, except for age group 4.

Most likely, this is not the only set of growth parameters which produces a growth curve that gives a certain correspondence to the peaks of Fig. 3.4.0.2a. One could have used the greatest length, 59.5 cm, as the estimate of L_{∞} , for example. (Note that considering the definition of L_{∞} as the average length of a very old fish, it would not in general be correct to use the largest fish observed as an estimate of L_{∞} . However, in this case with only 312 fish in the sample the largest fish may give a reasonable estimate of L_{∞} .) Using this $L_{\infty} = 59.5$ cm together with the same K value 0.4 and the same t_0 value -0.5 gives the lengths-at-age shown in column b of Table 3.4.0.1. Fig. 3.4.0.2b shows the corresponding mean lengths-at-age together with the length-frequency sample. Obviously, this choice of growth parameters produces a less convincing fit to the peaks than the one shown in Fig. 3.4.0.2a.

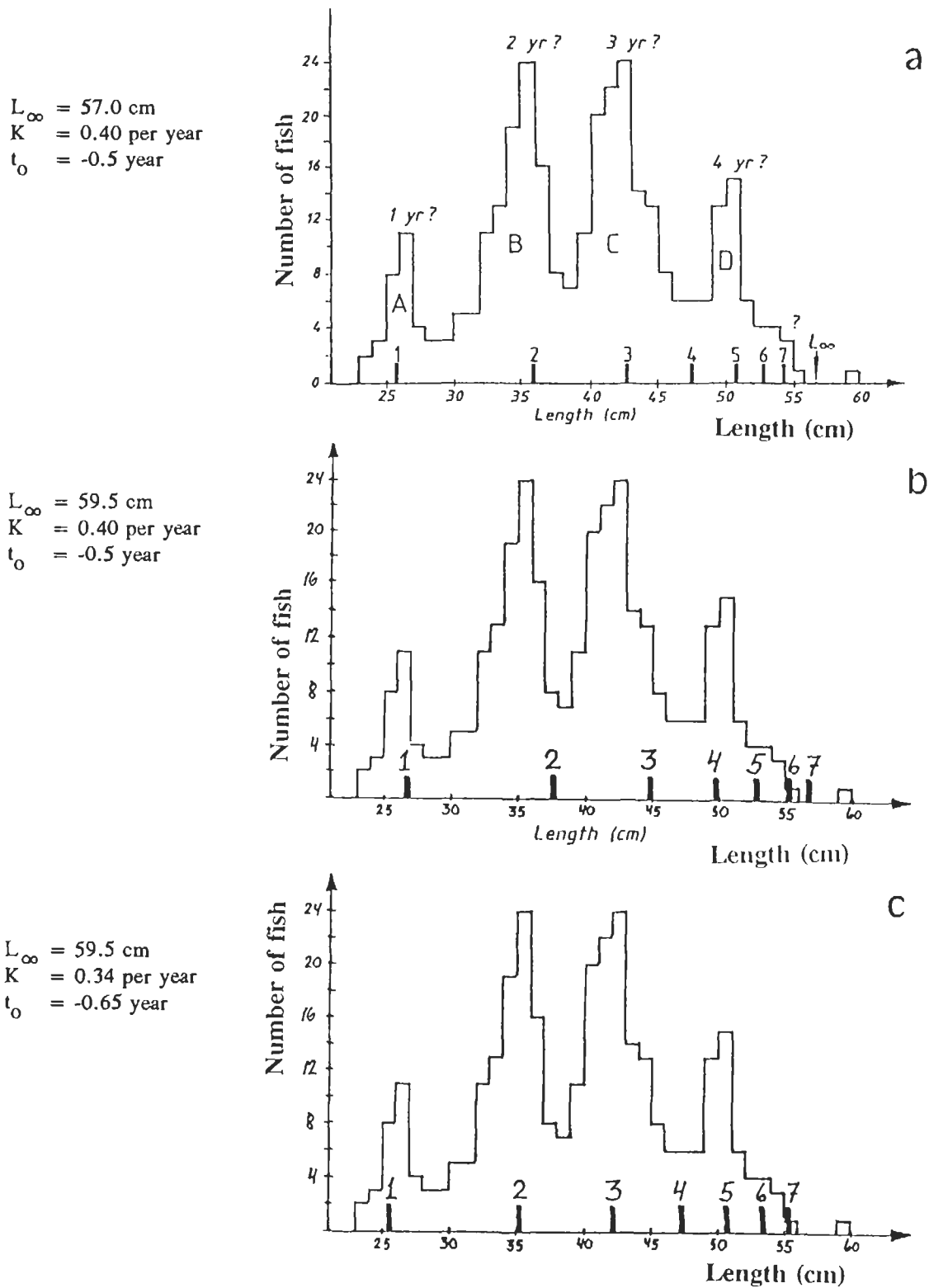


Fig. 3.4.0.2 Length-frequency sample of the coral trout (*Plectropomus leopardus*) from Goeden (1978). The small bars on the x-axis indicate the lengths-at-age corresponding to the growth parameters given in columns a, b and c of Table 3.4.0.1

Note: Mr. H. Weng, Brisbane, Australia, has drawn our attention to the fact that the coral trout (*Plectropomus leopardus*) changes sex, from male to female when reaching a length of 30 to 35 cm. This fact was mentioned by Goeden (1978), but overlooked by us. Although the results obtained in this example may not be the "real ones", the example still serves as an illustration of the method.

Reducing K to 0.34 and t_0 to -0.65 gives a much better agreement between peaks and mean lengths as shown in Fig. 3.4.0.2c. The corresponding mean lengths-at-age are given in column c of Table 3.4.0.1. Whether this fit is better than the fit shown in Fig. 3.4.0.2a is difficult to assess by visual inspection only.

In general, it is difficult to define a unique solution to this kind of problem. Different values of the growth parameters L_∞, K, t_0 may produce very similar growth curves. This becomes obvious when one observes that for a given value of L_∞ one can always determine corresponding values of the other two growth parameters K and t_0 so that the curve passes through two pre-specified points in the age/length coordinate system.

As an example let us give L_∞ the value 70 cm and let us determine K and t_0 so that the curve thus obtained comes close to the curve given in column c of Table 3.4.0.1. We do that by selecting K and t_0 so that the length at age $t = 1$, $L(1) = 25.5$ cm and $L(5) = 50.8$ cm.

Formulas for K and t_0 can be derived from Eq. 3.3.3.1 as follows:

$$-\ln\left(1 - \frac{L(t_1)}{L_\infty}\right) = -K*t_0 + K*t_1 \quad (a)$$

$$-\ln\left(1 - \frac{L(t_2)}{L_\infty}\right) = -K*t_0 + K*t_2 \quad (b)$$

Subtract (b) from (a).

Since $\ln a - \ln b = \ln \frac{a}{b}$ and after some rearranging, we get

$$\ln \frac{L_\infty - L(t_1)}{L_\infty - L(t_2)} = K*(t_2 - t_1) \quad \text{or}$$

$$K = \frac{1}{t_2 - t_1} * \ln \frac{L_\infty - L(t_1)}{L_\infty - L(t_2)} \quad (3.4.0.1)$$

The formula for t_0 is simply obtained by rearranging Eq. 3.3.3.1. In the case of $t = t_1$ it becomes

$$t_0 = t_1 + \frac{1}{K} * \ln\left(1 - \frac{L(t_1)}{L_\infty}\right) \quad (3.4.0.2)$$

Thus, for $t_1 = 1$ and $t_2 = 5$ corresponding to $L(1)$ and $L(5)$ respectively we get:

$$K = \frac{1}{5-1} * \ln \frac{70.0-25.5}{70.0-50.8} = 0.21 \text{ per year} \quad \text{and}$$

$$t_0 = 1 + \frac{1}{0.21} * \ln\left(1 - \frac{25.5}{70.0}\right) = -1.15 \text{ year}$$

These growth parameters produced the lengths-at-age shown in column d of Table 3.4.0.1. The two growth curves corresponding to columns c and d of Table 3.4.0.1 are shown in Fig. 3.4.0.3. It appears, that it is difficult to decide which of the two growth curves gives the best fit to the length-frequency sample in Fig. 3.4.0.2.

It is often extremely difficult to obtain an unambiguous interpretation of a data set of length-frequencies of tropical fish, in particular when there is only one complex length-frequency sample available and not a time series (see Venema, Christensen and Pauly, 1988). Additional

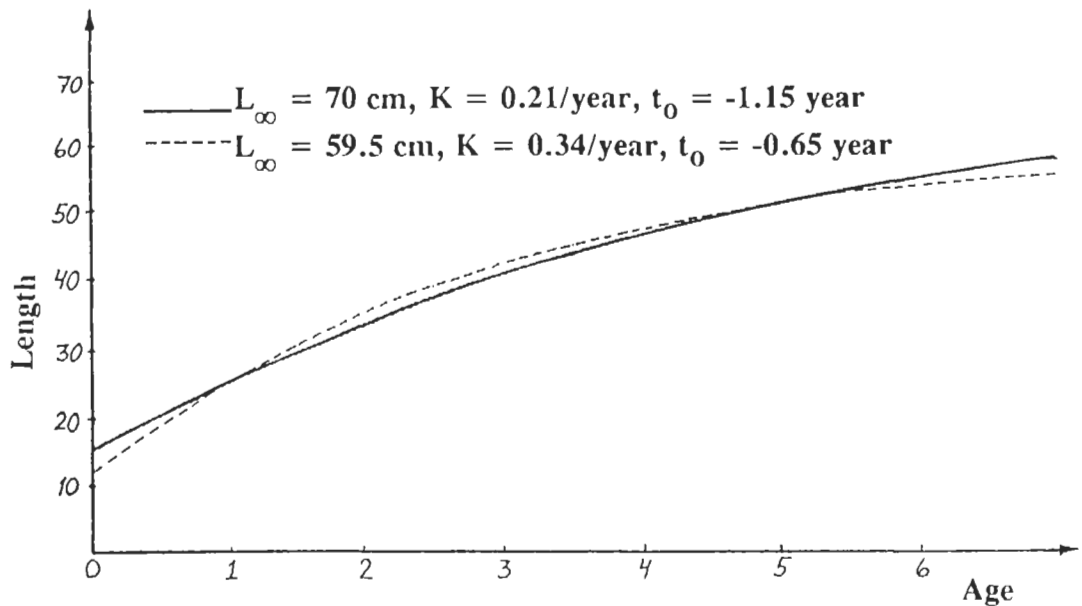


Fig. 3.4.0.3 Example of two growth curves which are approximately equal, but have quite different growth parameters. Derived from columns c and d of Table 3.4.0.1

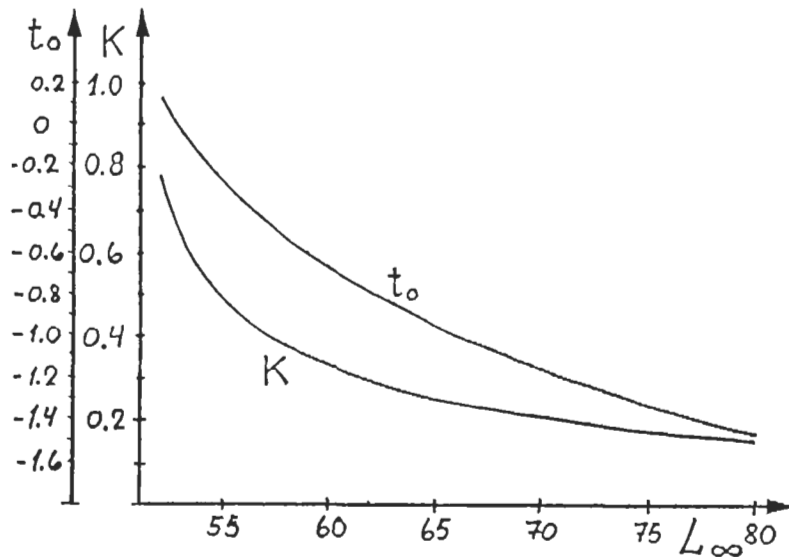


Fig. 3.4.0.4 The growth parameters K and t_0 as a function of L_∞ for growth curves fulfilling the condition $L(1) = 25.5$ cm and $L(5) = 50.8$ cm

information on the biology of the species in question may help a lot in correctly interpreting the data.

The interrelationship between the growth parameters, L_∞ , K and t_0 , is further demonstrated in Fig. 3.4.0.4 which shows K and t_0 as a function of L_∞ for growth curves fulfilling the condition $L(1) = 25.5$ cm and $L(5) = 50.8$ cm. Note that K and t_0 decrease as L_∞ increases. Thus, when comparing different estimates of K , L_∞ and t_0 the comparison should not be made on the basis of one individual parameter but on the basis of the resulting growth curves. In the example of columns c and d in Table 3.4.0.1 we would say that the two parameter sets:

$$(L_\infty, K, t_0) = (59.5, 0.34, -0.65) \quad \text{and}$$

$$(L_\infty, K, t_0) = (70.0, 0.21, -1.15)$$

are approximately equal in the sense that they produce nearly the same growth curves within the range of ages covered.

The more old fish in the sample, the better is the estimate of L_{∞} and the estimate of K becomes less dependent of the estimate of L_{∞} .

The above discussion leads to a warning: do not always consider estimates of growth parameters to be directly related to the physiology of the fish. Only when the sample is large and unbiased can you expect the estimated parameters to reflect their physiological interpretation.

Comparison of growth curves

The growth parameter K is related to the metabolic rate of the fish. Pelagic species are often more active than demersal species and have a higher K . The metabolic rate is also a function of temperature: tropical fishes have higher K values than cold-water fishes. These relations are clouded, however, by a correlation of K and L_{∞} such that small species have higher K values than large species at the same level of activity. A further complication is the statistical correlation of K and L_{∞} described above: different combinations of K and L_{∞} can give almost the same fit to data except when a wide range of ages is represented. Again, a high value of K combines with a low value of L_{∞} and *vice versa*. These two types of correlation are mixed up in the literature. We shall look into them separately.

The statistical correlation of K and L_{∞} within species

When the mean lengths of several age groups have been calculated from samples these means have a variance. It is therefore not surprising if two samples of the same population provide mean lengths differing by for instance 0.5 cm. We shall see what happens to the estimation of K and L_{∞} when the lengths-at-age are altered by that amount. Column B of Table 3.4.0.2 shows lengths at ages of 1-5 years of a fish with $L_{\infty} = 60$ cm, $K = 0.24$, $t_0 = 0$. Increments are calculated in column C. L_{∞} and K are estimated from these by Chapman's method (Section 3.3.2). Differences from the true values are due to rounding errors. Column D gives the lengths with 0.5 cm alternately added and subtracted as indicated by arrows. The ensuing L_{∞} is higher and K is lower. In column F the original lengths are again changed by 0.5 cm, but in the opposite direction. L_{∞} is lower and K higher.

We now have three estimates of L_{∞} and K assumed to be based on three samples from the same population. Fitting the equation

$$\ln K = a + b \cdot \ln L_{\infty}$$

to the three "observations" of L_{∞} and K gives

$$\ln K = 6.67 - 1.98 \cdot \ln L_{\infty}$$

Resulting from such a small trial the estimate of a slope of about -2 is not reliable of course, but Pauly (1979) did similar estimations for more than 100 species of fish for which at least three pairs of L_{∞} and K had been published. He calculated the slope for each species, averaged and found a mean extremely close to -2. W_{∞} in stead of L_{∞} was used in the actual estimation assuming $W_{\infty} = q \cdot L_{\infty}^3$. Also, base 10 logarithms were used which does not alter the slope. Pauly found:

Table 3.4.0.2 The effect of inaccurate determination of mean lengths at age upon the estimation of K and L_{∞} . Arrows indicate whether the original lengths are changed upwards or downwards

A	B	C	D	E	F	G
			mean length altered 0.5 cm			
age years	mean length cm	$L(t+1) - L(t)$	1st alteration		2nd alteration	
t	L	ΔL	L	ΔL	L	ΔL
1	12.80	10.07	↑ 13.20	9.17	↓ 12.30	11.07
2	22.87	7.92	↓ 22.37	8.92	↑ 23.37	6.92
3	30.79	6.24	↑ 31.29	5.24	↓ 30.29	7.24
4	37.03	4.90	↓ 36.53	5.90	↑ 37.53	3.90
5	41.93	-	↑ 42.43	-	↓ 41.43	-
L_{∞}	60.16		67.46		53.82	
K	0.2392		0.1931		0.3018	
$\ln K = 6.6731 - 1.976 \ln L_{\infty}$						

$$\log K = \phi - 0.67 \cdot \log W_{\infty} \quad (3.4.0.5)$$

$$\log K = \phi' - 2 \cdot \log L_{\infty} \quad (3.4.0.6)$$

in which

$$\phi' = \phi - 0.67 \cdot \log q \quad (3.4.0.7)$$

ϕ and particularly ϕ' (phi-prime) are much used as the probably best means of averaging growth parameters of a particular species. ϕ' is calculated for each data set and averaged. Inserting a value of L_{∞} , for instance the mean of all estimates, into Eq. 3.4.0.6 gives a value of K corresponding to the L_{∞} inserted. Whenever L_{∞} and K are estimated from a new set of data for the same species a calculation of ϕ' indicates if the new pair of L_{∞} and K is in accordance with previous results. The new ϕ' should be close to the previous estimates, because ϕ' is the constant in the regression of $\log K$ upon $\log L_{\infty}$. If it is markedly different there is reason to suspect the reliability of the new estimates of K and L_{∞} .

Pauly's data material includes not only the statistical correlation illustrated in Table 3.4.0.2, but also real differences between years and localities - perhaps mainly caused by differences of temperature and food availability. The relation of L_{∞} to K when differences of growth curves are caused by environmental conditions might be investigated utilizing ϕ' . Several pairs of L_{∞} and K should be combined to one estimate for each environmental condition. This would reduce or eliminate the effects of statistical correlation.

Inter-specific correlation of K and asymptotic size

Investigation of the correlation of K and asymptotic size between species requires data for species of very different sizes, otherwise the statistical and environmental effects discussed above may interfere with the analysis. W_{∞} should be used unless the species investigated are of almost the same shape, i.e., have the same q.

Fig. 3.4.0.6 shows $\ln K$ plotted against $\ln W_{\infty}$ for 81 species of fish for which estimates of W_{∞} range from 0.8 g to 852 kg. The relationship is

$$\ln K = 0.071 - 0.200 \cdot \ln W_{\infty}$$

The slope is about one third of the one found above for the statistical correlation within species, Eq. 3.4.0.5. The estimate has a high variance because fish of very different metabolic level are included in the analysis. Pelagic fishes with a high metabolic rate are responsible for most of the points above the line in Fig. 3.4.0.6. Demersal and deep sea fishes from cold water account for many points below the line.

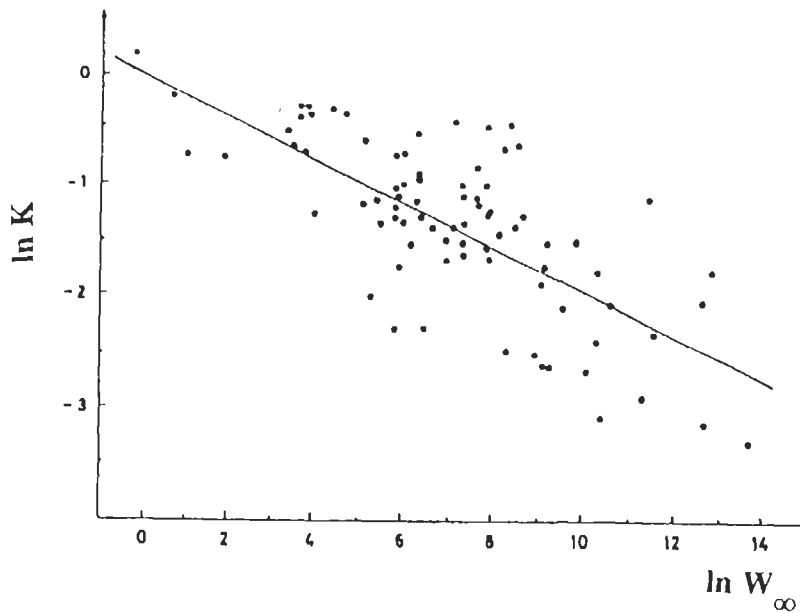


Fig. 3.4.0.6 The relationship of K and W_{∞} in 81 species of fish ranging from $W_{\infty} = 0.8$ g to $W_{\infty} = 852$ kg. From Ursin (1968) redrawn

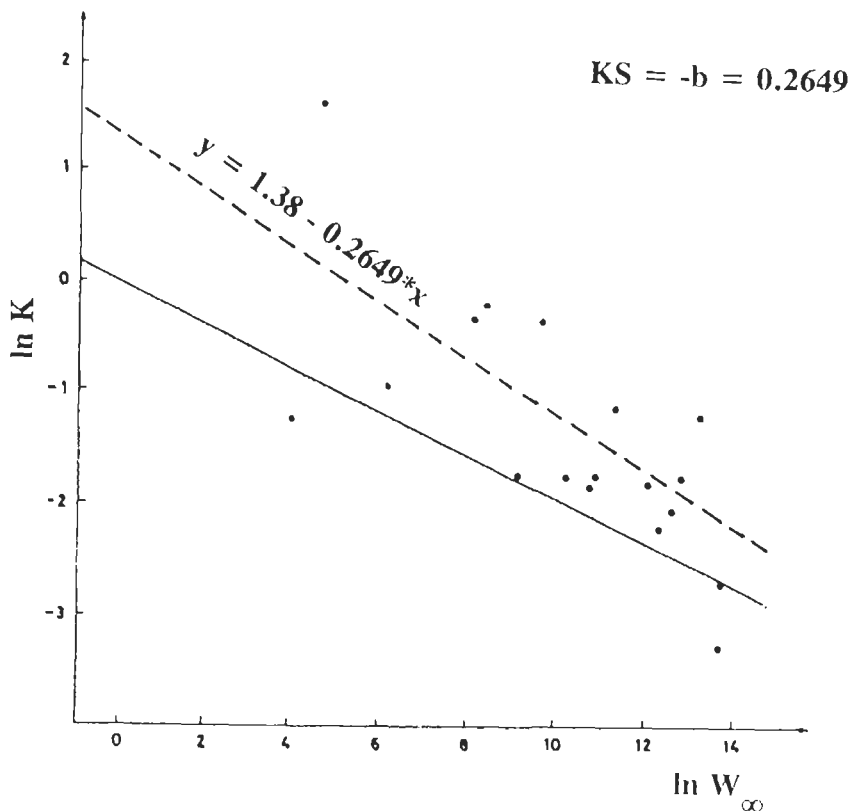


Fig. 3.4.0.7 High level of K in pelagic species: Scombridae. The regression line from Fig. 3.4.0.6 is inserted (drawn line), together with the regression line for the Scombridae alone (dotted line). Data as in Table 3.4.0.3

Analysis by individual fish families reduces the variance because most families are either predominantly pelagic or predominantly demersal. This is illustrated by a plot for Scombridae (mackerel and tuna) in Fig. 3.4.0.7, where the regression line from Fig. 3.4.0.6 has also been inserted.

The relationship described can be written as

$$\ln K = \ln K_0 - KS \cdot \ln W_\infty$$

or

$$K = K_0 \cdot W_\infty^{-KS} \quad (3.4.0.8)$$

in which $\ln K_0$ is comparable to ϕ of the within-species correlation, Eq. 3.4.0.5. For the estimation of K_0 is used

$$\ln K_0 = \ln K + KS \cdot \ln W_\infty$$

or

$$K_0 = K \cdot W_\infty^{KS} \quad (3.4.0.9)$$

K_0 is an index of the metabolic rate of a fish moving normally, but not recently fed (Ursin, 1968). The metabolic rate can be expressed by, for instance, oxygen consumption or weight loss during starvation. K_0 is independent of the size of the species: plotting $\ln K_0$ against $\ln W_\infty$ gives a regression line of slope zero, whereas plots of ϕ or ϕ' against $\ln W_\infty$ (or $\log W_\infty$) have a steep slope as illustrated in Fig. 3.4.0.8.

Fig. 3.4.0.7 shows a plot of $\ln K$ against $\ln W_\infty$ for various species of the family Scombridae. The slope of the regression line pertaining to the points is -0.26 (dotted line), the corresponding value of $KS = 0.26$. This result and those of four other families are listed in Table 3.4.0.3. The average KS value is 0.22, which is one third of the value of the slope in Pauly's ϕ formula, 0.67 in Eq. 3.4.0.5.

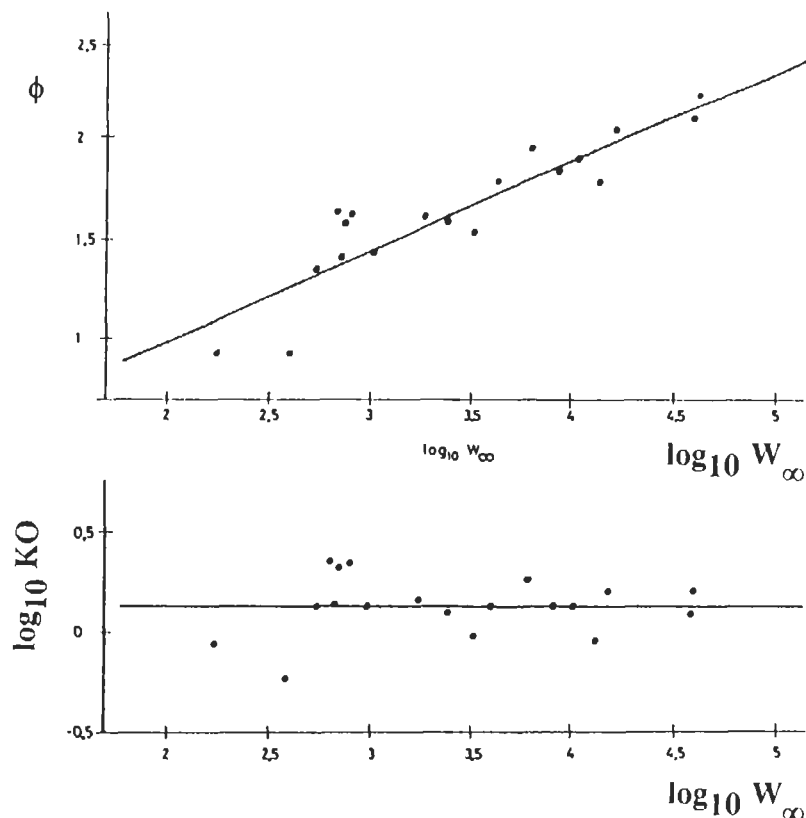


Fig. 3.4.0.8 Plot of ϕ (top) and $\log K_0$ (bottom) against $\log W_\infty$ for Serranidae. Data from Munro, 1983a

Table 3.4.0.3 Estimation of the parameters of Eq. 3.4.0.9 for five families of fish. Data from Ursin, 1968, Pauly, 1980b, and Munro, 1983a

family	no of pairs of K and W_{∞}	- slope KS	mean of $\ln K0$	mean metabolic index $K0$
Myctophidae	5	0.28	-0.21	0.81
Pleuronectidae	7	0.17	0.31	1.36
Gadidae	12	0.21	0.34	1.40
Scombridae	18	0.26	1.01	2.75
Serranidae	19	0.20	0.31	1.36
mean		0.22		

The mean metabolic index $K0$ for these five families varies from 0.81 in the mesopelagic Myctophidae to 2.75 in the epipelagic Scombridae.

$K0$ is useful to get a first estimate of a growth curve for a species which has not been satisfactorily investigated. W_{∞} may be guessed from the size of the larger fish in the catch, while K is determined from Eq. 3.4.0.8 using the mean $K0$ estimated for the family and the over-all estimate of KS for all families combined ($KS = 0.22$ in Table 3.4.0.3).

As an example, consider a species of the family Serranidae for which $q = 0.02$ and L_{∞} is guessed to be 30 cm. W_{∞} is then $0.02 \cdot 30^3 = 540$ g, which combined with $KS = 0.22$ and $K0 = 1.36$ (see Table 3.4.0.3) gives

$$K = 1.36 \cdot 540^{-0.22} = 0.34$$

Like the ϕ formula Eq. 3.4.0.8 can be changed into a function of L_{∞} by inserting $W_{\infty} = q \cdot L^3$, which gives

$$\ln K = \ln K0' - 3 \cdot KS \cdot \ln L_{\infty} \quad (3.4.0.10)$$

where

$$\ln K0' = \ln K0 - KS \cdot \ln q \quad (3.4.0.11)$$

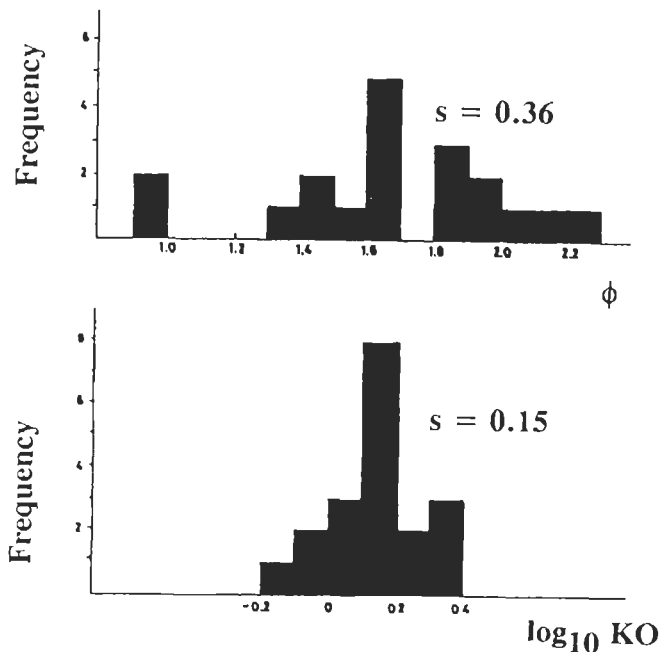


Fig. 3.4.0.9 Frequency distributions of estimates of ϕ (bottom) and of $\log_{10} K0$ (top) for species of Serranidae, cf. Fig. 3.4.0.8

Here, just like ϕ' , $\ln K0'$ is a function of q and therefore dependent on the shape of the fish. Whereas ϕ' is used within a species, and consequently with q constant or nearly so, $K0$ and $K0'$ are used for comparisons between different species where q will often be a variable. In such cases, where q is not constant, Eq. 3.4.0.8 should be used.

To emphasize the different uses of ϕ (for stocks of the same species) and $K0$ (for species within a family) consider Fig. 3.4.0.8 in which ϕ and $\log K0$ are plotted against $\log W_\infty$. (Base 10 logarithms are used because ϕ is such defined.)

For $\log K0$ the slope is zero, because $K0$ is independent of the size of the species, whereas ϕ has a slope of 0.47. This is approximately the difference between the slope 0.67 of the ϕ equation (Eq. 3.4.0.5) and the slope of the $K0$ equation $KS = 0.22$ (Eq. 3.4.0.8 and Table 3.4.0.3).

The standard deviations about the ϕ regression line in Fig. 3.4.0.8 (top) and about the $\log K0$ line (Fig. 3.4.0.8, bottom) are the same. However, the full distribution of ϕ has a higher standard deviation when the linear relationship with $\log W_\infty$ is ignored, see Fig. 3.4.0.9 (top).

Several authors have published histograms similar to Fig. 3.4.0.9 (top) of the distribution of ϕ for certain families of fish. In those histograms one must expect that the low values represent small species and the high values large species.

3.4.1 Bhattacharya's method

In Section 2.2 several means of graphical representation of a normal distribution were introduced. One of these is the Bhattacharya (1967) method, which is useful for splitting a composite distribution into separate normal distributions, i.e. when several age groups (cohorts) of fish are contained in the same sample. This method will be discussed in detail based on the hypothetical example of Table 3.2.1.1. In this case we know the solution: the set of normal distributions of which the total is composed. It is therefore possible to check the validity of the result of the analysis.

Basis of the computation procedures of the Bhattacharya method

The Bhattacharya method consists basically of separating normal distributions, each representing a cohort of fish, from the overall distribution, starting on the left-hand side of the total distribution. Once the first normal distribution has been determined it is removed from the total distribution and the same procedure is repeated as long as it is possible to separate normal distributions from the total distribution. The whole process can be divided into the following stages:

- Stage 1:** Determine an uncontaminated (clean) slope of a normal distribution on the left side of the total distribution.
- Stage 2:** Determine the normal distribution of the first cohort by means of a transformation into a straight line.
- Stage 3:** Determine the numbers of fish per length group belonging to that first cohort and then subtract them from the total distribution.

Stage 4: Repeat the process for the next normal distribution from the left, until no more clean normal distributions can be found.

Stage 5: Relate the mean lengths of the cohorts determined in stages 1 and 4 to the age difference between the cohorts.

As has already been shown in Section 2.6, a normal distribution is transformed into a straight line when: 1) numbers are replaced by their logarithms and 2) differences are calculated between consecutive logarithmic values. Let N designate the number in a length-frequency sample belonging to the length group:

$$[x - dL/2, x + dL/2]$$

where dL is the interval size, x is the interval midpoint, $x - dL/2$ is the lower and $x + dL/2$ is the upper limit of the interval.

If a certain length range in the sample contains only one cohort, this part of the frequency sample should conform to a normal distribution (e.g. from 10 cm to 21 cm in the sample shown in Fig. 3.2.2.2). In that case the linear relationship (cf. Eq. 2.6.5):

$$\Delta \ln N = a + b \cdot (x + dL/2)$$

would hold between the difference of the logarithm of the number in a certain length class and the logarithm of the number in the preceding class or

$$\Delta \ln N = \ln N(x + dL/2, x + 3dL/2) - \ln N(x - dL/2, x + dL/2)$$

as the dependent variable, y ,

and the upper limit of the smallest length group:

$$x + dL/2$$

as the independent variable x (compare Figs. 2.6.4a and 2.6.5).

Recall that the standard deviation of the normal distribution and the mean are obtained by:

$$s = \sqrt{-dL/b} \quad \text{and} \quad \bar{x} = -a/b \quad (\text{compare Eqs. 2.6.6 and 2.6.7})$$

Example 10: A Bhattacharya analysis of a constructed data set

The computation procedures related to steps 1 to 5 in the previous section will be illustrated below on the hand of the constructed data set presented earlier in Table 3.2.1.1 and the related graphical representation in Fig. 3.2.2.1. This data set was created from 6 normally distributed components, as shown in Fig. 3.2.2.2.

We will now try to use the Bhattacharya method to analyse the "total" column of Table 3.2.1.1, trying to break it up into the six normal contributions from which it has been composed. The advantage of using a constructed data set is that it is then possible to compare the results of the Bhattacharya analyses with the exact input data. The possibilities and limitations of the method can thus be illustrated. The computation procedure will be followed step by step in general terms, with examples drawn from the data set of Table 3.2.1.1. Unless indicated otherwise, the examples refer to Table 3.4.1.1, which is the first of a number of work sheets.

Step 1: Create a work sheet like Table 3.4.1.1 and complete column A, the length groups and column B, the corresponding frequencies from the available data set.

Example: Columns A and B in Table 3.4.1.1, taken from Table 3.2.1.1. Column B is labelled "N1+", because it contains the distribution of the first cohort (N1) plus all the other cohorts. In general, the symbol "Na+" stands for the number in the a'th plus older cohorts.

Step 2: Create column C by taking logarithms of the frequencies of N1+ (column B).

Examples: $\ln 1 = 0$
 $\ln 4 = 1.386$

Step 3: Column D contains the differences between the logarithms of two adjacent frequencies

$$\Delta \ln N1+ = \ln N1+ \text{ of the current line minus } \ln N1+ \text{ of the previous line}$$

Complete column D. Start on the second line, subtract the ln value of the first line of column C from that of the second line of column C and place it on the second line of column D. The first place remains open, since a difference between the first point and a foregoing point does not exist. Take care to use at least three decimal points. Continue by determining the differences ($\Delta \ln$) between the third and the second line, etc.

Step 4: Complete column E. Recall from Section 2.6 that $\Delta \ln N1+$ should be plotted against the upper limit of the smallest length group of the two from which $\Delta \ln N1+$ is calculated. Insert the mid-point, the upper limit of the smallest of the two classes, or the lower limit of the largest of the two at the same level as the corresponding $\Delta \ln N$.

See example Step 3.

Step 5: Make a complete plot of the length (column E, at the x-axis) against $\Delta \ln N1+$ (column D, at the y-axis).

Example: Fig. 3.4.1.1

Step 6: Inspect the plot and determine which points lie on a straight line. Mark these points in column E. Do not include points that may be affected by the next distribution. The further the points are lying to the right, the higher the chance is that they are influenced by the next distribution.

Example: Visual inspection of Fig. 3.4.1.1 shows that a straight line can be fitted to the first seven points (indicated by "*" in Table 3.4.1.1). Even the eighth point lies on the same line, but because it may be influenced by the next distribution it was not included in the subsequent calculations.

This straight line corresponds to the first normally distributed component, N1, which is interpreted as the 1983 spring cohort. That the straight line corresponding to N1 comes out so nicely is not surprising, since the first component has very little overlap with the next component, as can be observed in Figs. 3.2.2.1 and 3.2.2.2.

Table 3.4.1.1 Bhattacharya method: Estimation of first cohort, N1 (the 1983 spring cohort). # in columns B, C, G and H indicates where to start the calculations of N1 (cf. Fig. 3.4.1.2). Column I contains the remainder of the sample (N1+ - N1 = N2+)

A	B	C	D	E	F	G	H	I
L1-L2	N1+	ln N1+	$\Delta \ln N1+$ (y)	L (x)	$\Delta n N1$	ln N1	N1	N2+
12-13	1	0.000	-	-	-	-	1	0
13-14	4	1.386	1.386	13*	1.375	-	4	0
14-15	11	2.398	1.012	14*	1.059	-	11	0
15-16	24	3.178	0.780	15*	0.743	-	24	0
16-17	38#	3.638#	0.460	16*	0.427	3.638#	38#	0
17-18	42	3.738	0.100	17*	0.111	3.749	(42.48)	(-0.48)
18-19	33	3.497	-0.241	18*	-0.205	3.545	(34.61)	(-1.61)
19-20	20	2.996	-0.501	19*	-0.521	3.023	(20.55)	(-0.55)
20-21	7	1.946	-1.050	20	-0.837	2.186	(8.90)	(-1.90)
21-22	3	1.099	-0.847	21	-1.153	1.033	2.81	0.19
22-23	3	1.099	0.000	22	-1.469	-0.436	0.65	2.35
23-24	5	1.609	0.511	23	-1.785	-2.211	0.11	4.89
24-25	8	2.079	0.470	24	-	-	-	8
25-26	11	2.398	0.318	25	-	-	-	11
26-27	14	2.639	0.241	26	-	-	-	14
27-28	17	2.833	0.194	27	-	-	-	17
28-29	16	2.773	-0.060	28	-	-	-	16
29-30	15	2.708	-0.065	29	-	-	-	15
30-31	14	2.639	-0.069	30	-	-	-	14
31-32	11	2.398	-0.241	31	-	-	-	11
32-33	11	2.398	0.000	32	-	-	-	11
33-34	10	2.303	-0.095	33	-	-	-	10
34-35	9	2.197	-0.106	34	-	-	-	9
35-36	10	2.303	0.106	35	-	-	-	10
36-37	11	2.398	0.095	36	-	-	-	11
37-38	10	2.303	-0.095	37	-	-	-	10
38-39	10	2.303	0.000	38	-	-	-	10
39-40	11	2.398	0.095	39	-	-	-	11
40-41	11	2.398	0.000	40	-	-	-	11
41-42	9	2.197	-0.201	41	-	-	-	9
42-43	7	1.946	-0.251	42	-	-	-	7
43-44	7	1.946	0.000	43	-	-	-	7
44-45	5	1.609	-0.337	44	-	-	-	5
45-46	6	1.792	0.183	45	-	-	-	6
46-47	5	1.609	-0.183	46	-	-	-	5
47-48	3	1.099	-0.510	47	-	-	-	3
48-49	2	0.693	-0.406	48	-	-	-	2
49-50	2	0.693	0.000	49	-	-	-	2
50-51	2	0.693	0.000	50	-	-	-	2
51-52	1	0.000	-0.693	51	-	-	-	1

Class interval, dL = 1. Total number in cohort N1: 183.57

*) points used in the regression analysis, with results:

$$a = 5.4834, \quad b = -0.3160, \quad \bar{L}(N1) = -1/b = 17.35,$$

$$s(N1) = \sqrt{-dL/b} = 1.78$$

Example:

A	B	C	D	E
L1-L2	N1+	ln N1+	$\Delta \ln N1+$	L
12-13	1	0		
13-14	4	1.386	1.386-0 = 1.386	13
14-15	11	2.398	2.398-1.386 = 1.013	14

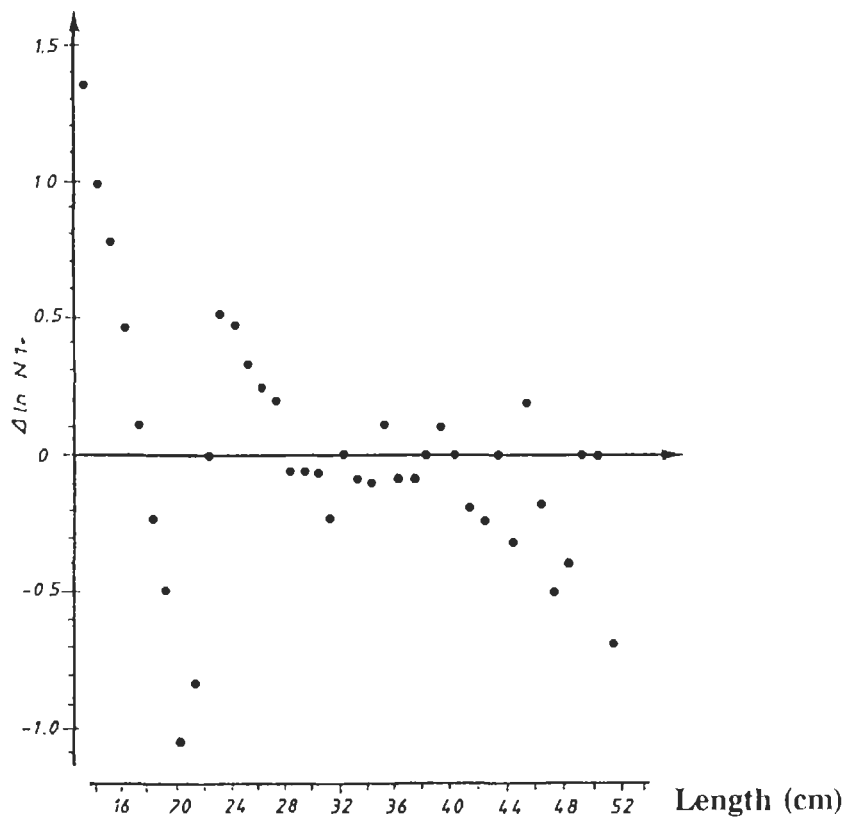


Fig. 3.4.1.1 Bhattacharya method: plot corresponding to columns D (y-axis) and E (x-axis) of Table 3.4.1.1

Step 7: Calculate the straight line that fits to the points by regressing column E against column D for the selected points (asterisks). Determine a (intercept) and b (slope) and calculate the mean length

$$\bar{L}(N1) = -a/b \text{ and the standard deviation } s(N1) = \sqrt{-1/b}$$

Example:

$$y = a + b \cdot x$$

where $y = \Delta \ln N1+$ (in column D) and $x = L$ (in column E)

$$a \text{ (intercept)} = 5.4834, b \text{ (slope)} = -0.3160$$

$$\bar{L}(N1) = -a/b = 17.35 \text{ cm and } s(N1) = \sqrt{-1/b} = 1.78$$

The regression line is shown in Fig. 3.4.1.2.

We have now determined the line that represents a normal distribution, which should to a large extent correspond with the left side of the actual distribution we have in our sample. The line should represent the first cohort, N1. In order to determine how far this is true we must first calculate the theoretical values of $\Delta \ln N1$, those corresponding to the line we have just determined and then reverse the process and convert the differences ($\Delta \ln N1$) into $\ln N1$ and then into numbers (N1). This process is illustrated in columns F, G and H of Table 3.4.1.1.

The second part of the computation procedure consists of the following steps:

Step 8: The formula $\Delta \ln N = a + bL$ can now be used to calculate the theoretical value $\Delta \ln N1$. This is done for as many length groups as one can expect to find in the first cohort (normal distribution).

Example:

$\Delta \ln N_1$ for the length groups 12-13 and 13-14 with mid length 13 is determined from

$$a + b \cdot 13 = 5.4834 - 0.3160 \cdot 13 = 1.375,$$

which is the first value in column F, the next value is

$$a + b \cdot 14 = 1.059, \text{ etc.}$$

Step 9: In order to be able to convert "a difference", $\Delta \ln N$, into its two components, $\ln N$ of a certain length group and $\ln N$ of the length group above it, we need a starting point. This starting point should be based on a frequency that is not contaminated by overlap with the following cohort (normal distribution). Therefore a frequency should be chosen on the left side of the first normal distribution. Preferably the frequency should also not be too low.

Example:

The frequency 38 of the length group 16-17 cm was chosen as the clean starting point, as indicated by "#". It is placed in column H as the first entry for N_1 , the numbers in the length-frequency distribution of the first cohort. The real starting point is actually the logarithm of 38, viz., 3.638 (see column C). This value is inserted in column G. The choice of 38 as a "clean" frequency also implies that the frequencies lying to its left, those above it in the table, viz., 1, 4, 11 and 24, are also considered to be clean. In other words none of these frequencies are supposed to overlap with the next cohort, so they are all clean frequencies of the first normal distribution N_1 (the 1983 spring cohort).

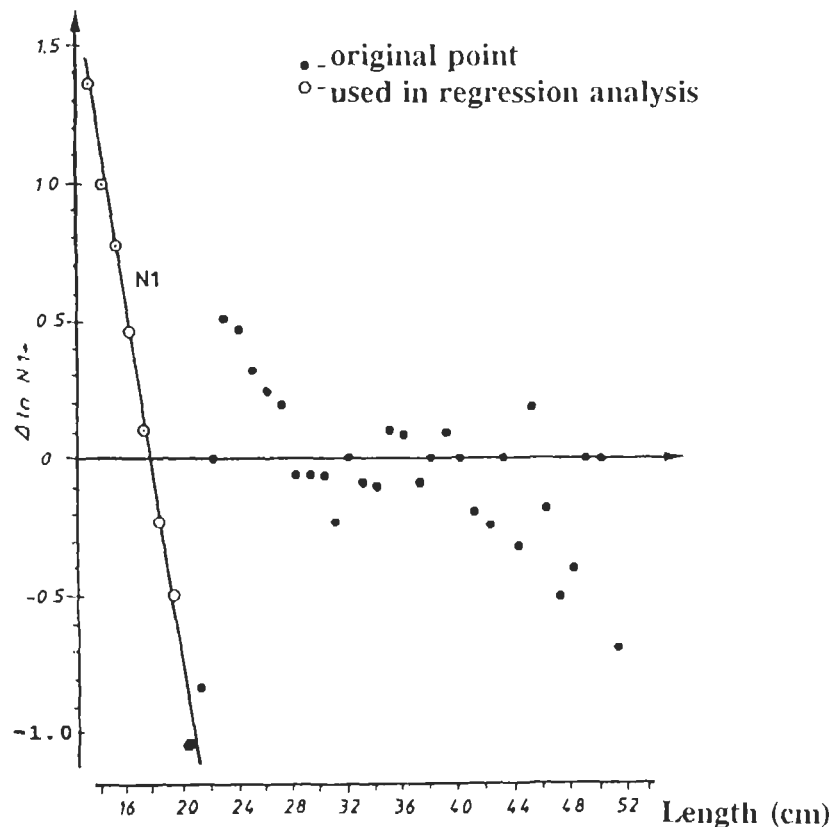


Fig. 3.4.1.2 Bhattacharya method: regression line estimated for the first cohort (compare to columns D and E in Table 3.4.1.1)

Step 10: We now have $\Delta \ln N_1$ corresponding to two adjacent length classes in column F and the first $\ln N_1$ of the lower length class in column G, which permits us to calculate the $\ln N_1$ of the next length class up using the formula

$$\ln N_1 (\text{upper length class}) = \ln N_1 (\text{lower length class}) + \text{the corresponding } \Delta \ln N_1$$

Example:

$$\begin{aligned} \ln N_1(17-18) &= \ln N_1(16-17) + \Delta \ln N_1(17-18 \text{ and } 16-17) \\ \ln N_1(17-18) &= 3.638 + 0.111 = 3.749 \\ \ln N_1(18-19) &= 3.749 + (-0.205) = 3.544 \end{aligned}$$

The new values are entered in column G.

Step 11: By taking antilogs the numbers corresponding to $\ln N_1$ in column G can be found and inserted in column H. This column is stopped when the number in column H approaches zero.

Examples:

$$\begin{aligned} \text{for length group 17-18: } N_1(17-18) &= \exp(3.749) = 42.48 \\ \text{for length group 18-19: } N_1(18-19) &= \exp(3.544) = 34.61 \end{aligned}$$

The results (in column H) are not exactly the same as the observed frequencies given in column B, because observations always deviate somewhat from the theoretical values. In the present case of a hypothetical data set, the deviations are due to rounding errors. With "real" data there are also deviations caused by "random noise". Even if the sample is a perfect random sample the observations will fluctuate around the true length distribution (the length distribution of the population).

Step 12: The numbers of fish per length group belonging to the youngest (1983 spring) cohort or N_1 , in column G, can now be subtracted from the Total distribution, or N_1+ , in column B. The new distribution obtained is placed in column I and called N_2+ , the frequency distribution of fish in the second cohort plus all the subsequent cohorts.

$$N_1+ \text{ minus } N_1 = N_2+ \text{ or column B - column H = column I}$$

In practice it may well happen that the figures in column I become negative because of random variation of the observations. However, this can be adjusted. Whenever the estimate of N_2+ numbers becomes negative we assign the value zero to N_2+ (in column I) while the N_1 is given the value of column B.

Examples:

$$42 - 42.48 = -0.48, \text{ which is adjusted to } 0 \text{ in column I, and to } 42 \text{ in column H}$$

$$33 - 34.61 = -1.61, \text{ which is adjusted to } 0 \text{ in column I and to } 33 \text{ in column H}$$

The results of the whole analysis of the first normal distribution will be:

A	B	H	I
L1-L2	N1+	N1	N2+
12-13	1	1	0
13-14	4	4	0
14-15	11	11	0
15-16	24	24	0
16-17	38	38	0
17-18	42	42	0
18-19	33	33	0
19-20	20	20	0
20-21	7	7	0
21-22	3	2.81	0.19
22-23	3	0.65	2.35
23-24	5	0.11	4.89
24-25	8	0	8
Total number in cohort N1 =		183.57	

$$\bar{L}(N1) = 17.35 \text{ and } s(N1) = 1.78$$

Since this is based on a constructed data set we can compare these results with the real values, which are given in Table 3.2.1.1 (spring 1983):

$$\text{Total number (N1)} = 182, \bar{L}(N1) = 17.3 \text{ and } s(N1) = 1.7$$

In this case the results obtained from the analysis are very close to the real values. What we have obtained are all the necessary elements to describe the first normal distribution, viz.,

$$\bar{L}(N1), s(N1) \text{ and } n(N1)$$

The whole process is now repeated in order to obtain those values for the next normal distribution, the one pertaining to the cohort that was born in the autumn of 1982 (see Table 3.2.1.1).

We have come to the end of the use of Table 3.4.1.1. By eliminating all values pertaining to N1 we create the next work sheet (Table 3.4.1.2) with N2+ (column I of Table 3.4.1.1) as the new column B. The whole procedure can then be repeated.

Fig. 3.4.1.3 shows the Bhattacharya plot for N2+ together with the estimated line for N1. Only the points to the right of the dotted line of Fig. 3.4.1.3 are used in the analysis now. The N1-line is shown for comparison only. Some points have been moved due to the subtraction of N1. The "old" points (i.e. those from the N1+ plot) are indicated by "x" and the "new" points by a triangle, in cases where the movement is visible. The two first points (corresponding to lengths 21 and 22 cm) are disregarded, because they refer to very small numbers of specimens.

The selection of points to fit a straight line is now a bit more difficult than in the case of the first cohort. In Fig. 3.4.1.3 the six points from lengths 23 to 28 cm were chosen. One can question why these points were chosen in preference to the points, for example, from lengths 24 to 29 cm or from lengths 24 to 28 cm. The choice is a subjective one. The results of the Bhattacharya method may sometimes be dependent on the person who actually performs the analysis. If, for example, only the points from lengths 24 to 27 cm were used the estimated mean length would be 28.7 cm and the standard deviation 3.2 cm. The actual choice made

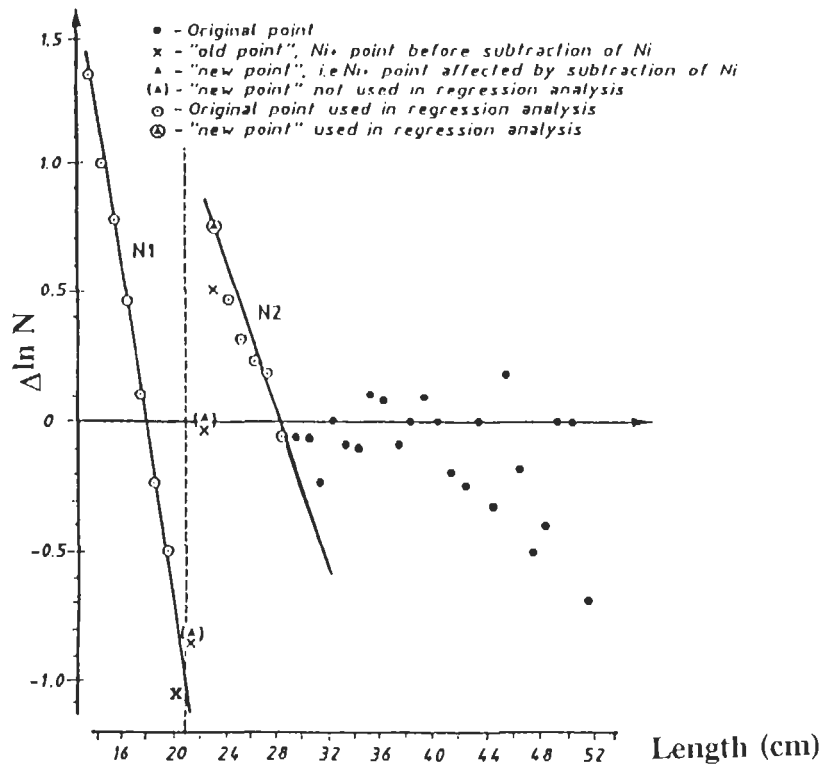


Fig. 3.4.1.3 Bhattacharya method: regression line estimated for the second cohort (compare columns D and E in Table 3.4.1.2)

in Fig. 3.4.1.3 gives a mean value of $\bar{L}(N2) = 27.77$ cm and a standard deviation $s(N2)$ of 2.66 cm which are both very close to the true values (see Tables 3.2.1.1 and 3.4.1.2). However, this cannot be used as a justification for this choice, because in real life we would not know what the true values should be. Also the selection of the "clean" value of $\ln N2$ from which $N2$ and $N3+$ are calculated is a subjective one. The more the observations deviate from the calculated frequencies the more pronounced the element of subjectivity.

In summary, the results obtained so far are:

cohort N1: mean length 17.35 cm, standard deviation 1.78 cm
(Table 3.4.1.1)

cohort N2: mean length 27.77 cm, standard deviation 2.66 cm
(Table 3.4.1.2)

Now that the first two mean lengths of cohorts have been estimated we are in a position to obtain a first rough estimate of the von Bertalanffy parameter K , provided we also have an estimate of the age difference between the two cohorts. We use Eqs. 3.4.0.1 and 3.4.0.2 with a time difference between the two cohorts equal to $t2-t1 = 0.5$ year. Further a rough estimate of L_∞ is obtained from the length-frequency sample, which tells us that the fish rarely get longer than 50 cm, so it is assumed that $L_\infty = 50$ cm. From Eq. 3.4.0.1 we get:

$$K = \frac{1}{t2-t1} * \ln \frac{L_\infty - \bar{L}(t1)}{L_\infty - \bar{L}(t2)} = \frac{1}{0.5} * \ln \frac{50 - 17.35}{50 - 27.77} = 0.77$$

and from Eq. 3.4.0.2:

$$t_o = t1 + \frac{1}{K} * \ln \left(1 - \frac{\bar{L}(t1)}{L_\infty} \right) = 0.5 + \frac{1}{0.77} * \ln \left(1 - \frac{17.35}{50} \right) = -0.05$$

Table 3.4.1.2 Bhattacharya method: Estimation of the second cohort, N2 (the 1982 autumn cohort). # in columns B, C, G and H indicates where to start the calculations of N2 (compare Fig. 3.4.1.3)

A	B	C	D	E	F	G	H	I
L1-L2	N2+	1n N2+	$\Delta 1n N2+$ (y)	L (x)	$\Delta 1n N2$	1n N2	N2	N3+
12-13	0	-	-	-	-	-	0	0
13-14	0	-	-	13	-	-	0	0
14-15	0	-	-	14	-	-	0	0
15-16	0	-	-	15	-	-	0	0
16-17	0	-	-	16	-	-	0	0
17-18	0	-	-	17	-	-	0	0
18-19	0	-	-	18	-	-	0	0
19-20	0	-	-	19	-	-	0	0
20-21	0	-	-	20	-	-	0	0
21-22	0.19	-1.661	-	21	-	-	0.19	0
22-23	2.35	0.854	2.515	22	-	-	2.35	0
23-24	4.89	1.587	0.733	23*	-	-	4.89	0
24-25	8	2.079	0.492	24*	-	-	8	0
25-26	11	2.398	0.319	25*	-	-	11	0
26-27	14#	2.639#	0.241	26*	0.248	2.639#	14#	0
27-28	17	2.833	0.194	27*	0.107	2.746	15.58	1.42
28-29	16	2.773	-0.060	28*	-0.034	2.712	15.06	0.94
29-30	15	2.708	-0.065	29	-0.175	2.537	12.64	2.36
30-31	14	2.639	-0.069	30	-0.316	2.221	9.22	4.78
31-32	11	2.398	-0.241	31	-0.457	1.764	5.84	5.16
32-33	11	2.398	0.000	32	-0.598	1.166	3.21	7.79
33-34	10	2.303	-0.095	33	-0.740	0.426	1.53	8.47
34-35	9	2.197	-0.106	34	-0.881	-0.455	0.63	8.37
35-36	10	2.303	0.106	35	-1.022	-1.477	0.23	9.77
36-37	11	2.398	0.095	36	-1.163	-2.640	0.07	10.93
37-38	10	2.303	-0.095	37	-	-	-	10
38-39	10	2.303	0.000	38	-	-	-	10
39-40	11	2.398	0.095	39	-	-	-	11
40-41	11	2.398	0.000	40	-	-	-	11
41-42	9	2.197	-0.201	41	-	-	-	9
42-43	7	1.946	-0.251	42	-	-	-	7
43-44	7	1.946	0.000	43	-	-	-	7
44-45	5	1.609	-0.337	44	-	-	-	5
45-46	6	1.792	0.183	45	-	-	-	6
46-47	5	1.609	-0.183	46	-	-	-	5
47-48	3	1.099	-0.510	47	-	-	-	3
48-49	2	0.693	-0.406	48	-	-	-	2
49-50	2	0.693	0.000	49	-	-	-	2
50-51	2	0.693	0.000	50	-	-	-	2
51-52	1	0.000	-0.693	51	-	-	-	1

Total number in cohort N2: 104.44
 *) points used in the regression analysis, with results:
 $a = 3.9168$, $b = -0.1411$, $\bar{L}(N2) = -a/b = 27.77$
 $s(N2) = \sqrt{-dL/b} = 2.66$

where the value $t_1 = 0.5$ is an arbitrary age.

Thus, as a first rough estimate of the growth curve we have

$$L(t) = 50 * [1 - \exp(-0.77 * (t + 0.05))]]$$

The estimation procedure presented above is not generally recommended. It has been given here to demonstrate how little data are actually required to roughly estimate a growth curve. Such a first estimate, however, may be used to predict the next mean length, i.e. the mean length of cohort N3.

Assuming cohort N3 to be 1.5 years old we get:

$$L(1.5) = 50 * [1 - \exp(-0.77 * (1.5 + 0.05))]] = 34.8 \text{ cm}$$

Table 3.4.1.3 Bhattacharya method: Estimation of the third cohort, N3 (the 1982 spring cohort). # in columns B, C, G and H indicates where to start the calculations of N3 (compare Fig. 3.4.1.4)

A	B	C	D	E	F	G	H	I
L1-L2	N3+	ln N3+	$\Delta \ln N3+$ (y)	L (x)	$\Delta \ln N3$ (x)	ln N3	N3	N4+
.....								
25-26	0	-	-	25	-	-	0	0
26-27	0	-	-	26	-	-	0	0
27-28	1.42	0.351	-	27	-	-	1.42	0
28-29	0.94	-0.062	-0.413	28	-	-	0.94	0
29-30	2.36	0.859	0.921	29	-	-	2.36	0
30-31	4.78	1.564	0.705	30*	-	-	4.78	0
31-32	5.16	1.641	0.077	31*	-	-	5.16	0
32-33	7.79#	2.053#	0.412	32*	-	2.053#	7.79#	0
33-34	8.47	2.137	0.084	33*	0.111	2.164	8.71	0
34-35	8.37	2.125	-0.012	34*	-0.032	2.132	8.43	0
35-36	9.77	2.279	0.154	35	-0.175	1.957	7.08	2.69
36-37	10.93	2.392	0.113	36	-0.318	1.639	5.15	5.78
37-38	10	2.303	-0.089	37	-0.460	1.179	3.25	6.75
38-39	10	2.303	0.000	38	-0.603	0.576	1.78	8.22
39-40	11	2.398	0.095	39	-0.746	-0.170	0.84	10.16
40-41	11	2.398	0.000	40	-0.888	-1.058	0.35	10.65
41-42	9	2.197	-0.201	41	-1.031	-2.089	0.12	8.88
42-43	7	1.946	-0.251	42	-1.174	-3.263	0.04	6.96
43-44	7	1.946	0.000	43	-	-	-	7
44-45	5	1.609	-0.337	44	-	-	-	5
45-46	6	1.792	0.183	45	-	-	-	6
46-47	5	1.609	-0.183	46	-	-	-	5
47-48	3	1.099	-0.510	47	-	-	-	3
48-49	2	0.693	-0.406	48	-	-	-	2
49-50	2	0.693	0.000	49	-	-	-	2
50-51	2	0.693	0.000	50	-	-	-	2
51-52	1	0.000	-0.693	51	-	-	-	1
Total number in cohort N3:							58.20	
*) points used in the regression analysis, with results:								
a = 4.8196, b = -0.1427, $\bar{L}(N3) = -a/b = 33.77$,								
s(N3) = $\sqrt{-dL/b} = 2.65$								

Table 3.4.1.3 has been prepared for the analysis of N3+ and the related Bhattacharya plot is shown in Fig. 3.4.1.4. The selection of points used for the regression for N3 is now even more questionable than the one made for cohort N2. However, the estimated mean value of $L(N3) = 33.8$ cm came out reasonably well compared to the value of 34.8 cm calculated above and which we happen to know is close to the true value of 35.3 cm (Table 3.2.1.1).

With three mean lengths we are now in a position to apply the von Bertalanffy plot (cf. Section 3.3.3, Eq. 3.3.3.1), again assuming the arbitrary age of 0.5 years for the first cohort. The input data for the estimation of K and t_0 from the von Bertalanffy plot are shown in Table 3.4.1.4, together with the results of the regression analysis. The estimate of $t_0 = -0.13$ year is an arbitrary value, since we used arbitrary ages. Nevertheless it puts us in a position to calculate length at other arbitrary ages because the shape of the growth curve is independent of t_0 . With the new growth parameters estimated in Table 3.4.1.4 the expected mean length of cohort N4 with arbitrary age 2.0 years becomes:

$$L(2.0) = 50 * [1 - \exp(-0.7 * (2.0 + 0.13))] = 38.7 \text{ cm}$$

We now continue with the Bhattacharya method to estimate N4. Table 3.4.1.5 and Fig. 3.4.1.5 show the Bhattacharya analysis for N4+. It is difficult to see a straight line. Selecting the five points corresponding to lengths 37 - 41 cm would give a mean length of 40.0 cm,

which is a reasonable value. (Actually, this value is very close to the true value of 40.2 cm (cf. Table 3.2.1.1), but we are not supposed to have that information.)

At this stage one should probably consider the fit of a straight line as being so poor that the analysis should be terminated. When to stop is largely a matter of taste, although some objective criteria for limitations of the Bhattacharya method can be devised as will be discussed in Section 3.5.4. Anyway, we stop here to bring this example to an end.

Table 3.4.1.4 Estimation of K and t_0 by the von Bertalanffy plot using arbitrary input ages and the mean lengths estimated in Tables 3.4.1.1 to 3.4.1.3 (compare Table 3.3.3.1)

t (x)	$\bar{L}(t)$	$-\ln(1-\bar{L}(t)/50)$ (y)
0.5	17.4	0.428
1.0	27.8	0.812
1.5	33.8	1.127

a (intercept) = 0.09
b (slope) = 0.699, $K = 0.7$ per year
 $t_0 = -a/b = -0.13$ year

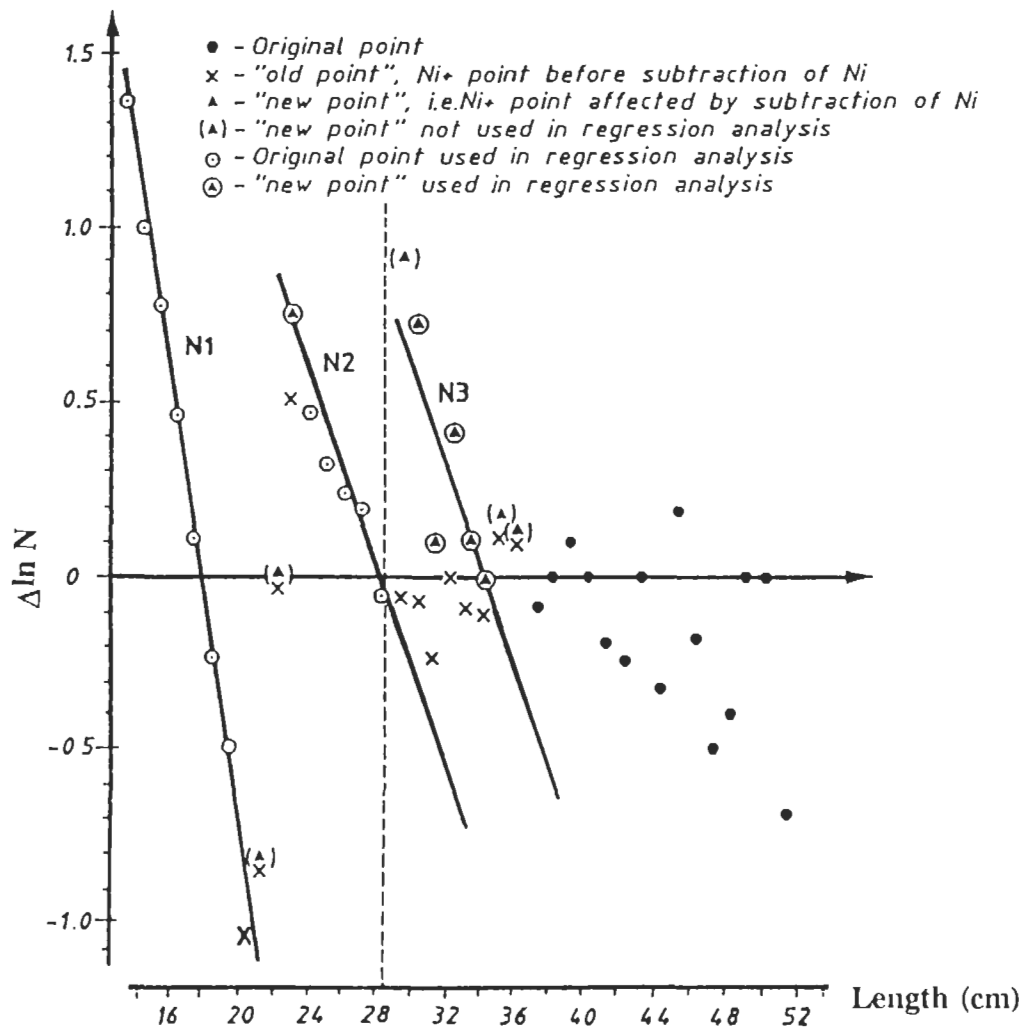


Fig. 3.4.1.4 Bhattacharya method: regression line estimated for the third cohort (compare columns D and E in Table 3.4.1.3)

Table 3.4.1.5 Bhattacharya method: Attempt to estimate the fourth cohort, N4 (the 1981 autumn cohort), compare Fig. 3.4.1.5

A	B	C	D	E	F	G	H	I
L1-L2	N4+	1n N4+	$\Delta 1n N4+$ (y)	L (x)	$\Delta 1n N4$	1n N4	N4	N5+
34-35	0	-	-	34	-	-	0	0
35-36	2.69	0.990	0.990	35	?	?	?	?
36-37	5.78	1.754	0.764	36	?	?	?	?
37-38	6.75	1.910	0.156	37 (*)	?	?	?	?
38-39	8.22	2.107	0.197	38 (*)	?	?	?	?
39-40	10.16	2.318	0.211	39 (*)	?	?	?	?
40-41	10.65	2.366	0.048	40 (*)	?	?	?	?
41-42	8.88	2.184	-0.184	41 (*)	?	?	?	?
42-43	6.96	1.940	-0.244	42	?	?	?	?
43-44	7	1.946	0.006	43	?	?	?	?
44-45	5	1.609	-0.337	44	?	?	?	?
45-46	6	1.792	0.183	45	?	?	?	?
46-47	5	1.609	-0.183	46	?	?	?	?
47-48	3	1.099	-0.510	47	?	?	?	?
48-49	2	0.693	-0.406	48	?	?	?	?
49-50	2	0.693	0.000	49	?	?	?	?
50-51	2	0.693	0.000	50	?	?	?	?
51-52	1	0.000	-0.693	51	?	?	?	?

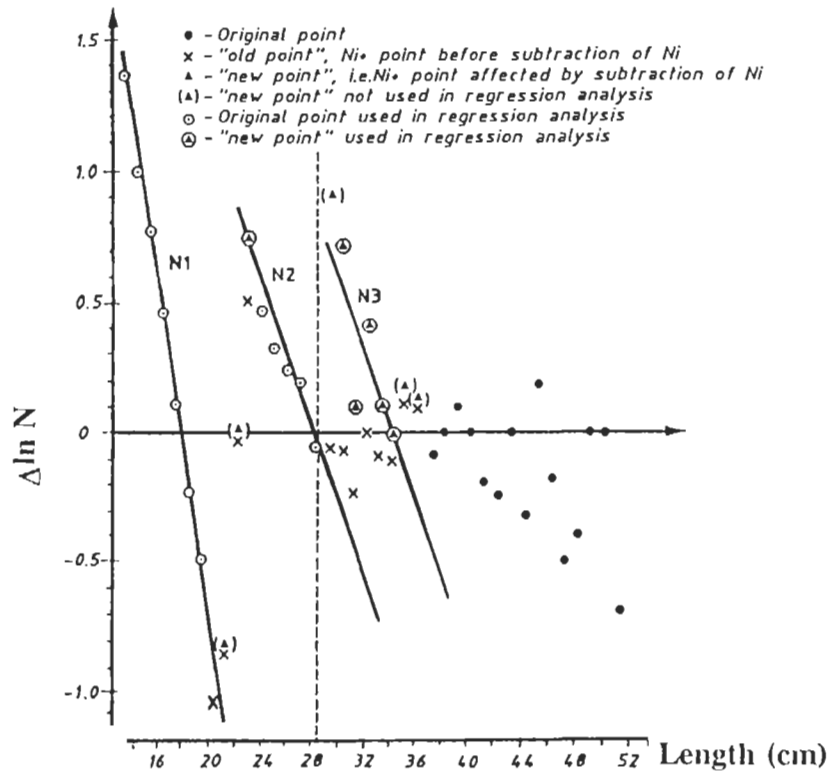


Fig. 3.4.1.5 Bhattacharya method: plot for estimation of the fourth cohort (compare columns D and E in Table 3.4.1.5). In this case the fit was considered too dubious

Bias

Input data for the Bhattacharya analysis are often biased due to gear selection and recruitment, i.e. the small fish are under-represented in the frequency samples, either because they escape through the meshes of the gear, or because they have not yet migrated from the nursery grounds to the fishing grounds (cf. Section 7.1). Aspects connected with bias caused by selection will be discussed in Chapter 6, where also a method to adjust length-frequency

samples for selection will be presented. In many cases the Bhattacharya analysis should be preceded by an adjustment for selection.

Another source of bias is observed for migratory fish species. Sometimes components are lacking because the cohort was not present in the area where the samples were taken. This aspect will be discussed in Chapter 11.

Computer programs

As you may have noticed, the Bhattacharya exercise takes some time to do by "paper-and-pencil". With the aid of a computer (which may be a microcomputer), however, the method is not hard to work with in practice.

The program "BHATTAC" in the LFSA package of microcomputer programs (see Chapter 15) closely follows the set-up explained in the foregoing. With a little experience you can do the exercise of Section 3.4.1 with BHATTAC in a few minutes. The program has a number of additional features: Whenever you have estimated a component BHATTAC displays a graph like Fig. 3.2.2.2 on the screen, allowing you to evaluate the fit to the original data. BHATTAC also checks whether your results are reasonable or not by calculating the "*separation index*", described in Section 3.5.4. Perhaps the most important feature of BHATTAC, compared to the "paper-and-pencil-method", is that it allows you to do the analysis several times, each time with a different set of input data. You may for example want to try out a range of alternative ways to fit the straight lines in the Bhattacharya plot.

One of the weak points of the "paper-and-pencil" version of the Bhattacharya method is the estimation of the numbers of fish in each cohort, since it is based on the subjective selection of one "clean point", from which the values of $\ln N_{a+}$ are calculated. A more rigorous statistical approach would be to apply all points used for the estimation of the regression line. In fact this more correct procedure is applied in BHATTAC.

When doing the Bhattacharya analysis on the computer you should always, as a matter of routine, try out different length class intervals (cf. Exercise 3.4.1), since it often happens that the structure of the points on the Bhattacharya plot emerges only for an optimal length class interval, which you may find simply by trying out various alternatives. Similar improvements may be obtained by pooling data over longer periods. In most cases you will be working with time series of length-frequencies (to be dealt with in Section 3.4.2), for example in the form of monthly length-frequency samples. You will then have the choice between, say, working with samples representing one month or to pool the data of three months to represent a quarter of the year. Such alternative aggregations of the basic data can easily be made by computer.

The "COMPLEAT ELEFAN" package contains a program "MPA", which also does the Bhattacharya exercise. FiSAT contains the same program.

Pauly and Caddy (1985), have developed a slightly different version of the Bhattacharya method for use with a programmable calculator. In their version the lines are determined by three successive points only, which are chosen so that they have the highest negative correlation coefficient. Their version is an attempt to turn the Bhattacharya method into an objective method, i.e. a method producing results independent of the person carrying out the analysis.

(See Exercise(s) in Part 2.)

3.4.2 Modal progression analysis

Example 10 used in Section 3.4.1 was based on one length-frequency sample collected during one survey. It was demonstrated that a somewhat rough estimate of the growth equation could be obtained from such a data set. L_{∞} and K could be estimated, whereas t_0 could only be determined relative to the arbitrary ages chosen for the cohorts.

Example 11: Modal progression analysis, based on the data of Example 4

Now suppose we had the type of data described in Example 4 in Section 3.2.1, i.e. length-frequency samples from each month or quarter during one or several years. The example illustrated in Fig. 3.2.1.1 consists of 12 length-frequency samples collected during surveys carried out in the months January, April, July and October during three years (1982 to 1984). Such a time series puts us in a much better position to estimate growth parameters than in the case of a single sample (October) as used in Example 10 to illustrate the Bhattacharya analysis.

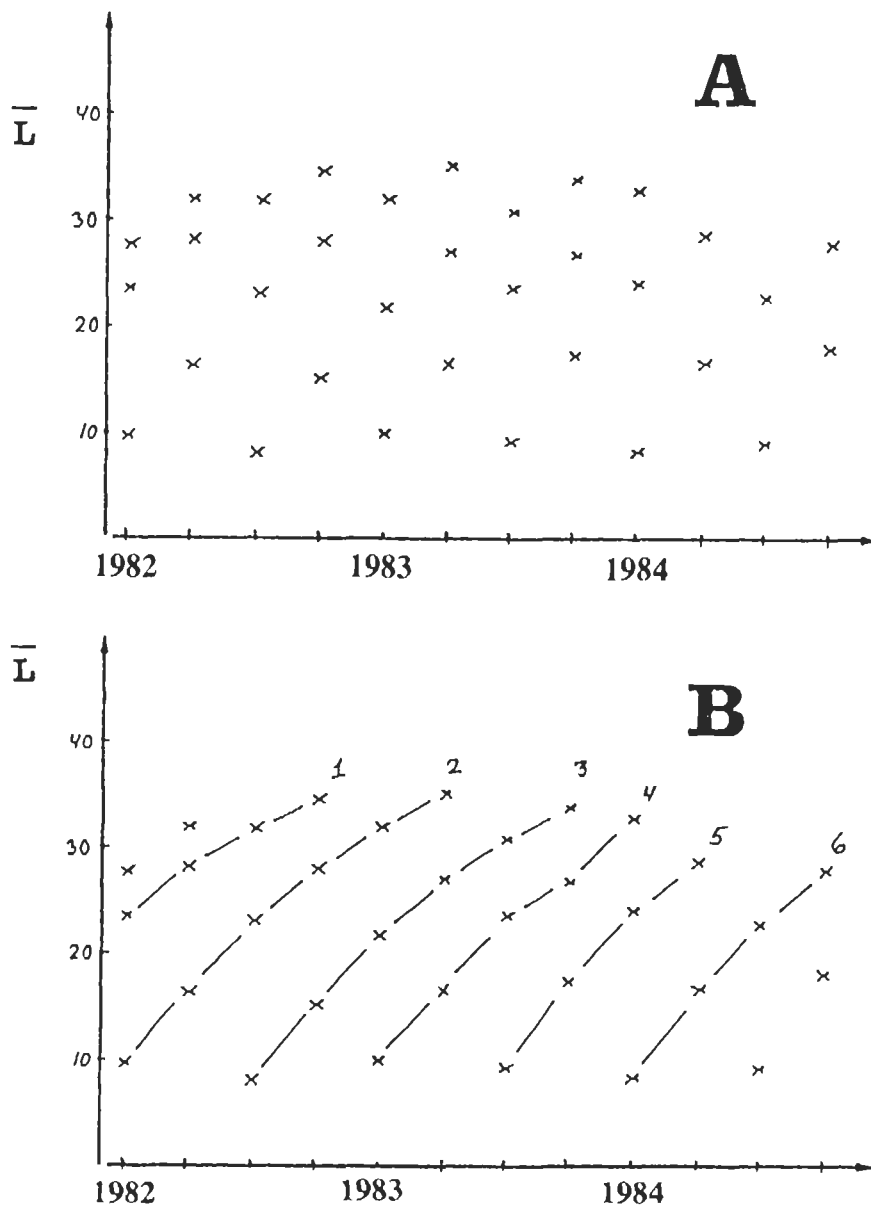


Fig. 3.4.2.1 Modal progression based on the results of the Bhattacharya analyses
 A: Mean lengths of components from Bhattacharya plots
 B: Mean lengths connected to represent growth curves of assumed cohorts

Table 3.4.2.1 Results of Bhattacharya analyses of the time series of length-frequency samples illustrated in Fig. 3.2.1.1

date of sample	third component	second component	first component
JAN 82	27.9	23.5	9.8
APR 82	32.0	28.1	16.5
JUL 82	31.8	23.1	8.0
OCT 82	34.6	28.0	15.3
JAN 83	32.0	21.8	10.0
APR 83	35.1	27.0	16.5
JUL 83	30.9	23.5	9.2
OCT 83 *)	33.8	27.8	17.4
JAN 84	32.9	24.0	8.3
APR 84	-	28.2	16.8
JUL 84	-	22.9	9.0
OCT 84	-	27.9	18.0

*) from Table 3.4.1.4

Table 3.4.2.2 The mean lengths from Table 3.4.2.1 rearranged into cohorts (see Fig. 3.2.1.1)

date of sample	COHORTS, $\bar{L}(t)$ in cm					
	1 spring 1981	2 autumn 1981	3 spring 1982	4 autumn 1982	5 spring 1983	6 autumn 1983
JAN 82	23.5	9.8	-	-	-	-
APR 82	28.1	16.5	-	-	-	-
JUL 82	31.8	23.1	8.0	-	-	-
OCT 82	34.6	28.0	15.3	-	-	-
JAN 83	-	32.0	21.8	10.0	-	-
APR 83	-	35.1	27.0	16.5	-	-
JUL 83	-	-	30.9	23.5	9.2	-
OCT 83	-	-	33.8	27.8	17.4	-
JAN 84	-	-	-	32.9	24.0	8.3
APR 84	-	-	-	-	28.2	16.8
JUL 84	-	-	-	-	-	22.9
OCT 84	-	-	-	-	-	27.9

Suppose that each of the twelve samples of the time series is given the same treatment as the single October 1983 sample. The results of the twelve Bhattacharya analyses could then be those given in Table 3.4.2.1. In each of the first nine samples three components have been found (as was the case in Section 3.4.1), whereas the last three samples were more difficult to analyse so that only two components could be identified. Please note that the number of components (cohorts) that could be identified is much lower than the actual number present, which are represented by dots in Fig. 3.2.1.1.

We may assume that the various cohorts remain in the sea for some time and thus that they are sampled at different stages of growth from the time of recruitment to the fishing (or sampling) area to their extinction. We may also assume that a mean length of a cohort, as determined for example by the Bhattacharya method, will correspond to a somewhat larger mean length in a sample taken a few months later and so forth. By plotting those mean lengths from a series of samples against a time axis and connecting them a growth curve can be obtained.

In Fig. 3.4.2.1A the mean lengths of the components have been plotted against the sample date. In Fig. 3.4.2.1B those mean lengths which we believe to correspond to the same cohorts, have been connected. Excluding the two first and the two last points we have thus identified six cohorts.

The connection of points to produce cohorts is a subjective process although in the present case the choice appeared quite easy to make. In practice it may not always be so simple.

It appears from Fig. 3.4.2.1B that there are two cohorts per year, for instance cohorts No. 3 and No. 4 which recruited in 1982. Assuming seasons of the northern hemisphere, No. 3 will be called the 1982 spring cohort and No. 4 the autumn cohort. The various growth curves drawn for each cohort enable us to interpret and rearrange the results of the twelve Bhattacharya analyses (Table 3.4.2.1) by cohorts as shown in Table 3.4.2.2.

The estimation of K and L_{∞}

The data in Table 3.4.2.2 are of the type that make it possible to apply the Gulland and Holt plot (cf. Section 3.3.1) by calculating:

$$\frac{\Delta L}{\Delta t} = \frac{L(t+\Delta t) - L(t)}{(t+\Delta t) - t} \quad \text{and}$$

$$\bar{L}(t, t+\Delta t) = \frac{L(t) + L(t+\Delta t)}{2}$$

The time difference $\Delta t = 0.25$ years, remains constant in this case, so it would also be possible to apply Chapman's method (Eq. 3.3.2.2).

The values of $\Delta L/\Delta t$ and $\bar{L}(t, t+\Delta t)$ are shown in Table 3.4.2.3. To illustrate the calculations we consider cohort No. 1, recruited in the spring of 1981 (see Fig. 3.4.2.1). For the two first samples we get:

$$\frac{\Delta L}{\Delta t} = \frac{L(\text{Apr } 82) - L(\text{Jan } 82)}{t(\text{Apr } 82) - t(\text{Jan } 82)} = \frac{28.1 - 23.5}{0.25} = 18.4 \quad \text{and}$$

$$\bar{L}(\text{Jan } 82, \text{Apr } 82) = \frac{L(\text{Jan } 82) + L(\text{Apr } 82)}{2} = (23.5 + 28.1)/2 = 25.8$$

It would be possible to make separate Gulland and Holt plots for each of the six cohorts, each with three to five points only. However, under the assumption that the growth parameters remain constant over the entire sampling period, all the 23 data pairs given in Table 3.4.2.3 may be combined into one single Gulland and Holt plot.

The regression of all 23 $\Delta L/\Delta t$ values on $\bar{L}(t, t+\Delta t)$ values gives the following results:

$$a \text{ (intercept)} = 41.84 \text{ and } b \text{ (slope)} = -0.8740$$

from which we get

$$L_{\infty} = -a/b = 47.9 \text{ say } 48 \text{ cm and}$$

$$K = -b = 0.87 \text{ per year with a 95\% confidence interval [0.72, 1.02]} \\ \text{(see Table 3.4.2.3)}$$

The Gulland and Holt plot is shown in Fig. 3.4.2.2. Estimates of L_{∞} and K have thus been obtained based on the entire time series.

Table 3.4.2.3 Input data and regression analysis for Gulland and Holt plot derived from Table 3.4.2.2. Note $\Delta t = 0.25$ year

sample date	cohort											
	1 spring 1981		2 autumn 1981		3 spring 1982		4 autumn 1982		5 spring 1983		6 autumn 1983	
	\bar{L}	$\frac{\Delta L}{\Delta t}$	\bar{L}	$\frac{\Delta L}{\Delta t}$	\bar{L}	$\frac{\Delta L}{\Delta t}$	\bar{L}	$\frac{\Delta L}{\Delta t}$	\bar{L}	$\frac{\Delta L}{\Delta t}$	\bar{L}	$\frac{\Delta L}{\Delta t}$
JAN 82	25.8	18.4	13.2	26.8		-		-		-		-
APR 82	30.0	14.8	19.8	26.4		-		-		-		-
JUL 82	33.2	11.2	25.6	19.6	11.7	29.2		-		-		-
OCT 82		-	30.0	16.0	18.6	26.0		-		-		-
JAN 83		-	33.6	12.4	24.4	20.8	13.3	26.0		-		-
APR 83		-		-	29.0	15.6	20.0	28.0		-		-
JUL 83		-		-	32.4	11.6	25.7	17.2	13.3	32.8		-
OCT 83		-		-		-	30.4	20.4	20.7	26.4		-
JAN 84		-		-		-		-	26.1	16.8	12.6	34.0
APR 84		-		-		-		-		-	20.0	24.4
JUL 84		-		-		-		-		-	25.4	20.0
OCT 84		-		-		-		-		-		

$a = 41.84$ $b = -0.8740$
 $sb^2 = \frac{1}{n-2} * ((sy/sx)^2 - b^2) =$
 $\frac{1}{23-2} * ((6.62333/7.08467)^2 - (-0.8740)^2) = 0.005242$
 $K = -b = 0.87$ per year, $L_\infty = -a/b = 47.9$ cm
 $sb = 0.072$, $t_{21} = 2.09$ (see Table 2.3.1)
 95% confidence interval of b ($= -K$): $[-1.02, -0.72]$ and of K : $[0.72, 1.02]$

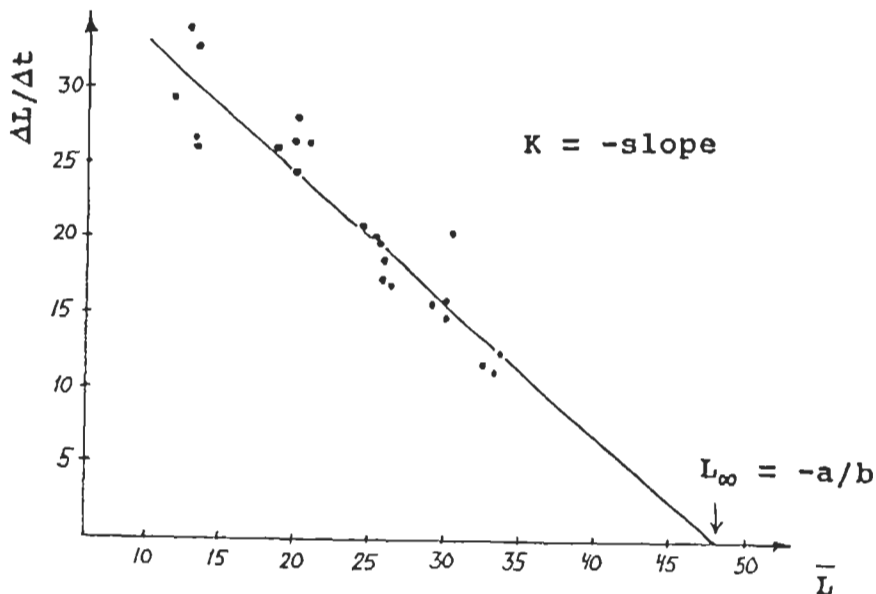


Fig. 3.4.2.2 Gulland and Holt plot based on data in Table 3.4.2.3

The estimation of t_0

The next step is to estimate the arbitrary initial condition parameters t_{01} for the spring cohorts and t_{02} for the autumn cohorts using the von Bertalanffy plot. We allot an arbitrary age of one year to the spring cohort of 1981 in January 1982, 1.25 years in April 1982, etc. The spring cohort of 1982 No. 3 is similarly allotted an age of one year in January 1983, etc. the procedure is the same for the autumn cohorts.

Table 3.4.2.4 contains the arbitrary ages, $t(i)$ of each cohort together with the dependent variable of the von Bertalanffy plot:

$$y = -\ln(1 - \bar{L}(t)/L_{\infty})$$

Values of $\bar{L}(t)$ are taken from Table 3.4.2.2. There are two regression analyses to be carried out:

$$\begin{aligned} \text{Spring cohorts: } & y = -K*t_{01} + K*t(i) , i = 1,3,5 \\ \text{Autumn cohorts: } & y = -K*t_{02} + K*t(i) , i = 2,4,6 \end{aligned}$$

where $t(i)$, the independent variable, is the arbitrary age of cohort no. i , as defined in Table 3.4.2.4. In this case six cohorts are considered simultaneously, and we believe that there are three spring cohorts and three autumn cohorts. As shown in Table 3.4.2.4 the two regression analyses gave the results:

	a (intercept)	b (slope)	$t_{01} = -a/b$ year	K (per year)
Spring cohorts:	-0.2055	0.8433	0.24	0.84
Autumn cohorts:	-0.7305	0.9169	0.80	0.92

As expected, the difference between t_{01} and t_{02} became close to half a year, as explained in Section 3.2.1 (see Table 3.2.1.2) for this example. The mean of the two K-values is 0.88 (close to the value of 0.87 estimated from the Gulland and Holt plot). A statistical test would show that the two estimates are not significantly different and we would therefore use the common value $K = 0.88$ per year. Thus the two equations:

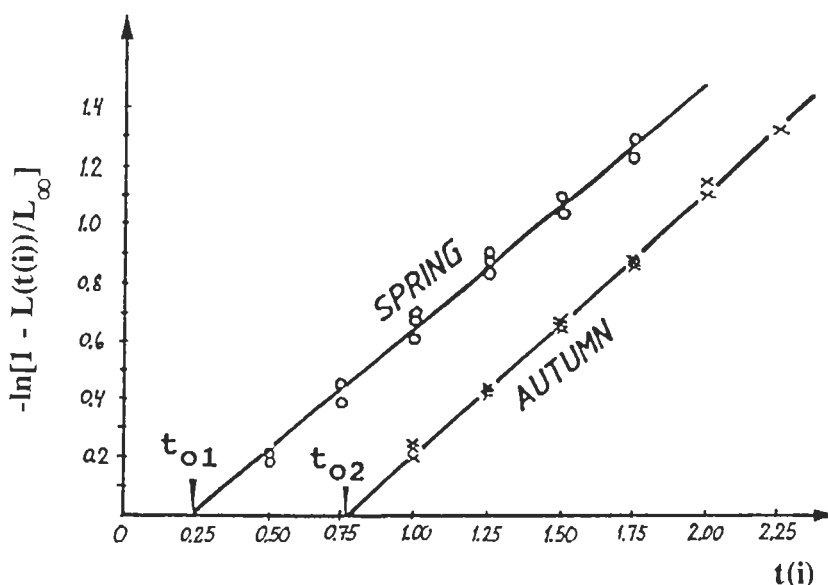


Fig. 3.4.2.3 The two von Bertalanffy plots based on data from Table 3.4.2.4

Spring cohorts: $L(t) = 48*[1 - \exp(-0.88*(t-0.24))]$

Autumn cohorts: $L(t) = 48*[1 - \exp(-0.88*(t-0.80))]$

can be used to calculate the length of spring cohorts and autumn cohorts for different arbitrary ages. We may stop the analysis at this stage, or we may continue trying to estimate the birthday of the cohorts.

Table 3.4.2.4 Input data and regression analysis for von Bertalanffy plot. Mean lengths of the components, $L(t)$ derived from Table 3.4.2.2, $L_{\infty} = 48$ cm

A: spring cohort							
date of sample	no. 1 spring 1981		no. 3 spring 1982		no.5 spring 1983		time of sampling T = (x)
	t(1)	y *)	t(3)	y *)	t(5)	y *)	
JAN 82	1.00	0.673	-	-	-	-	1982.00
APR 82	1.25	0.880	-	-	-	-	1982.25
JUL 82	1.50	1.086	0.50	0.182	-	-	1982.50
OCT 82	1.75	1.276	0.75	0.384	-	-	1982.75
JAN 83	-	-	1.00	0.605	-	-	1983.00
APR 83	-	-	1.25	0.827	-	-	1983.25
JUL 83	-	-	1.50	1.032	0.50	0.213	1983.50
OCT 83	-	-	1.75	1.218	0.75	0.450	1983.75
JAN 84	-	-	-	-	1.00	0.693	1984.00
APR 84	-	-	-	-	1.25	0.886	1984.25
JUL 84	-	-	-	-	-	-	1984.50
OCT 84	-	-	-	-	-	-	1984.75
spring cohorts: n = 14 a = -0.2055, b = 0.8433, so K = 0.84 per year $t_{o1} = -a/b = 0.24$ year $sb^2 = \frac{1}{n-2} * [(sy/sx)^2 - b^2] = \frac{1}{14-2} * [(0.35016/0.41313)^2 - 0.84332^2] = 0.0006005$ sb = 0.0245, $t_{12} = 2.18$ (see Table 2.3.1) 95% confidence interval of b (= K): [0.79 , 0.90]							
B: autumn cohorts							
date of sample	no. 2 autumn 1981		no. 4 autumn 1982		no. 6 autumn 1983		time of sampling T = (x)
	t(2)	y *)	t(4)	y *)	t(6)	y *)	
JAN 82	1.00	0.228	-	-	-	-	1982.00
APR 82	1.25	0.241	-	-	-	-	1982.25
JUL 82	1.50	0.656	-	-	-	-	1982.50
OCT 82	1.75	0.875	-	-	-	-	1982.75
JAN 83	2.00	1.099	1.00	0.234	-	-	1983.00
APR 83	2.25	1.314	1.25	0.421	-	-	1983.25
JUL 83	-	-	1.50	0.673	-	-	1983.50
OCT 83	-	-	1.75	0.866	-	-	1983.75
JAN 84	-	-	2.00	1.157	1.00	0.190	1984.00
APR 84	-	-	-	-	1.25	0.431	1984.25
JUL 84	-	-	-	-	1.50	0.648	1984.50
OCT 84	-	-	-	-	1.75	0.870	1984.75
autumn cohorts: n = 15 a = -0.7305, b = 0.9169, so K = 0.92 per year $t_{o2} = -a/b = 0.80$ year $sb^2 = \frac{1}{n-2} * [(sy/sx)^2 - b^2] = \frac{1}{15-2} * [(0.36593/0.39491)^2 - 0.91692^2] = 0.0001375$ sb = 0.037, $t_{13} = 2.16$ (see Table 2.3.1) 95% confidence interval of b (= K): [0.84 , 1.00]							
*) $y = -\ln(1 - \bar{L}(t)/L_{\infty})$							

Estimation of the birthday

To estimate the birthday, the idea is to extrapolate the growth curve beyond the first data point and see where it intersects with the time axis as illustrated in Fig. 3.4.2.4. This figure shows cohort no. 3 as an example. The curve cuts the time axis at the point 1982.24. On the arbitrary age axis the intersection point is $t_{01} = 0.24$. The point 1982.24 (29th of March) must be somewhere in the neighbourhood of the birthday. Because the von Bertalanffy growth curve does not conform to the early life stages of fish (cf. Section 3.1) this is an approximation. An alternative way of finding the approximate birthday is to use gonadal maturity stage data.

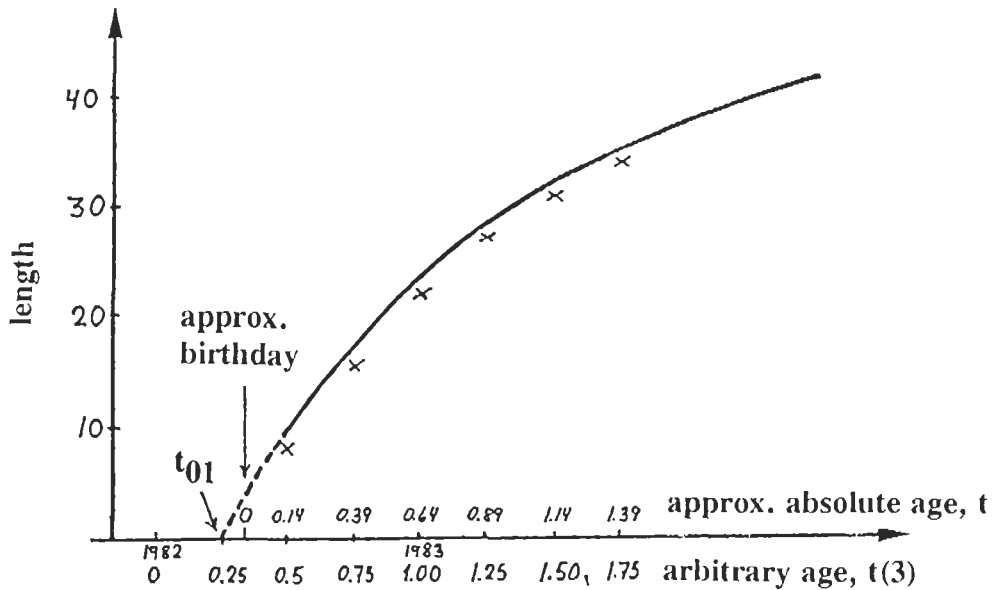


Fig. 3.4.2.4 Illustration of how the approximate birthday is estimated

The use of data on gonadal maturity

Another method of estimating the birth day is to estimate the spawning season from maturity stages of the adults. Fig. 3.4.2.5 shows an example of maturity stage data (from Wyatt, 1983). In this case the percentages of the three main stages of gonadal maturity are presented.

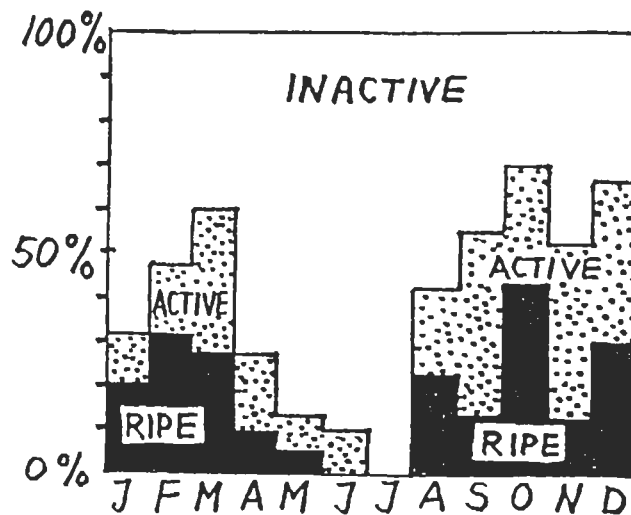


Fig. 3.4.2.5 Maturation stages observed for the squirrel fish *Holocentrus rufus*) from Wyatt (1983). Based on samples of 1331 fish

From maturation stage data, e.g., a graph of the percentage of ripe fish, we can define one (or two) mean spawning day(s), in the same way as the mean recruitment day was defined in Chapter 1, if the graph is unimodal (or bimodal). The histogram for the percentage of ripe fish in Fig. 3.4.2.6 could be interpreted as two spawning seasons with peaks in February and October. The mean spawning day may then be used as an estimate of the birth day (perhaps corrected for a time lag). However, the results of such analyses should be treated with a certain reservation as fluctuations in spawning are not the only factor which determine the fluctuations of recruitment. The success of a larva to feed and grow into a recruit and at the same time to avoid being eaten by predators is a complex process affected by a variety of environmental (biotic and abiotic) factors. The survival rate could, for instance, be almost nil for one spawning season and high for another. For a discussion of these matters see, for example, Bakun *et al.* (1982).

The application of modal progression analysis

The estimates obtained by following the progression of the modes (= cohorts) in the length-frequencies is considered superior to the method based on one single sample (Section 3.4.1). Further, there are cases where the single sample approach is not applicable at all. This is the case for short-lived species where there is only one (or only two) cohorts in a length-frequency sample. Such an example is shown in Fig. 3.4.2.6. It deals with commercial catches of the shrimp *Penaeus semisulcatus* in Kuwait waters (from Mohamed *et al.*, 1979). This species has a life span of one to two years and there are two cohorts per year. Most of the samples contain only one mode, so that the single sample approach is not applicable. However, to follow the progression of the modes appears a simple thing in this case. Modal progression analysis is especially useful for such short-lived species.

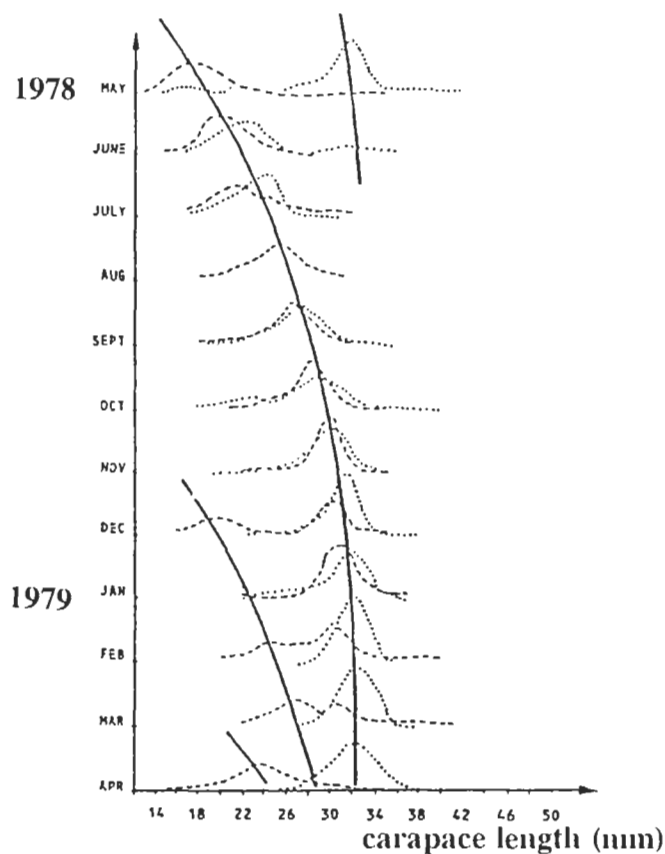


Fig. 3.4.2.6 Example of modal progression analysis. Size distributions of catches of *Penaeus semisulcatus* in the artisanal (—) and industrial (...) catches in Kuwait waters. (From Mohamed *et al.*, 1979)

Computer programs

The program "MODALPR" in the LFSA package can execute the modal progression analysis as described above. The LFSA package also allows you to continue from the Bhattacharya analysis (program "BHATTAC") with a least squares estimation of the growth parameters (program "VONBER", cf. Section 3.3.4) instead of the Gulland and Holt plot. The "COMPLEAT ELEFAN" package contains a program "MPA" to do the modal progression analysis. A similar program has been incorporated in FiSAT. There are several other computer programs available which attempt to solve the problem dealt with in this section, some of which will be discussed in Section 3.5.

Data massage

Running the Bhattacharya analysis and the modal progression analysis on a computer, one should always as a routine try out different aggregations of data, i.e. the so-called "*data-massage*" or "data-squeezing". Table 3.4.2.5 illustrates the process of data-massage. Part A contains the original data, i.e. a time series of fourteen monthly length-frequency samples grouped into sixteen 1-cm groups. From part A to part B the data have been squeezed into eight 2-cm length groups. From part B to part C data have been further squeezed into five 3-monthly groups. Sometimes a data-massage makes the structure of the data more apparent. (With "structure" is meant the straight lines in the Bhattacharya plots and the modal progression.)

If the data are grouped in such small classes that the "random noise" within each cell of the table hides the structure of the data one should massage the data. We may also observe the opposite problem, namely that the data are grouped in class intervals which are too large so that the structure becomes concealed behind the grouping. If your basic data are grouped in such large class intervals (in length or in time) there is nothing you can do to solve the problem. Therefore you should always record your basic data in as fine a grouping as practical. For example, if you are in doubt whether to use 1-cm groups or 2-cm groups, then use 1-cm groups. You can easily convert 1-cm groups into 2-cm groups, whereas you cannot do the opposite transformation. The grouping of data often simply has to be "just right" before you can successfully carry out a combined Bhattacharya/Modal Progression Analysis.

(See Exercise(s) in Part 2.)

3.4.3 The probability paper and parabola methods

There are other ways of analysing composite normal distributions which, like the Bhattacharya analysis, are basically paper-and-pencil methods and contain a certain amount of subjectivity.

One is the probability paper method introduced by Harding (1949) and further developed by Cassie (1954). It is based on the fact that a normal distribution becomes linear when plotted on probability paper. A mixture of several normal distributions provides a more complex line with inflexion points. As with the Bhattacharya method the individual normal distributions can be removed one by one.

Another approach is the parabola method introduced by Hald (1952) and used in fisheries research by Tanaka (1953). The mathematical base is the transformation of a normal distribution into a parabola by taking logarithms, see Section 2.6, Eq. 2.6.3. With this

method, parabolas are fitted to the log-transformed numbers of composite length-frequency data. The procedure is otherwise as with the Bhattacharya method which is a more sophisticated version based on the fact that differences between equi-distanced points on a parabola form a straight line.

The Bhattacharya method seems to leave less to subjective decisions on the researcher's part than the other methods do. However, persons skilled in the application of either the probability paper method or the parabola method also seem to reach plausible results.

Table 3.4.2.5 Illustration of the process of "data-massage". For further explanation, see text

A: BASIC DATA: 1-cm length groups by month

Length class	----- 1981 -----												----- 1982 -----			
	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR		
4- 5																
5- 6																
6- 7																
7- 8																
8- 9				18	24	12										
9-10				21	51	16										
10-11																
11-12																
12-13																
13-14																
14-15																
15-16																
16-17																
17-18																
18-19																
19-20																

B: MASSAGED DATA: 2-cm length groups by month

Length class	----- 1981 -----												----- 1982 -----			
	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR		
4- 6																
6- 8																
8-10				39	75	28										
10-12																
12-14																
14-16																
16-18																
18-20																

C: MASSAGED DATA: 2-cm length groups by 3 months

Length class	--- 1981 ---				--- 1982 ---			
	MAR	JUN	SEP	DEC	MAR	MAY	AUG	NOV
4- 6								
6- 8								
8-10		142						
10-12								
12-14								
14-16								
16-18								
18-20								

3.5 FITTING GROWTH CURVES BY MEANS OF COMPUTER PROGRAMS

The methods presented in Section 3.4, the "paper-and-pencil" methods and their computer-based counterparts basically treat the data sample by sample. Often the tracing of the growth curves becomes easier when the entire time series is considered. Some samples may be easy to resolve into cohort components and to interpret in terms of growth in an unambiguous way. By using the findings from the "easy" samples we may also be able to give unambiguous interpretations of samples we would otherwise not be able to interpret.

Figs. 3.5.0.1 and 3.5.0.2 illustrate this feature. The January sample in Fig. 3.5.0.1 seems easy to resolve into two components as shown in Fig. 3.5.0.2, whereas the September sample shows no structure whatsoever. The May sample appears more problematic than the January sample, but it is still possible to interpret. However, together the January and the May samples show a clear picture from which a growth curve can be estimated. By extrapolating the growth curve to the September sample we are now also in a position to split that into cohorts.

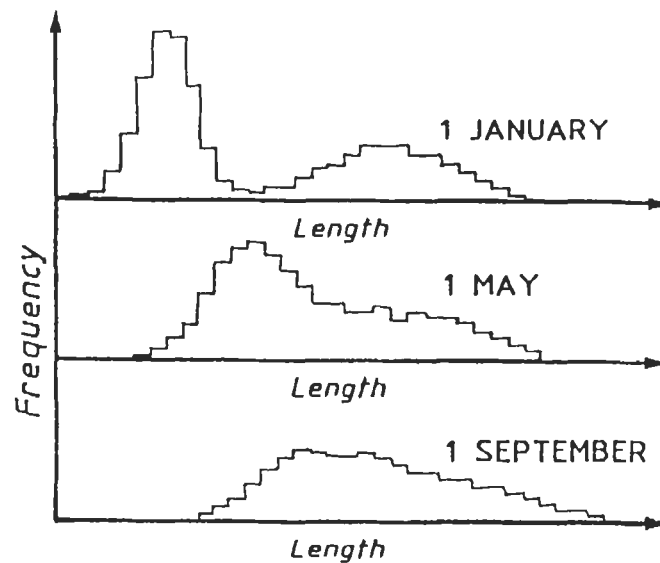


Fig. 3.5.0.1 Examples of an "easy" sample (January) and a "difficult" sample (September)

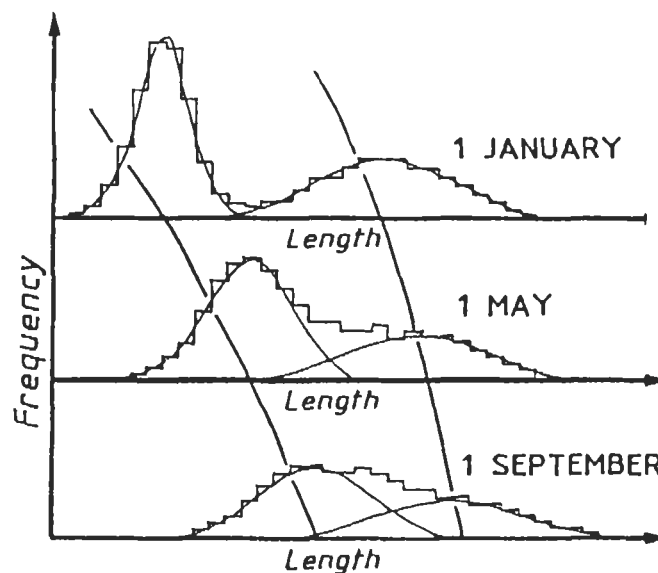


Fig. 3.5.0.2 Hypothetical example of how an "easy" sample (January) is used to treat a "difficult" sample (September)

This approach may be applied when using the "paper-and-pencil" method, especially when aided by a computer. It is however possible to leave more work to the computer and to let it do the analysis using a more sophisticated technique, such as a least squares estimation technique, (cf. Section 3.3.4).

The computer-based methods to be dealt with here require so many computations that it is almost impossible to do them by paper-and-pencil. We present two alternative approaches:

1. The "ELEFAN I" method (Electronic Length-Frequency ANalysis)
2. The "maximum-likelihood" method.

The first was introduced by Pauly and David (1981). The second may be considered a computerized version of the Bhattacharya method. It is based on the traditional theory on statistical analysis of frequency samples - a method which you may consider a generalized version of linear regression analysis. The basic philosophies behind the two methods are similar.

A detailed discussion of computer-based methods is considered outside the scope of the manual. The main purpose is to present some basic features of the methods which hopefully encourage the reader to go into further studies in this field.

3.5.1 ELEFAN I

The "ELEFAN I" program deals with estimation of growth parameters using length-frequency analysis (Pauly and David, 1981; and Pauly, 1987). The most recent description of the entire package will be found in Pauly (1987).

Example 12: The application of ELEFAN I to the coral trout data

To illustrate ELEFAN I we use the data on coral trout shown in Fig. 3.4.0.2. ELEFAN I consists of two major stages:

Stage 1: Restructuring of length-frequencies

Stage 2: Fitting of a growth curve

Stage 1, the restructuring process is illustrated in Fig. 3.5.1.1 where part "a" shows the original data as presented by Goeden (1978) in 0.5 cm length groups. To smooth out small irregularities the data have been rearranged in 2 cm length groups as shown in part "b". The curve in part "b" is the "moving average frequency" over 5 length groups. The method to obtain a moving average is illustrated for the length interval 26-28 cm:

interval	frequency	
18-20	0 *	
20-22	0 *	
22-24	2	} >
24-26	11	
26-28	15	
28-30	6	
30-32	10	
.....		

$$\text{moving average} = \frac{2+11+15+6+10}{5} = 8.8$$

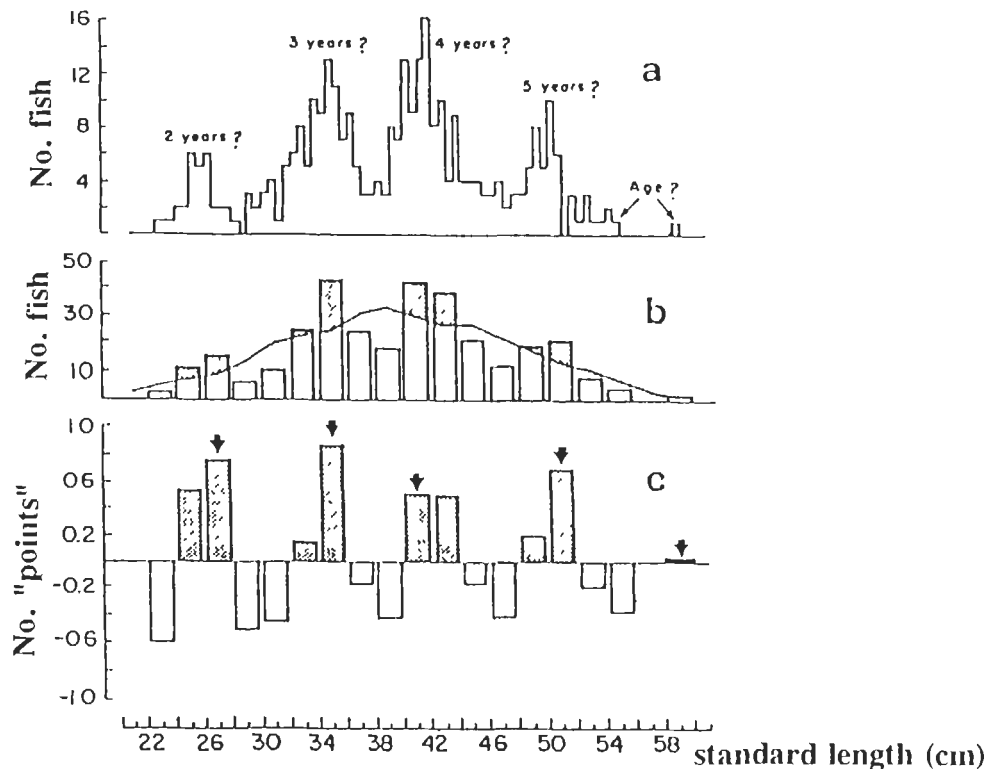


Fig. 3.5.1.1 Example of the ELEFAN I restructuring of a length-frequency sample (from Pauly & David, 1981). Data from Goeden, (1978), on the coral trout (*Plectropomus leopardus*)

The values, for the first length groups 22-24 and 24-26 cm are calculated by adding two zeroes and one zero respectively as indicated by "*". (A similar procedure is applied to the last length groups.) The curve that results from this procedure is used to emphasize peaks (shaded bars above moving average) and intervening troughs. In part "c" the original frequencies of part "b" have been divided by the moving average and 1 has been subtracted. Consider again as an example length group 26-28 cm. Here we get:

$$15/8.8 - 1 = 0.7 \text{ "points"}$$

Actually, some additional minor adjustments have also been made but we shall not go into that. Using the restructuring process the peaks and the troughs became well-structured and easy to identify by the "points" allotted. Note that clear peaks have been allotted a similar number of points irrespective of the number of fish they represent.

Stage 2, the fitting of a growth curve is illustrated in Figs. 3.5.1.2 and 3.5.1.3.

In the present example for coral trout only one sample was used. To do the ELEFAN I type of fitting growth curves we should preferably have a time series of samples. Basically, ELEFAN I is a modal progression analysis. However, if a time series is not available we can circumvent the problem by assuming one, simply by repeating the sample for a suitable range of years, the assumption being that all cohorts follow the same growth curve. Thus, ELEFAN I can be applied to both the single sample case and the time series case. If the constructed time series over the ten years shown in Fig.3.5.1.2 had been a real time series we would have got slightly different frequencies each year. Fig. 3.5.1.3 shows eight repetitions of the restructured sample arranged similarly to Fig. 3.5.1.2. It is difficult to fit a curve to the

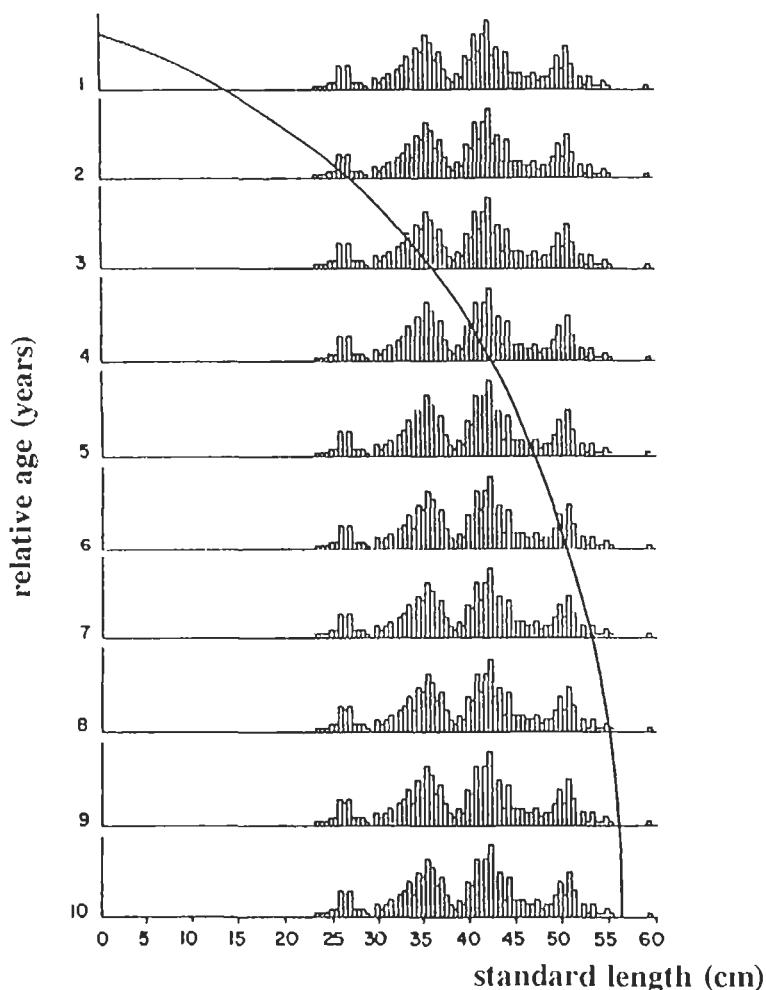


Fig. 3.5.1.2 The length-frequency sample of Fig. 3.5.1.1a repeated over 10 years for simulation of time series of samples (compare Fig. 3.5.1.3)

original frequencies in Fig. 3.5.1.2 and it is not possible to give an objective criterion whether one curve fits better than another if one uses an eye fit only. The restructured samples in Fig. 3.5.1.3 however, are easier to fit because peaks and troughs have been exaggerated.

With the restructured data (the "points" shown in Fig. 3.5.1.1c) it has become possible to define an objective measure for goodness of fit, for which Pauly and David (1981) suggested the ratio "ESP/ASP", where "ESP" stands for "*Explained Sum of Peaks*" and "ASP" for "*Available Sum of Peaks*".

To understand the concept of "ESP" consider Fig. 3.5.1.3. The most convincing fit of a growth curve is one which hits all the peaks indicated by arrows. However, there may not exist such a von Bertalanffy growth curve, and therefore a "score" concept has been introduced to measure how close a curve can come to the best fit. Whenever a curve hits a bar at the axis, either positive or negative it scores "points" (cf. Fig. 3.5.1.1). The total score of a growth curve is the sum of the points scored from each sample as shown in Fig. 3.5.1.3.

"ASP" (available sum of peaks) is the maximum score a curve can reach, i.e. the sum of the positive peaks indicated with arrows. Such an arrow occurs whenever there is a sequence of positive bars. (In this connection a "sequence" may be a single bar.) The ratio ESP/ASP thus becomes a measure for how close a curve is to the best possible fit.

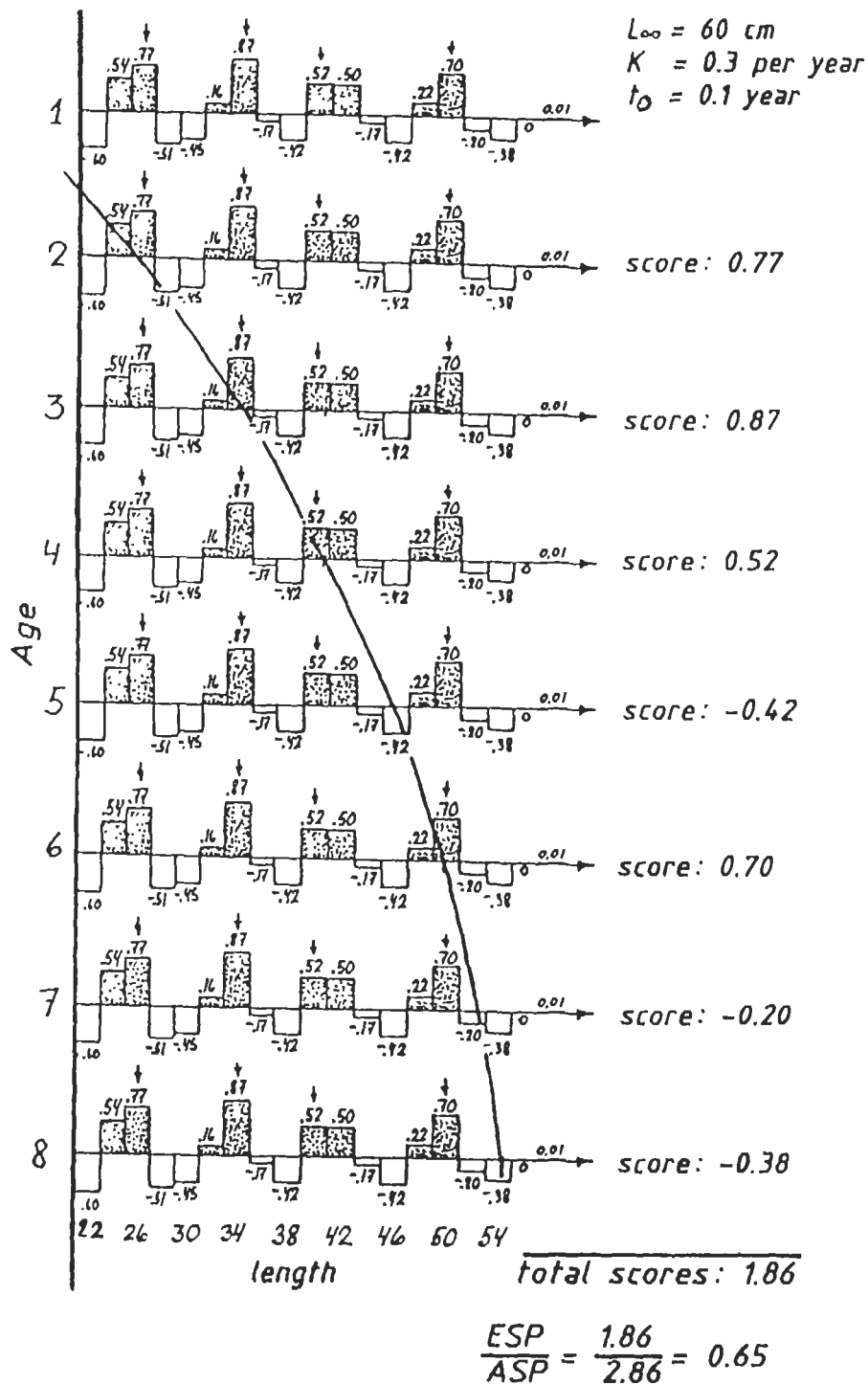


Fig. 3.5.1.3 The restructured length-frequency sample of Fig. 3.5.1.1c repeated over eight years to simulate a time series of samples (compare Fig. 3.5.1.2). A single growth curve determined by the parameters $L_{\infty} = 60 \text{ cm}$, $K = 0.3 \text{ per year}$ is tested for goodness of fit (ESP/ASP)

The computational procedure described so far may be carried out by paper and pencil for a single growth curve within a reasonable time. But after that it is no longer possible (in practice) to follow ELEFAN I by paper and pencil. One of the main features of ELEFAN I is that many (say, thousands) of different growth curves are tested in the way described in Fig. 3.5.1.3. Among the thousands of possible growth curves the one that produces the highest value of ESP/ASP is selected.

(See Exercise(s) in Part 2.)

3.5.2 The seasonalized von Bertalanffy growth equation

Fig. 3.5.2.1 shows an application of ELEFAN I to a penaeid shrimp. This growth curve estimated by ELEFAN I is clearly not a von Bertalanffy growth curve because $\Delta L/\Delta t$ does not decrease linearly with age (cf. Section 3.1). The explanation is that ELEFAN I works with the "seasonalized von Bertalanffy growth equation" (Pitcher and Macdonald, 1973; Cloern and Nichols, 1978 and Pauly and Gaschütz, 1979):

$$L(t) = L_{\infty} * [1 - \exp\{-K*(t-t_0) - (CK/2\pi) * \sin(2\pi*(t-t_s))\}] \quad (3.5.2.1)$$

This is the usual von Bertalanffy equation (Eq. 3.1.0.1) with an extra term:

$$(CK/2\pi) * \sin(2\pi*(t-t_s)) \quad (\text{where } \pi = 3.14159\dots)$$

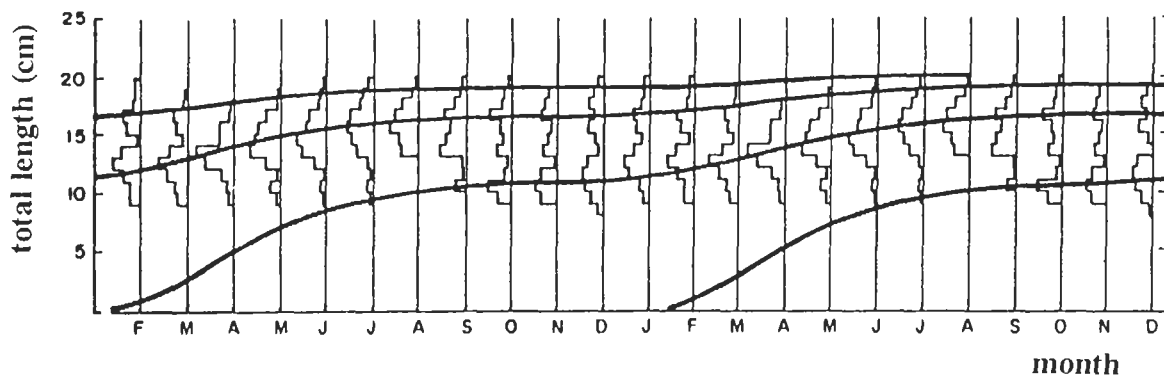


Fig. 3.5.2.1 Example of a seasonally oscillating growth curve estimated by ELEFAN I (from Pauly, 1981). Data from Rodriguez (1977) on female shrimp (*Penaeus kerathurus*) off Cadiz, Spain. Note that data were available for one year, and these have been repeated to simulate two years of sampling. Estimated parameters are: $L_{\infty} = 21.0$ cm (total length), $K = 0.8$ per year, $C = 0.9$, $t_w = 0.8$ (winter point), $ESP/ASP = 0.46$

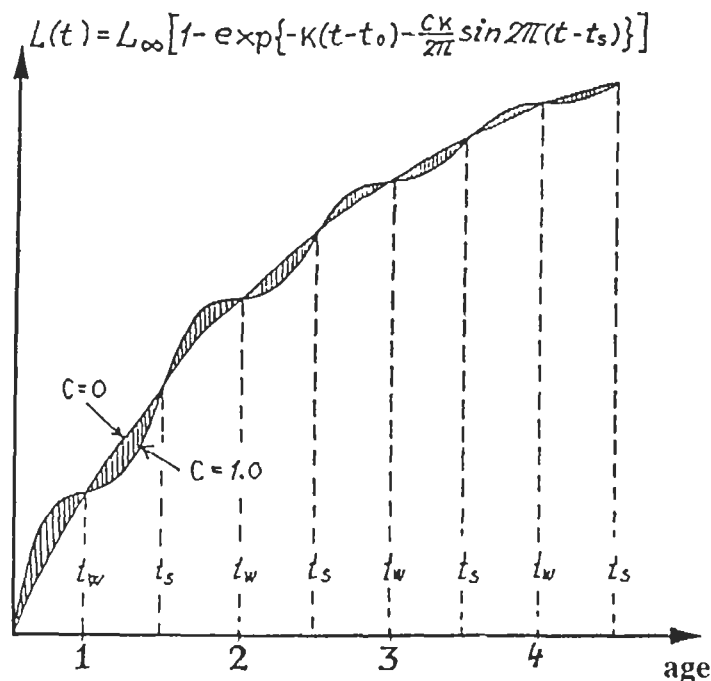


Fig. 3.5.2.2 The seasonalized von Bertalanffy growth equation. Note that for $C = 1$ the growth rate is zero at the winter points

This term produces seasonal oscillations of the growth rate, actually by changing t_0 during the year. The parameter "ts" is called the "*summer point*" and takes values between 0 and 1. At the time of the year when the fraction ts of the year has elapsed the growth rate is highest. At time $t_w = ts+0.5$, the "*winter point*", the growth rate is lowest. The parameter C, the "*amplitude*", also usually takes values between 0 and 1. If $C = 0$ Eq. 3.5.2.1 reduces to the ordinary von Bertalanffy equation, that is $C = 0$ implies that there is no seasonality in the growth rate. The higher the value of C the more pronounced are the seasonal oscillations. If $C = 1$ the growth rate becomes zero at the winter point. Fig. 3.5.2.2 shows a seasonalized growth curve with $C = 1$ together with an ordinary von Bertalanffy curve ($C = 0$). All other seasonalized curves with different C's (but with other parameters kept constant) will be in the shaded area.

3.5.3 Maximum likelihood methods

The calculation of a mean value as described in Section 2.1 and the least squares method described in Section 3.3.4 are applications of the "*maximum likelihood principle*".

The method to be described in this section aims at solving the same problem as the ELEFAN I method and some other problems. The main difference lies in the definition of the goodness of fit. ELEFAN I uses the ratio ESP/ASP (cf. Section 3.5.1) whereas the "*maximum likelihood method*" uses the (weighted) sum of the squares of the deviations between model and observations (or measures with similar properties). In principle this measure of goodness of fit is the same as the one used in linear regression analysis (cf. Eq. 2.4.3 and Fig. 2.4.2).

The full statistical theory behind this method is complicated and so is the computer program. However, a fishery scientist running the program does not need to know all the technical details. If the basic principles behind the method are understood, few difficulties in using the program should be encountered.

The basic idea of ELEFAN I, to follow the progression of modes and test a large number of alternative combinations of growth parameters, is also the basic idea behind the maximum likelihood approach. The measure for goodness of fit used in the maximum likelihood method is closely related to the so-called "*chi-squared criterion*" which is conceptually simple and therefore used in the following explanation of the method.

In Fig. 3.5.3.1 a length-frequency sample is presented that we assume to be composed of two cohorts. When using the maximum likelihood computer program on that sample, we would obtain a result as illustrated in Fig. 3.5.3.2, where the dotted curves represent the two cohorts and the full line the sum of the calculated frequencies of the two cohorts. The dots indicate the original, observed frequencies, and the bars the differences between observed and calculated frequencies.

In addition to the growth parameters the maximum likelihood method also works with the following parameters (in the case of two cohorts):

- N1 = total number of observations in first cohort
- N2 = total number of observations in second cohort
- s1 = standard deviation of first cohort
- s2 = standard deviation of second cohort

The mean lengths, \bar{L}_1 and \bar{L}_2 follow from the growth parameters (cf. Fig. 3.5.3.2, where L_1 and L_2 corresponding to arbitrary ages t_1 and t_2 are shown as an example). From the

parameters the calculated (theoretical) frequency of each cohort, $fc_1(L)$ and $fc_2(L)$ and the total frequency

$$fc_{total}(L) = fc_1(L) + fc_2(L)$$

of each length group can be calculated as explained in Section 2.2.

The measure of goodness of fit, the "chi-squared criterion", is defined as:

$$\chi^2 = \sum \frac{\{f_{obs}(L) - fc_{total}(L)\}^2}{fc_{total}(L)} \quad (3.5.3.1)$$

which is the sum over all $fc_{total}(L)$ values > 0

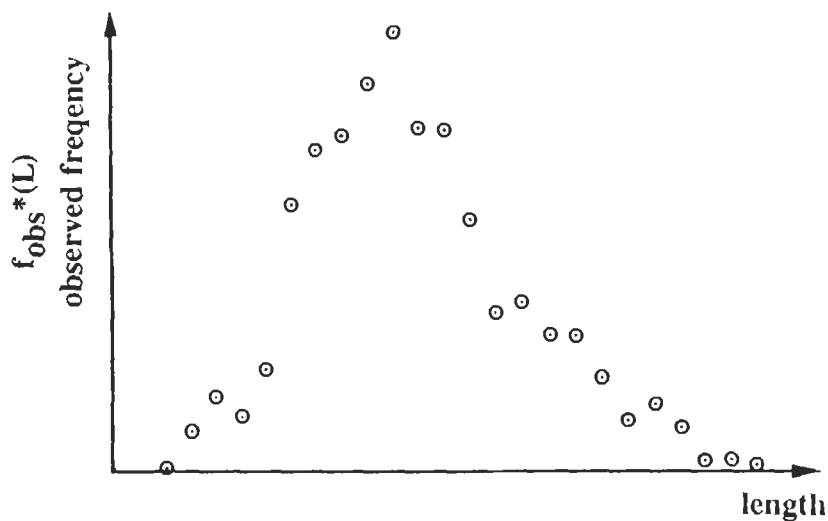


Fig. 3.5.3.1 The basic data from which the resolution into normally distributed components in Fig. 3.5.3.2 is derived

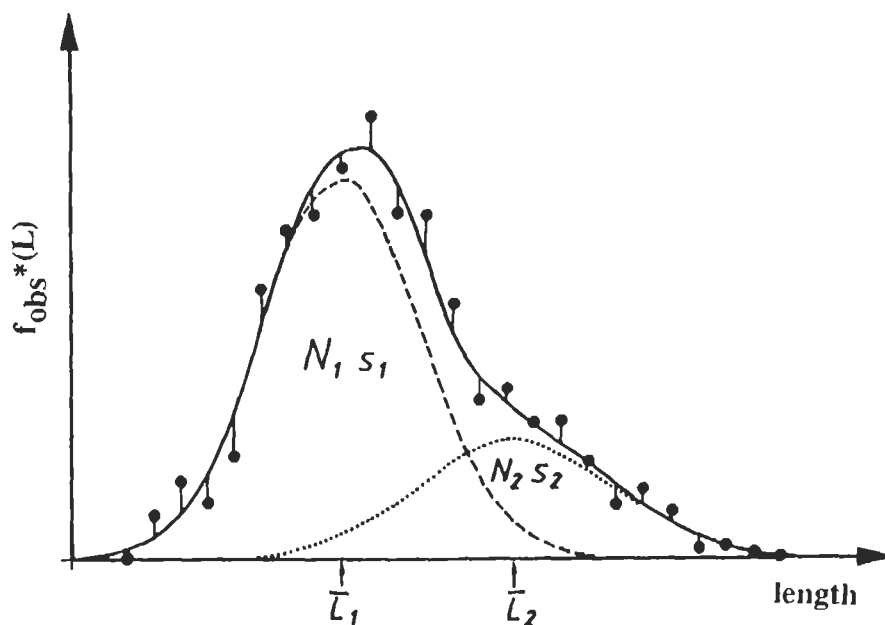


Fig. 3.5.3.2 Illustration of the chi-squared criterion. Input data are from Fig. 3.5.3.1. Also the number of cohorts must be given as input

where $f_{obs}(L)$ stands for the observed frequency in length group L (= interval midpoint). It is used to minimize the differences between observed and calculated frequencies over the entire length range of the sample. The maximum likelihood program determines that set of parameters (L_{∞} , K , t_0 , $N1$, $N2$, $s1$ and $s2$) which minimizes the chi-squared criterion. A comparison with Eq. 2.4.3 (" $f_{c_{total}}$ " and " f_{obs} " correspond to " $a + b*x(i)$ " and " $y(i)$ ", respectively) illustrates the relationship between the chi-squared criterion and linear regression. Fig. 3.2.2.2 shows another example of what the maximum likelihood method would get out of a length-frequency sample if it were given that the number of cohorts was six.

As the chi-squared criterion is a standard measure for goodness of fit when dealing with frequencies the list of references dealing with this concept is nearly endless. A good introduction to the theory (written for biologists) is given in Sokal and Rohlf (1981, Chapter 17).

In addition to the growth parameters the maximum likelihood method also gives numbers and standard deviations. The program requires the same input as the ELEFAN I program, but it also requires the number of cohorts in the sample as input. Often one has to guess that number. However, this extra input appears not to create great practical problems.

The maximum likelihood program works as an "*iterative process*". That is, it must be fed an initial guess on the solution which is then improved in a number of iterative steps. Thus, to start the maximum likelihood estimation procedure we need an approximation to the solution of the exercise. Such an initial solution can be obtained from, for example, the Bhattacharya analysis and the modal progression analysis described in Sections 3.4.1 and 3.4.2. The maximum likelihood method does not make the "paper-and-pencil" methods superfluous. We still need these methods to start the iteration process and, perhaps most important, to evaluate the results. The search for an acceptable set of initial values often is the most time-consuming part of the task.

Fig. 3.5.3.3 illustrates the procedure of the maximum likelihood estimation. Usually, the starting point is called the "*initial guess*" at the solution. However, calling it a "guess" might not be appropriate, as it has to be rather close to the final solution to make the iterative process converge. Therefore, it is important to have a simple and dependable method to get a first "good guess" at the solution. For example the Bhattacharya method and the modal progression analysis could be used.

Another feature of the maximum likelihood method is that it gives estimates of the confidence limits of all the parameters, which the Bhattacharya method and the modal progression analysis are unable to do. The confidence limits from the modal progression analysis given in Table 3.4.2.3 are based on the assumption that the estimates from the Bhattacharya analysis have zero variance. The maximum likelihood method does not require such (highly unrealistic) assumptions.

We conclude this brief discussion of the maximum likelihood method with a few words on its historical development. The first work in the field is nearly as old as Petersen's pioneering work on length-frequencies of fish (cf. Section 3.4.) as Pearson in 1894 presented his work on separation of frequencies into normally distributed components.

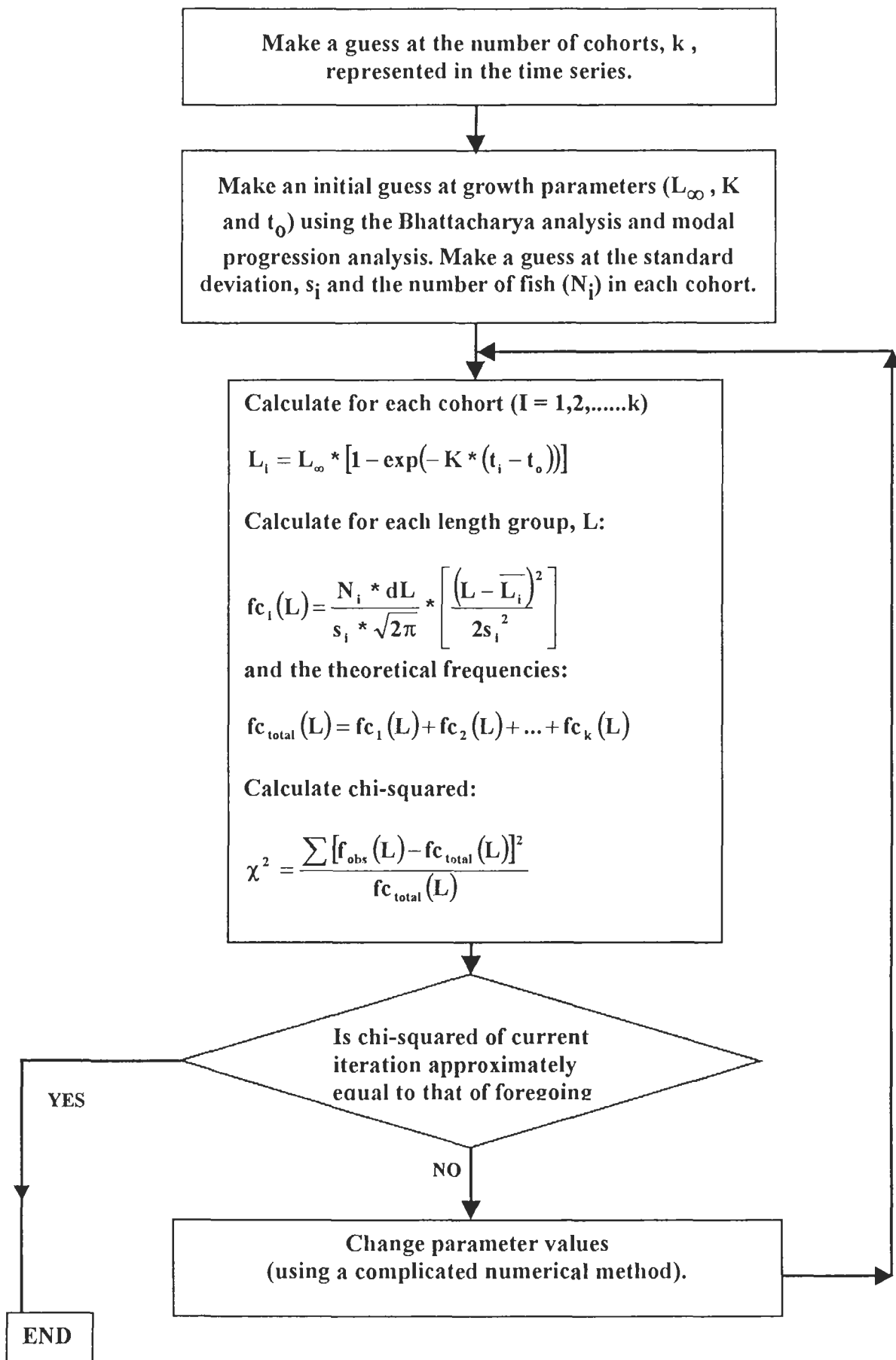


Fig. 3.5.3.3 The iterative process of the maximum likelihood estimation procedure (see also Fig. 3.5.3.2)

Computer programs

One of the first computer programs to separate frequencies into normally distributed components using maximum likelihood techniques is the "NORMSEP" program by Hasselblad and Tomlinson (1971). NORMSEP was based on the work by Hasselblad (1966). Another

important contribution on separation of fish length-frequencies into normally distributed components was given by Macdonald and Pitcher (1979). This work was extended by Schnute and Fournier (1980) to include estimation of growth parameters in the single sample case. This contribution in turn was extended by Sparre (1987a) to deal with the time series case and the seasonalized von Bertalanffy growth curve and a few other things, the theory of which is dealt with in the following section. The NORMSEP program is included in the FiSAT package.

3.5.4 Limitations of length-frequency analysis

As appears from the examples (Sections 3.4 and 3.5) it is often difficult to resolve a mixed distribution. The old fish (the longest fish) especially create problems. Intuitively, one expects the separation into components to be troublesome when mean values of neighbouring components are located close to each other compared to the size of the standard deviations.

Applying more rigorous statistical methods than those presented in this manual, Hasselblad (1966), McNew and Summerfelt (1978) and Clark (1981) have shown that the "*separation index*"

$$I = \frac{\bar{L}(a+1) - \bar{L}(a)}{\{s(a+1) + s(a)\}/2} \quad (3.5.4.1)$$

is a relevant quantity to study when assessing the possibility for a successful separation of two neighbour components. \bar{L} stands for the mean value and s for the standard deviation (see Fig. 3.5.4.1). Without going into details the main findings of the three above-mentioned works can be summarized by the rule of thumb: If the separation index is less than two, $I < 2$, it is virtually impossible to separate the components.

Table 3.5.4.1 Separation indices calculated for the example of Section 3.4.1. The parameters marked by "*" cannot be estimated from length-frequency data alone (cf. Table 3.2.1.1 and Fig. 3.2.2.2)

a	$\bar{L}(a)$	s(a)	$I = 2 * \frac{\bar{L}(a+1) - \bar{L}(a)}{s(a+1) + s(a)}$
1	17.3	1.7	4.82
2	27.9	2.7	
3	35.3	3.4	2.43
4	40.2 *	3.6 *	1.40
5	43.3 *	3.8 *	0.84
6	45.5 *	3.6 *	0.59

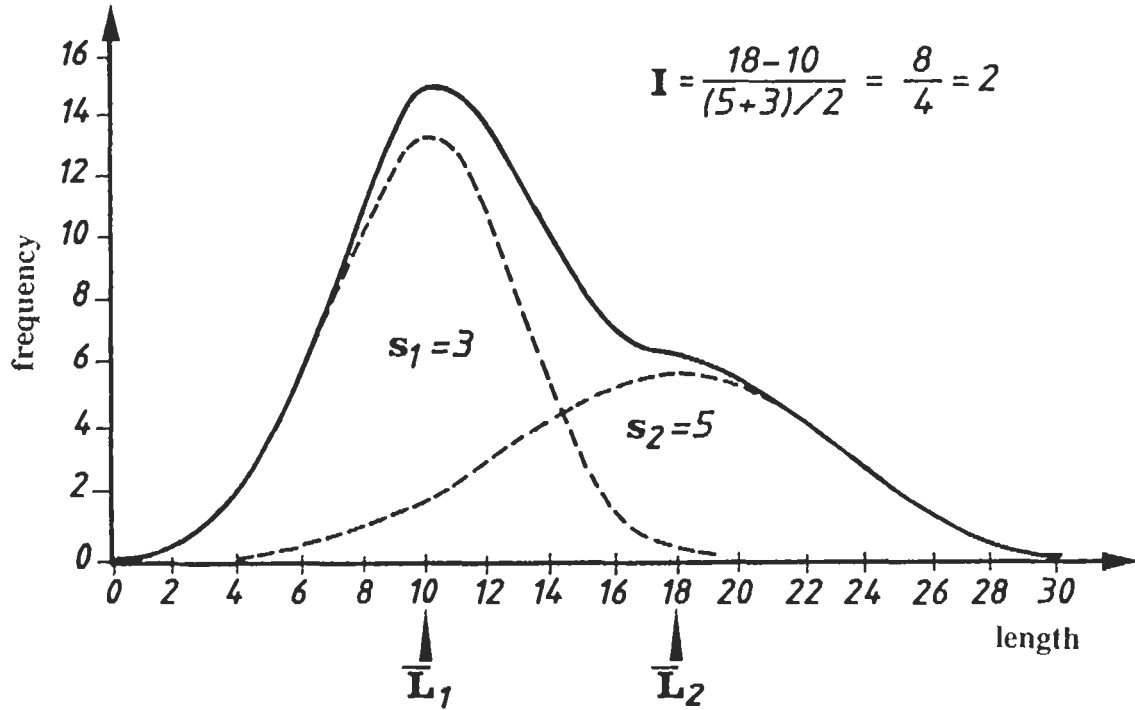


Fig. 3.5.4.1 Example of two normally distributed components with the critical separation index, I-value of 2

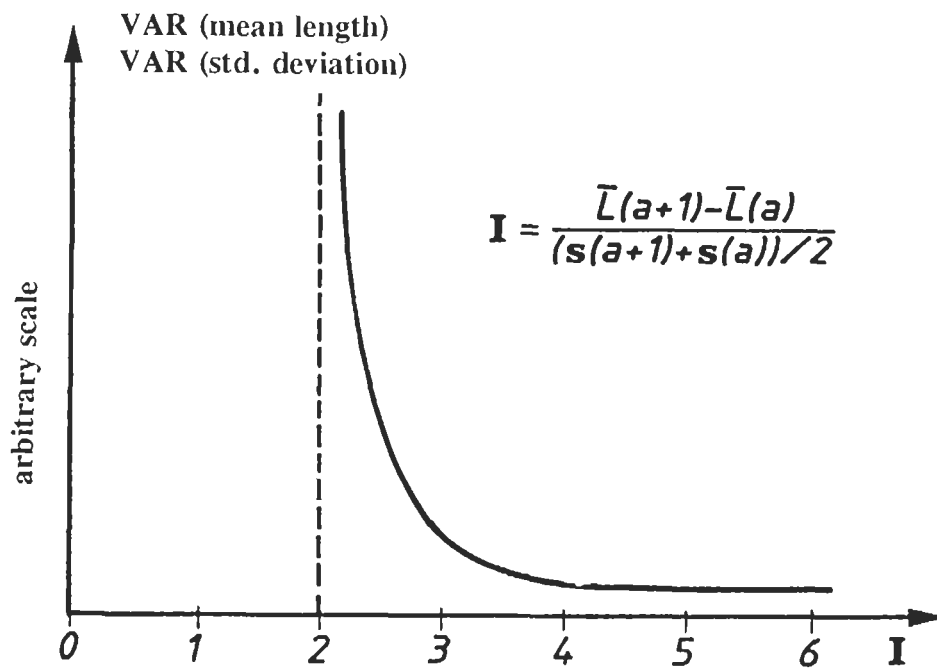


Fig. 3.5.4.2 General description of the functional relationship between separation index, I, and variances of estimates

Fig. 3.5.4.1 shows an example of two normally distributed components with $I = 2$. Fig. 3.5.4.2 shows the typical functional relationship between separation index and variance of the estimates. (For further details see, for example Hasselblad, 1966.)

As an example consider Table 3.2.1.1 (i.e. the hypothetical data used to illustrate the paper and pencil methods). In Table 3.5.4.1 the separation indices have been calculated for the six components. These are known because the data are hypothetical or constructed. Suppose the data had been real data for which we did not know the true parameters. In that case there would be hope for estimation of only three components with separation indices 4.82 and 2.43 respectively. This conclusion holds for all methods, including the most sophisticated computerized ones.

Another way of exploring the limitations of length-frequency analysis is the "*Monte Carlo simulation technique*". By this technique we simulate length-frequency samples using a computer (cf. Section 3.2.1). The technique is called "Monte Carlo" because it includes a component of "*random variability*", the principle of the "roulette", which is added to all the simulated observations. By making assumptions about the parameter values and the magnitude of the random component and by simulating the corresponding length-frequency samples we are in a position to evaluate the various methods. The procedure works as follows:

- Step 1:** Make assumptions on parameter values and the magnitude of the stochastic component.
- Step 2:** Simulate a time series of length-frequencies according to step 1.
- Step 3:** Analyse the simulated data (assuming the parameters to be unknown) using for example Bhattacharya analysis and modal progression analysis.
- Step 4:** Compare the results (if any) of step 3 to the "true" parameters from step 1.

Using this procedure we will be able to give statements like: If a fish stock has length distributions with certain parameters then we are able or not able to estimate the growth parameters with a certain prespecified accuracy.

Also difficulties in obtaining unbiased samples should be mentioned in connection with the limitations of length-frequency analysis. Probably the most important source of bias stems from the migration of fish. Limitations of length-based methods applied to migratory fish stocks are discussed in Chapter 11.