ESTIMATION OF GROWTH PARAMETERS

T. V. Sathianandan

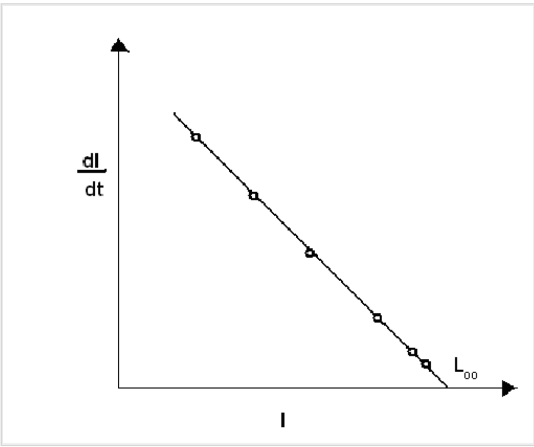
Fishery Resources Assessment Division

ICAR- Central Marine Fisheries Research Institute

15

A growth curve is an empirical model of the evolution of a quantity over time. Some growth curves for certain biological systems display periods of ***exponential growth***. **A Gompertz curve,** named after Benjamin Gompertz, is a sigmoid function. It is a type of mathematical model, where growth is slowest at the start and end of a time period. In biology, a growth model is a depiction of length or weight of animals as a function of age. In the case of fish populations, the study of growth is to determine the body size as a function of its age. The growth model developed by **von Bertalanffy** (1934) has been found to be suitable for the observed growth of most of the fish species. This model expresses length as a function of age of the animal.

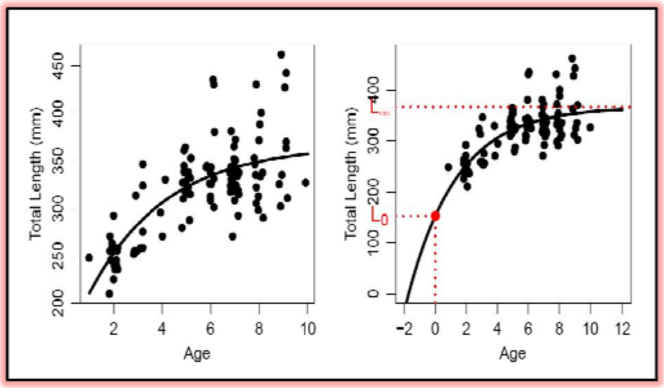
Fish increases in length as they grow older but their growth rate which is the increment in length per unit time decreases as they grow old. When the rate of growth is plotted against the length, in most cases it will look almost like a straight line with descending limb (negative slope). This line will cut the x-axis at a point where the rate of growth is zero. This is the point beyond which the fish will not grow further and the length of the fish at this point is known as the asymptotic length denoted by L_∞ .

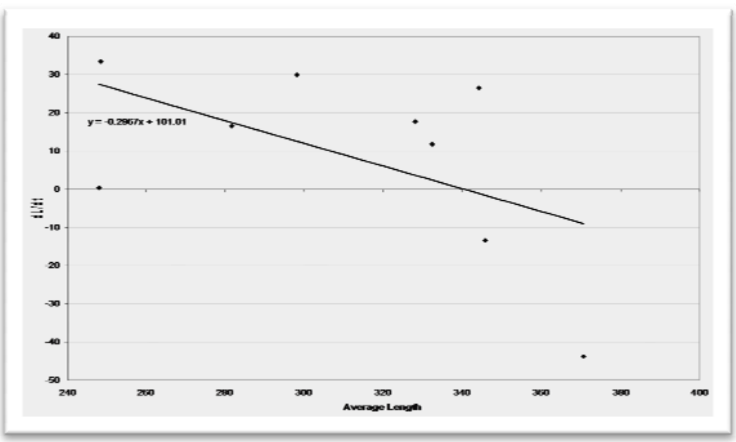


Example: Length-at-age for a portion of a sample of male Atlantic croakers (left) and average length-at-age are given in the following table. The figures show plots of the growth curve and growth rates.

Reprinted from the CMFRI, FRAD. 2014. Training Manual on Fish Stock Assessment and Management, p.150.

Indiv	*Age*	*Length*	*AvgAge*	*AvgLength*
1	1	248	1	248.0
2	2	210	2	248.4
3	2	225	3	281.8
4	2	236	4	298.3
5	2	240	5	328.2
6	2	245	6	345.9
7	2	255	7	332.5
8	2	258	8	344.3
9	2	263	9	370.8
10	2	270	10	327.0
11	2	292		
⋮	⋮	⋮		





To develop the growth model the above phenomenon can be represented by means of a differential equation

$$\frac{dl}{dt} = K(L_\infty - l)$$

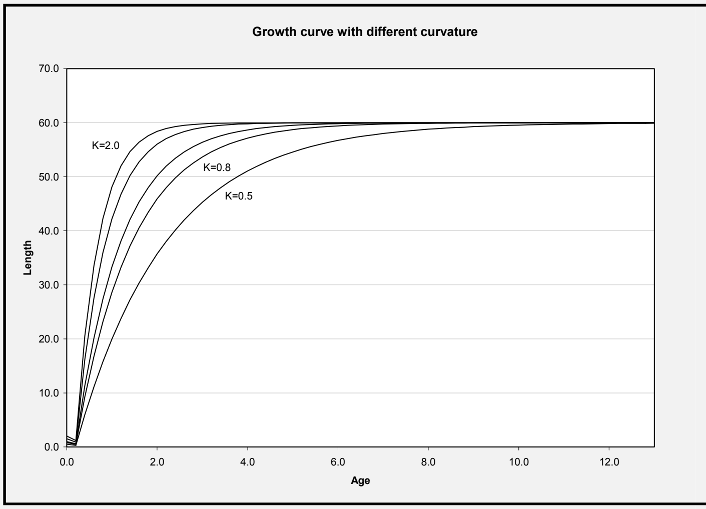
This can be rewritten as

$$\frac{dl}{L_\infty - l} = K\,dt$$

The required growth model is then obtained by integrating the above differential equation to yield,

$$-\log(L_\infty - l) = K\,t + C$$

where C is a constant to be determined.



Expressing this equation for the length l we get,

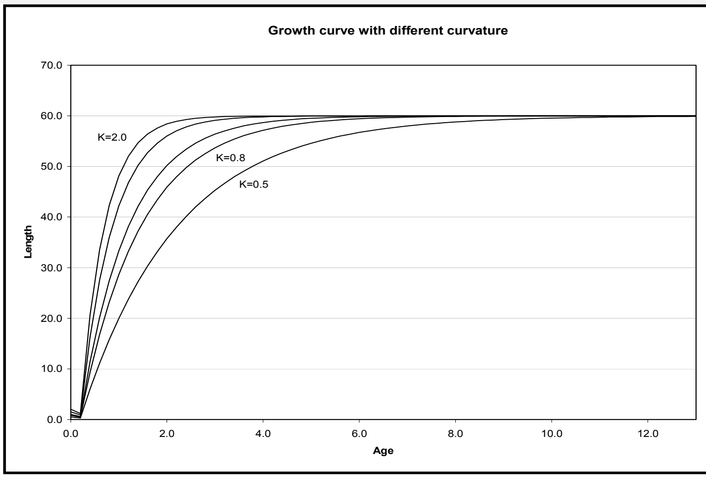
$$l = L_\infty - Ce^{-K\,t}$$

When $t = 0$ the length l also will be zero so that we get,

$$0 = L_\infty - C.$$

Hence $C = L_\infty$ and we get the equation as

$$l = L_\infty(1 - e^{-K\,t})$$



But usually the length will be zero at a different point $t = t_0$ so that we get the solution for the constant as

$$C = L_\infty\, e^{K\,t_0}$$

Hence the general growth equation is obtained by substituting the above value for C as $l = L_\infty(1 - e^{-K(t - t_0)})$

Parameters of the model are K, L_∞ and t_0. Here K is termed as the curvature, L_∞ is known as the asymptotic length and t_0 is the age at birth.

The data commonly used for fish stock assessment is the length frequency data collected periodically by sampling from commercial catches. The data so obtained will consist of animals of different age group. The animals born on same day (single spawning) is termed as a cohort and the animals of the same age will not have same length rather it will vary with a mean and variance. If we make a histogram of their length most of the animals will fall at the middle and it will have the well known bell shape. The sample collected at a time will be a mixture of such bell shaped distributions corresponding to different age groups. If we are able to trace the length distributions of each cohort separately from its initial age up to its life span then we would be able to work out its growth and growth model parameters. As the sample collected by us from commercial catch will be a mixture of cohorts of different age groups the problem reduces to resolution of individual components (known as normal distributions or Gaussian components) from the mixture.

Resolution of Gaussian Components from Polymodal Distributions

The frequency distribution of length obtained from a sample of fish is usually skew and polymodal. The modes corresponding to individual age groups are very useful in separating the different Gaussian components of which it is assumed to be composed off. Here the problem is to resolve a distribution into Gaussian components. Different procedures are available for resolution of a mixture into Gaussian components. These are probability paper method, parabola method and Bhattacharya's method. Among this the last method is most popular.

Probability Paper Method: Decomposition of polymodal frequency distributions using probability paper method was introduced by Harding (1949) and later modified by Cassie in 1950. This involves dissection of the distribution at points of inflexion of the probit plot, followed by correction for over lap of components. In this method, the cumulative percentages of the frequency distribution are first plotted against the mid points of the classes on a probability graph paper and the point of inflexion are marked. Cumulative percentages of these points are the keys for separation of the components and each segment between them are due to separate distributions. Each of these components is then extracted by adjusting the original cumulative percentages within in segments so that the total is 100. These adjusted values if plotted on the same probability paper will be linear. The means of each separated component are estimated from the actual frequencies falling in the corresponding region.

Parabola Method: If the frequency distribution of random variable distributed as normal has *y* as the frequency for a class with mid value *x* then we can express y as

$$y = N \int_{x-c/2}^{x+c/2} f(x)\,dx$$

where *f(x)* is the probability distribution of a normal random variable with mean μ, standard deviation σ, *c* the class interval and *N* the total frequency. An approximation for the relation is

$$y \approx \frac{(N c e^{-(x-\mu)^2})/2\sigma^2}{\sqrt{2\pi}\sigma}$$

$$\ln(y) \approx \ln(\frac{Nc}{\sqrt{2\pi}\sigma}) - \frac{(x-\mu)^2}{2\sigma^2}$$

The above equation is of the form which is a quadratic equation representing a parabola. The axis of symmetry of the above parabola will be at $x = \mu$. Hence, if we plot the natural logarithm of the class frequencies against the mid values of the classes we can represent the different peaks with different parabolas each corresponding to a normal distribution whose mean is the point where the axis of symmetry intersects the *x*-axis.

Bhattacharya's Method: If $y(x)$ denote the observed frequency of the class with *x* as its mid value and *h* the class width, then

$$y = \int_{x-h/2}^{x+h/2} \sum_{i=1}^{k} N_i f(x, \mu_i, \sigma_i)\, dx$$

$$\approx \int_{x-h/2}^{x+h/2} N_r f(x, \mu_r, \sigma_r)\, dx$$

$$\approx \int_{x-h/2}^{x+h/2} N_r \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu_r)^2}\, dx$$

$$\ln(y) = \ln(\frac{hN_r}{\sigma_r\sqrt{2\pi}}) - \frac{h^2}{24\sigma^2} - \frac{\sigma_r^2 - (h^2/12)}{2\sigma_r^2} t^2$$

by ignoring terms with higher orders in *h* and

$$t = \frac{(x-\mu_r)}{\sigma}$$

$$\Delta t^2 = 2h(x - \mu_r + \frac{h}{2}\sigma_r^2 \text{ and}$$

$$\Delta \ln(y) = -h(\sigma_r^2 - h^2/12) - (x - \mu_r + h/2)/\sigma_r^4$$

That is, the graph of $\Delta \ln(y)$ against the mid value of the class will be linear. If denote the x intercept and the angle the line makes with the –ve direction of the x-axis then the mean and standard deviationof the Gaussian component corresponding to this region are estimated as a plot of $\Delta \log y(x) = \log(y(x+h)) - \log(y(x))$ against *x* is to be made first. Then the number of regions where the graph look like straight lines with–ve slope, indicate the number of components (under certain conditions). By connecting the points in the regions fit straight lines for these regions. If θ_r is the angle it makes with the *x* axis and λ_r is the *x* intercept for the r^{th} region for $r = 1, \cdots, k$ then the mean and variance of the r^{th} component is estimated as

$$\mu_r = \hat{\lambda}_r + h/2$$

$$\hat{\sigma}_r^2 = (dh\ Cot\hat{\theta}_r / b) - h^2/12$$

where *b* and *d* denote the relative scales for *x* and $\angle \log y(x)$ respectively. The proportions of the mixture can be estimated as

$$\hat{p}_i = \hat{N}_i \Big/ \sum_{i=1}^{k} \hat{N}_i$$

where $\hat{N}_i$ is the total frequency of the i^{th} class and it is estimated by

$$\hat{N}_r = \sum y(x) \Big/ \sum \hat{p}_r$$

Here, the summation being restricted to the region under consideration and

$$\hat{p}_{ri} = P\left(\frac{x + h/2 - \hat{\mu}_i}{\hat{\sigma}_i}\right) - P\left(\frac{x - h/2 - \hat{\mu}_i}{\hat{\sigma}_i}\right)$$

where *P(x)* is the distribution function of standard normal variate.

Estimation of Growth Parameters

Once we have data on age and corresponding length obtained from the above procedure we may use any one of the following methods as per the situation to estimate the growth parameters.

Gulland and Holt Plot: For small values of Δt (need not be kept constant), the required expression is

$$\frac{\Delta L}{\Delta t} = K\,L_\infty - K\,\overline{L}_t \text{ where } \Delta L = L_{t+\Delta t} - L_t \text{ and } \overline{L}_t = \frac{(L_{t+\Delta t} + L_t)}{2}$$

By regressing $\frac{\Delta L}{\Delta t}$ on $\overline{L}_t$ (of the type $y = a + b\,x$) we can get estimates of the growth parameters as

$$\hat{K} = -\hat{b} \text{ and } \hat{L}_\infty = -\frac{\hat{a}}{\hat{b}}$$

Example: The first two columns of the following table pertain to the age and corresponding average length of animals of a cohort. The growth parameters can be estimated by calculations in the remaining columns and a followed regression. The steps followed are

1. Generate column dL as the increment in length (difference of consecutive values of L(t)
2. Generate column dt as the increment in age (difference of consecutive values of Age(t)
3. Compute values in column dL/dt as the ratio of values in dL and dt)
4. Compute the mean length Lbar(t) as the average of consecutive values of L(t)

Age (t)	L(t)	dL	dt	dL/dt	Lbar(t)
1	25.7	10.3	1	10.3	30.85
2	36.0	6.9	1	6.9	39.45
3	42.9	4.6	1	4.6	45.20
4	47.5	3.2	1	3.2	49.10
5	50.7	2.1	1	2.1	51.75
6	52.8	1.4	1	1.4	53.50
7	54.2				

Now regress the values in column dL/dt with the values in Lbar(t). That is carryout regression analysis with values in column dL/dt as Y values and values in Lbar(t) as X values and obtain the regression coefficients *a* and *b*.

Regression Statistics	
Multiple R	0.999922
R Square	0.999844
Adjusted R Square	0.999804
Standard Error	0.046844
Observations	6

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	22.36353	0.11182	199.9952	3.75E-09
Lbar(t)	-0.39163	0.00245	-159.872	9.18E-09

The estimates of coefficients in the regression model obtained through the regression analysis are a = 22.36353 and b = -0.39163 and the estimates of growth parameters are

$$\hat{K} = -\hat{b} = 0.39163 \text{ and } \hat{L}_\infty = -\frac{\hat{a}}{\hat{b}} = -\frac{22.36356}{-0.39163} = 57.1$$

Ford-Walford Plot: The growth equation can be brought into the form

$$L_{t+\Delta t} = a + b\,L_t \text{ where } a = L_\infty(1-b) \text{ and } b = e^{-K\,\Delta t}$$

When Δt is constant we can get estimates of a and b by regressing $L_{t+\Delta t}$ on L_t and the estimates of growth parameters can be obtained as

$$\hat{K} = -\frac{\ln(\hat{b})}{\Delta t} \text{ and } \hat{L}_\infty = \frac{\hat{a}}{(1-\hat{b})}$$

Example: For the same set of data the column L(t+1) is made with the next value of L(t). As per the Ford-Walford plot we regress the values in L(t+1) with values in L(t) and find the regression coefficients a and b.

Age (t)	L(t)	L(t+1)
1	25.7	36.0
2	36.0	42.9
3	42.9	47.5
4	47.5	50.7
5	50.7	52.8
6	52.8	54.2
7	54.2	

Regression Statistics	
Multiple R	0.999987
R Square	0.999974
Adjusted R Square	0.999968
Standard Error	0.039173
Observations	6

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	18.7018	0.074708	250.3308	1.53E-09
L(t)	0.672493	0.001713	392.5672	2.53E-10

The estimates of coefficients in the regression model obtained through the regression analysis are a = 18.7018 and b = 0.672493. Thus the estimates of growth parameters are

$$\hat{K} = -\frac{\ln(\hat{b})}{\Delta t} = -\frac{\ln(0.672493)}{1} = 0.3968 \text{ and } \hat{L}_\infty = \frac{\hat{a}}{(1-\hat{b})} = \frac{18.7018}{(1-0.672493)} = 57.1$$

Method of Chapman and Gulland: When Δt is constant, using the growth equation we can make the relation

$$L_{t+\Delta t} - L_t = c\, L_\infty - c\, L_t \text{ where } c = 1 - e^{-K\,\Delta t}$$

Through a regression of $(L_{t+\Delta t} - L_t)$ on L_t we can arrive at a regression relation of the form $y = a + b\,x$ and using the estimates of coefficients of this regression equation we can estimate the growth parameters as

$$\hat{L}_\infty = -\frac{\hat{a}}{\hat{b}} \text{ and } \hat{K} = -\frac{\ln(1+b)}{\Delta t}$$

Example: For the given data first we generate a column with values L(t+1)-L(t) and regress these values on L(t) to obtain the constants a and b in the linear regression equation.

Regression Statistics	
Multiple R	0.999945
R Square	0.999891
Adjusted R Square	0.999863
Standard Error	0.039173
Observations	6

	Coefficients	Standard Error	t Stat	P-value
Intercept	18.7018	0.074708	250.33083	1.528E-09
L(t)	-0.32751	0.001713	-191.182	4.49E-09

The estimates of a and b from the regression analysis are a = 18.7018 and b = -0.32751. The corresponding estimates of growth parameters are

$$\hat{L}_\infty = -\frac{\hat{a}}{\hat{b}} = -\frac{18.7018}{-0.32751} = 57.1 \text{ and } \hat{K} = -\frac{\ln(1+b)}{\Delta t} = \frac{-\ln(1-0.32751)}{1} = 0.3968$$

ELEFAN – Electronic Length Frequency Analysis

The first component ELEFAN-I in the system of ELEFAN is the program for estimation of growth parameters from length frequency data. It was first developed in 1978 and it consisted of (i) component for separation of samples into normally distributed components (ii) estimation of growth parameters by generating the growth curve and minimizing the sum of squared deviations from the means of the component distributions. Later versions incorporated an algorithm which by passes the sample separation step and fits the growth curve to peaks defined independently of any assumed underlying distribution.

- Data pre-processing: ELEFAN-I uses a simple high-pass filter to identify peaks and troughs in length frequency data. The high pass filter used is a running average over 5 classes which leads to the definition of peaks as those parts of the length frequency distribution that are above the corresponding moving average and those below the corresponding running average are the thoughs separating peaks.
- Steps involved in fitting of the growth curve in ELEFAN-I are
    i. Calculate the maximum sum of points available in a set of length frequency samples. These are points which can be accumulated by one singe growth curve. It is termed as available sum of peaks (ASP).
    ii. Trace through the set of length frequency tables sequentially arranged in time for any arbitrary input of growth parameters L_∞ and K. A series of growth curves starting from the base of each of the peaks are then projected forward

and backward in time to meet all other samples or the same sample again and again.

iii. Accumulate points obtained by each growth curve when passing through the troughs separating peaks.

iv. Select the curve which pass through most peaks and avoid most troughs and accumulate the largest number of points called Explained Sum of Peaks (ESP).

v. Decrement or increment the values of L_∞ and K until the ratio ESP/ASP reaches a maximum.

The growth model used in ELEFAN-I is the seasonally oscillating version of the generalized von Bertalanffy Growth Function (VBGF) of the form

$$L_t = L_\infty[1-\exp(-KD(t-t_0)+\frac{CKD}{2\pi}Sin(2\pi(t-t_s))]^{\frac{1}{D}}$$

where

L_t is the predicted length at age t.

L_∞ is the asymptotic length

K is the growth constant – stress factor by Pauly 1981.

D is another growth constant – termed as surface factor by Pauly 1981

C is a factor that express the amplitude of the growth oscillations.

t_0 is the age at which the fish would have had zero length

t_s sets the beginning of the sinusoidal growth oscillation with respect to t = 0

In ELEFAN-I the model is used with two of the original parameters replaced (i) t_s with winter point WP and (ii) t_0 is described as a factor used to adjust a growth curve to an absolute age scale. Here a parameter "T0" is internally used to fulfil the role of t_0. Winter point WP designates the period of the year, expressed as a function of a year when growth is slowest. In northern hemisphere WP is often found to be near 0.2 (February) while for the southern hemisphere WP is often a value close to zero. The relation between WP and t_s is

$$t_s + 0.5 = WP$$

When D = 1 and C = 0 the model will take the form of the normal VBGF used for fisheries research. When 0 < C < 1 growth oscillates seasonally and when C > 1 growth oscillates strongly.