

Part 2
**Introductory guides to the FMSP
stock assessment software**

6. LFDA software – Length Frequency Data Analysis

G.P. Kirkwood and D.D. Hoggarth

The LFDA (Length Frequency Data Analysis) package was originally produced by FMSP project R4517 and then extended to a Windows-based environment in project R5050CB. The software allows users to estimate non-seasonal and seasonal growth curves from length frequency data using three alternative fitting methods. Using these estimated growth curves, further analysis allows estimation of total mortality rates from a length converted catch curve and two other methods, and estimation of age frequency distributions based on “age slicing”.

6.1 FITTING VON BERTALANFFY GROWTH CURVES

Methods for analysis of length frequency data have tended to fall into two groups. The first group consists of methods that directly estimate growth parameters from the length frequencies, the most well-known of these being the ELEFAN method. The second group consists of what may be called “modal analysis” methods, i.e. methods that attempt to dissect the length frequency distributions into age frequency distributions. These can then be subjected to subsequent analysis using the very wide range of fishery assessment techniques that assume the fish can be aged accurately. The LFDA package includes three methods of estimating von Bertalanffy growth function (VBGF) parameters from the first group.

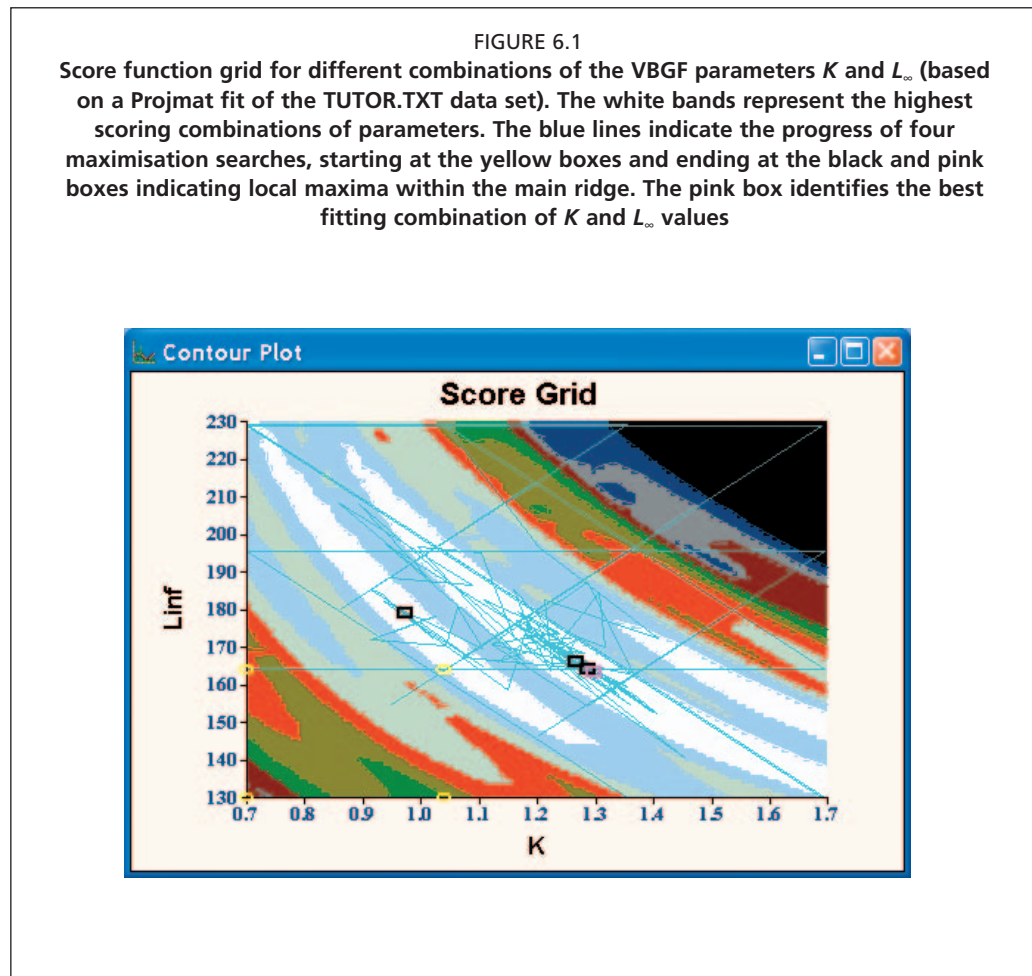
The available methods are Shepherd’s Length Composition Analysis (SLCA, Shepherd, 1987), the projection matrix method (PROJMAT, Rosenberg, Beddington and Basson, 1986) and the ELEFAN method (Pauly, 1987). Each of the three methods was originally developed to estimate the parameters of a non-seasonal von Bertalanffy growth curve; however they are all in principle also suitable for estimating the parameters of seasonal growth curves. In practice, it has been found that the SLCA method does not perform well when estimating seasonal growth parameters, and so LFDA allows only the PROJMAT and ELEFAN methods to be used for fitting seasonal growth curves.

The concept behind fitting VBGF models in LFDA using length frequency data is very simple. Given a single length frequency distribution or a set of length frequency distributions, the set of von Bertalanffy parameters is sought that leads to the best description of the distributions. This is done in a slightly different way for each method, and differently for fitting seasonal and non-seasonal curves, as explained in the Technical Appendix help file. In each case, the overall principle is that a “score function” measures the goodness of fit of the length frequency distributions for each combination of von Bertalanffy parameters. The higher the value of the score function – i.e. the better the goodness of fit – the more consistent that set of von Bertalanffy parameters is with the data. The final estimates of the parameters correspond to those that lead to the highest value of the score function.

The non-seasonal von Bertalanffy curve has three parameters: L_{∞} , the average maximum length, K , a measure of the growth rate (the rate at which L_{∞} is approached); and t_0 , the time (age) at which length is zero. As illustrated in Figure 6.1, strong correlation is usually found between the parameters L_{∞} and K , with one or more ridges of pairs of L_{∞} and K values giving almost equally good fits to the data. This problem

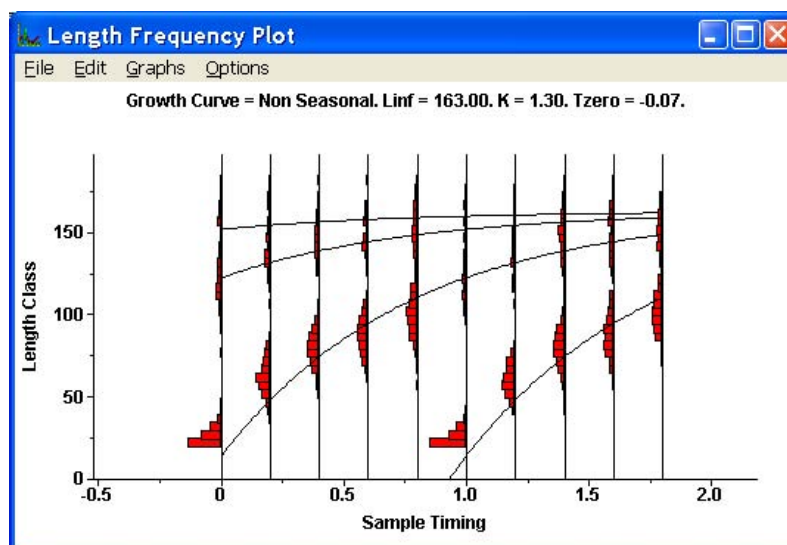
is most extreme when only a small size range of animals are included in the length frequency data. Two of the VBGF parameters, L_{∞} and t_0 , relate to the extreme values of the growth curve. With length frequency data for only a restricted range of lengths, very little information is available about these “anchor points” for the growth curve. This means that estimates of t_0 and L_{∞} represent considerable extrapolations beyond the range of the data and will be associated with high uncertainty. The parameter estimates obtained from length frequency data are thus those that provide the best fit to the data over the range of lengths available.

The LFDA help file describes a procedure for searching manually first of all for good combinations of K and L_{∞} (the highest scoring ridges), and then using a maximisation algorithm to find the best fit within that range (see Figure 6.1). Users are advised to run the maximisation process from a range of different starting points within the selected region to ensure that consistent values are obtained. It is also good practice to always plot the fitted VBGF curves against the length frequency data to confirm by eye that they give a reasonable tracking of the main modes in the data. As noted by Gulland and Rosenberg (1992), if clear modes are not visible in the data, length-based fitting methods should not be used (see Section 3.1.5).



A fitted non-seasonal growth curve is illustrated in Figure 6.2. In this case the data from the tutorial were actually generated with a seasonal simulation model. It can be seen that the curve passes through some of the modes nicely, but misses others quite badly. Where this is apparent one may consider fitting non-seasonal growth curves as guided below.

FIGURE 6.2
A non-seasonal VBGF growth curve fitted to the TUTOR.TXT data set



6.1.1 Fitting seasonal growth curves

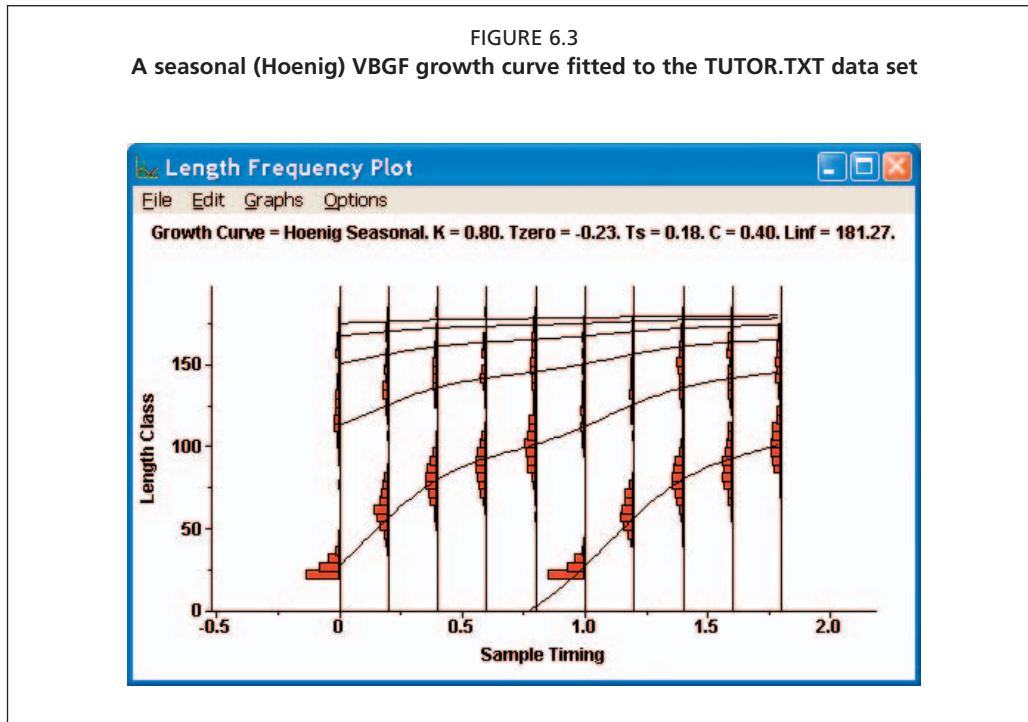
LFDA allows the fitting of seasonal VBGF curves based on the alternative formulations of Hoenig and Hanumara (1982) or Pauly *et al.* (1992) (see LFDA help files). Such seasonal growth curves have five parameters instead of three (the usual K , L_{∞} and t_0 , plus two further parameters that fix the position and amplitude of the seasonality). Fitting curves with five parameters is much harder than with only three and should not even be attempted unless it is fairly clear from the data that there is evidence of seasonal variation in growth rates (e.g. if there are times of year when a non-seasonal curve consistently underestimates or overestimates the observed modal lengths). Seasonal growth curves should not be fitted in other circumstances because they will nearly always show some degree of seasonal growth even if it is not really there at all. An apparently better fit can always be achieved with five parameters than with three.

In a normal statistical context, five parameters is not very many parameters to estimate. Those familiar with non-linear estimation might expect that one should simply carry out a standard numerical maximisation of the score function with respect to all five parameters simultaneously (actually this only would involve four, since the estimate of t_0 can be determined after the other four have been estimated). Experience has shown, however, that only with almost perfect simulated data is this consistently successful, and the option to estimate the parameters this way is not included in the LFDA package (see help files for further details). Rather, the seasonal growth parameters are estimated two-by-two by LFDA, using an iterative process, as described below.

In the case of the Hoenig and Choudary Hanumara curve, the best fitting pairs of values of L_{∞} and K are first sought for fixed values of C and t_s , and then the best estimates of C and t_s are sought for the fixed values of L_{∞} and K . This automated process is repeated over and over until convergence is reached. While this iterative process does normally converge, it is by no means certain that it will converge on the global maximum. In particular, because the process starts by fitting a non-seasonal growth curve, if the true growth curve is highly seasonal inappropriate estimates can arise. Users must exercise some independent judgement based on how well the estimated growth curve really fits the data, and some exploration of different ranges

of parameters is highly recommended. A seasonal curve fitted to the tutorial data set is shown in Figure 6.3.

A second less automated approach may also be taken to fitting seasonal growth curves, taking advantage of any supplementary information about the parameters that may exist. More reliable estimates may thus be found if it is possible to fix the values of L_{∞} or the winter point t_s . If strongly seasonal growth is suspected, it can also be useful to consider only large values of C . Manual grid searches can then be carried out with some of the seasonal growth parameters fixed (see LFDA Tutorial and Reference Guide help files).



6.1.2 Uncertainty in the LFDA growth parameter estimates

Confidence intervals for the growth parameter estimates could in principle be calculated using “bootstrap” resampling methods. However, these have not been implemented in the LFDA package, partly due to the difficulty of automatically finding the global maxima of the score function surface, but also because the resampling process itself is extremely complicated statistically and requires knowledge rarely available about the sampling process underlying the collection of the length frequency samples.

In the absence of estimates of confidence intervals, the best approach may be to identify sets of values of the growth parameters that lead to values of the score function that are close to the maximum, and treat these as informal surrogates for possible confidence regions. For example, in most cases, sets of parameters that lead to values of the score function that differ by only a few percent are probably all equally likely. Sets of growth parameters that lead to fits of the length frequency data that are virtually indistinguishable by eye over the length ranges available must also be considered as equally possible. Estimates of growth parameters fitted using length frequency data are always likely to be fairly uncertain (see Section 3.1.5).

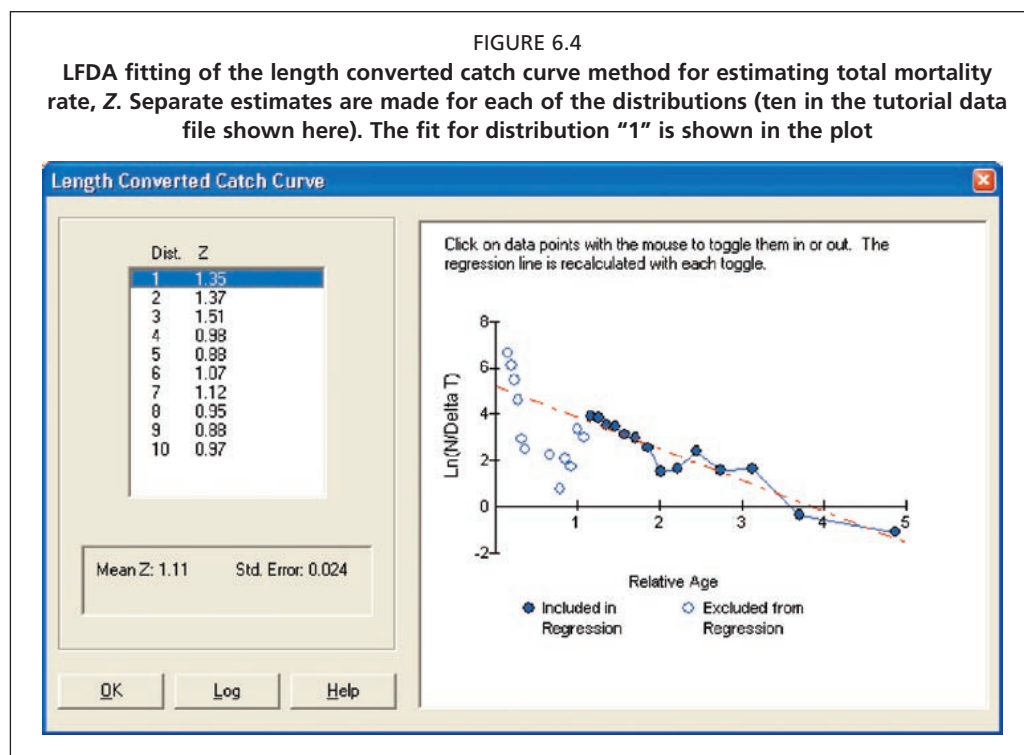
6.2 ESTIMATING TOTAL MORTALITY RATES (Z)

In addition to methods for estimating growth parameters, the LFDA package also includes three methods for estimating the total mortality rate Z (or Z/K), using the estimates of the von Bertalanffy parameters. The routines available in LFDA are the

Beverton-Holt method (Beverton and Holt, 1956), the Powell-Wetherall method (Powell, 1979; Wetherall, Polovina and Ralston, 1987) and a method based on a length converted catch curve. Full descriptions of these methods are given in the Technical Appendix help files. All of the three methods assume that the overall population is in a “steady state”, with constant mortality and recruitment over the ages represented by the lengths in the samples.

All three of the mortality estimators available in LFDA are based on non-seasonal von Bertalanffy growth curves, so cannot be used to estimate mortality for a stock displaying strongly seasonal growth. This is not normally a problem since the growth of most fish stocks can be reasonably described by the non-seasonal von Bertalanffy model. The case of strongly seasonal growth is a difficult one, for it is unlikely that such a stock would have non-seasonal mortality anyway. For this reason such stocks should be treated with great care.

The “catch curve” method produces a separate estimate of Z for each distribution in a length frequency data file by fitting regression lines through the right-hand side of a length-converted von Bertalanffy catch curve (see help files). The user is required to input values of K and L_{∞} (e.g. as estimated above). LFDA then presents estimate of Z for each of the distributions in the dataset, along with the mean and standard error of the estimates (see Figure 6.4). The user is also required to toggle off any points lying on the ascending arm of the catch curve to give a valid fit (see Technical Appendix help file). This gives a degree of subjectivity to the method that is common to length based approaches. Since the estimates of Z may vary over the year as the cohorts grow through the sample, it can be important to have length frequencies from all seasons.



The method described by [Beverton and Holt \(1956\)](#) for estimating Z from length frequency samples is perhaps the most well-known of the three LFDA methods. It relies on a simple algebraic relationship between the mean length in each sample, the length at first full exploitation, the von Bertalanffy growth parameters and the total mortality rate Z . Inputs are this time required for K , L_{∞} and L_c where L_c is defined here as the first length class which is fully exploited (not the same as the length at 50 percent

selectivity). The Beverton and Holt estimator can be quite reliable if the assumptions behind the method are met, and if L_c is well-estimated.

The Powell-Wetherall (1979) method assumes that the shape of the right hand tail of a length frequency distribution will be determined by the ratio between the total mortality rate Z and the growth rate K . Unlike the other methods, it does not directly calculate an estimate of Z ; rather it gives a series of estimates of L_∞ and the ratio K/Z . While it thus provides yet another means of estimating L_∞ , it is a little more complicated to get estimates of Z with this method (see help file). Users should take particular care if the estimate of L_∞ implied by the Powell-Wetherall method differs substantially from those obtained by the SLCA, PROJMAT or Elefan methods. An estimate of Z should only be made from the predicted K/Z by dividing by K from one of these three methods, if the estimates of L_∞ are similar for both of the methods.

As noted above, each of the Z estimators report a standard error for the Z estimate based on the values obtained from the different samples, perhaps taken over 12 or 24 months. Users should note such standard errors will underestimate the true uncertainty in the parameters, due to the use of constant input values of K and L_∞ . More realistic confidence intervals should be calculated by testing out a range of input values for K and L_∞ .