Multi-Distribution Annually-Pulsed Size-Frequency Data and Growth Increment Data Model Introductory User Manual

By Alison Meynert
Co-op Student
Pacific Biological Station
July 1999

Table of Contents

Introduction 1

Required data 2

Formatting data 3

Initializing the model 4

Loading, saving, and editing algorithm data 7

One function call 8

Minimization 9

Results Summary 11

Calculating Covariance 12

Using both increment and size-frequency data 13

Appendix A: Parameter Descriptions 14

Introduction

This manual is intended for use as an introductory guide to Barry Smith's Age Model program (AgeFrequencyModel.exe). It is not intended to serve as a resource for actual data analysis; rather, its goal is to guide a new user through the functions of the program they are most likely to use. We will follow an example data set through analysis, starting from raw data.

Required Data

This program takes two types of input data: raw size-frequency and growth increment data; therefore we will be using both types of data in our example. This data is for red sea urchins, sampled at Site 3 in the Queen Charlotte Islands in 1998.

Size-Frequency Data Sample:

0.5	0
1.5	0
2.5	1
3.5	0
4.5	2
5.5	3
6.5	7
7.5	12
8.5	13
9.5	19
10.5	20
11.5	16
12.5	18
13.5	18
14.5	15

Increment Data Sample:

1	8.620	8.336	152	516
1	9.009	8.179	152	516
1	9.339	7.896	152	516
1	9.688	11.076	152	516
1	9.832	9.004	152	516
1	9.908	7.420	152	516
1	10.042	9.272	152	516
1	10.070	5.326	152	516
1	10.432	8.643	152	516
1	10.803	6.524	152	516
1	11.028	11.461	152	516
1	11.267	9.595	152	516
1	11.356	7.623	152	516
1	11.407	6.483	152	516

The size-frequency data is in pairs, with the first column being the midpoint of the size interval, and the second being the number of animals found in that interval.

The growth increment data is somewhat more complicated. The first column is called the “set number”, and is usually set to 1 for a single data set. The second column is the initial size and the third is the growth increment. The fourth column is the date of initial measurement, in days from January 1st of that year (Jan. 1st = day 1). The fifth column is the date of final measurement, in days from January 1st of the year the initial measurements were done.

Note: If you do not have growth increment data, but you do have size-at-age data, you can use this data as growth increment data. Assign the set number column to the age of the animal, the initial size to zero, the growth increment to the measured size, the initial date to zero, and the final date to 365 times the animal’s age. If you wish, you can be more specific with the final date and correct for the month in which the animal was caught.

This data is from Microsoft Excel, and in its present form, cannot be read by the analysis program. Our next step, now that we have taken an initial look at what data is required, is to format the data so that it can be used by the program.

Formatting the data

Size-frequency data will need to be in the following form:
######,####,####, …
i.e. fixed width, comma delimited text. The first six characters are read as the midpoint of the size interval. The next four are read as the frequency of animals in that interval. At least one frequency set must be included, but you can have up to eight sets. To optimize model speed, sort the records in order of size interval.

Growth increment data will need to be in the following form:
######,######,######,######,######
i.e. fixed width, comma delimited text. Each set of six characters is read as follows: set number, initial size, growth increment, initial date, final date. You cannot include anything more than this. To optimize model speed, sort the increment records in order of initial size.

We recommend saving Excel (**.xls**) data sets as Comma Delimited Text (**.csv**) files. This will automatically add in the commas you will need, and save you considerable time. For the next formatting step, we recommend using TextPad editor. It is a shareware program available at www.textpad.com in an evaluation version.

Open your data (in **.csv** form) in a text editor such as TextPad or NotePad, and save it with the same filename but with extension **.dat**. This step is a pain to do. Each column of your data must be right-aligned, using spaces, not tabs. The following is an example of correctly formatted data.

Size-Frequency Data Sample:

```
   0.5,    0
   1.5,    0
   2.5,    1
   3.5,    0
   4.5,    2
   5.5,    3
   6.5,    7
   7.5,   12
   8.5,   13
   9.5,   19
  10.5,   20
  11.5,   16
  12.5,   18
  13.5,   18
  14.5,   15
```

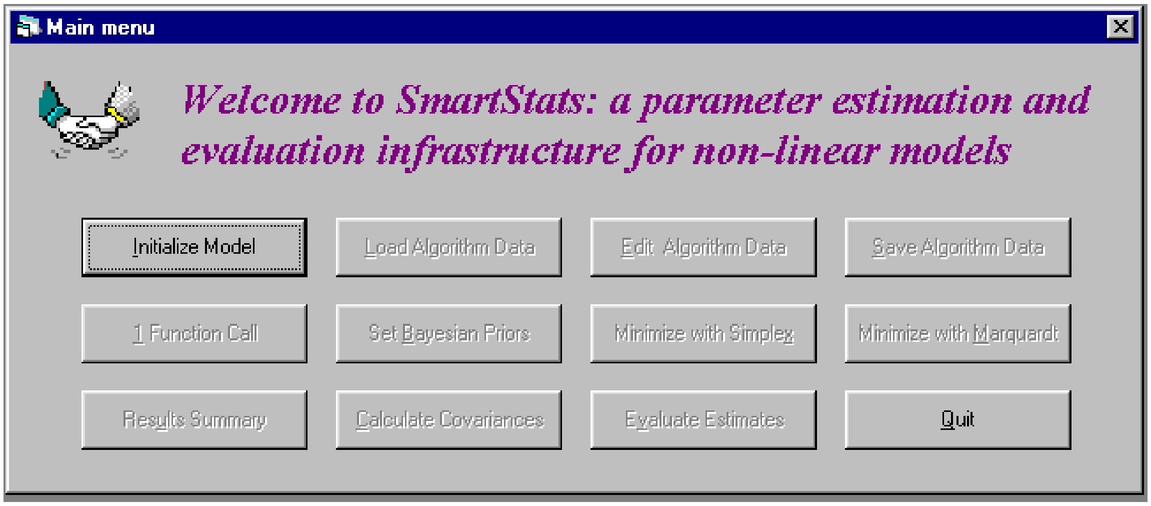
Growth Increment Data Sample:

```
     1, 8.620, 8.336,   152,   516
     1, 9.009, 8.179,   152,   516
     1, 9.339, 7.896,   152,   516
     1, 9.688,11.076,   152,   516
     1, 9.832, 9.004,   152,   516
     1, 9.908, 7.420,   152,   516
     1,10.042, 9.272,   152,   516
     1,10.070, 5.326,   152,   516
     1,10.432, 8.643,   152,   516
     1,10.803, 6.524,   152,   516
     1,11.028,11.461,   152,   516
     1,11.267, 9.595,   152,   516
     1,11.356, 7.623,   152,   516
     1,11.407, 6.483,   152,   516
     1,11.510, 8.574,   152,   516
```

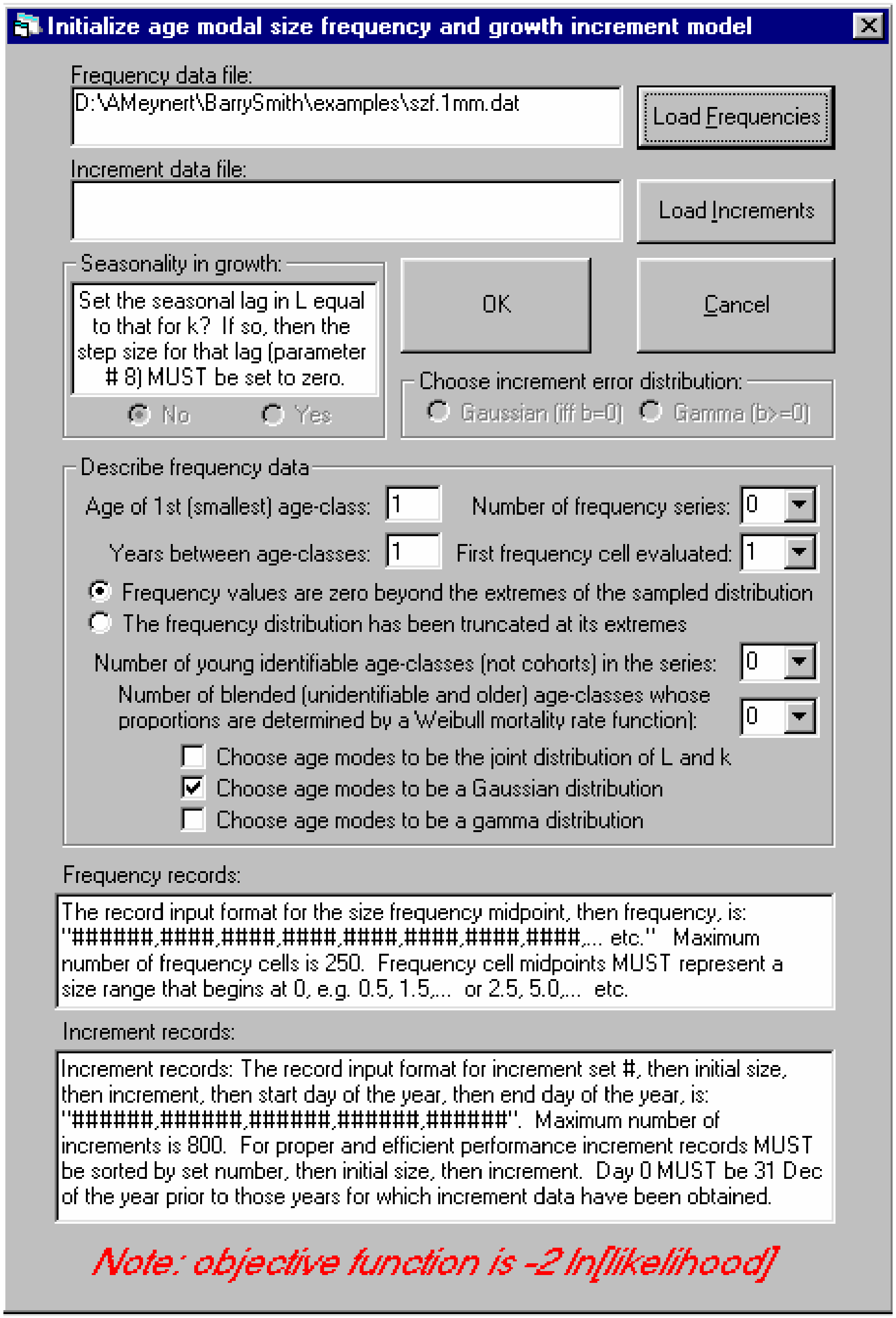
TextPad gives you an advantage when doing this part of the formatting. It has an option, under the Configure menu, that allows selection (and copy and paste) of vertical blocks of text. Ensure your data is correctly formatted, as it can cause you headaches later on if it is not.

Initializing the model

Now we are ready to start the program. It should be in your Start Menu (if you are running Windows 95 or higher), under Age Model. You will see a screen that looks like the following:



Choose Initialize Model from the menu. You will see the following dialog:



As you can see, we have already loaded one of the size-frequency data files. This is a good time to pull out your size-frequency diagrams. We will be referring to it. For our example data, we are going to choose the following options:

- Age of 1st (smallest) age-class: 1
    - This allows you to adjust the starting point of your data
- Years between age-classes: 1
    - This allows you to set the intervals between what you believe are age classes for your data

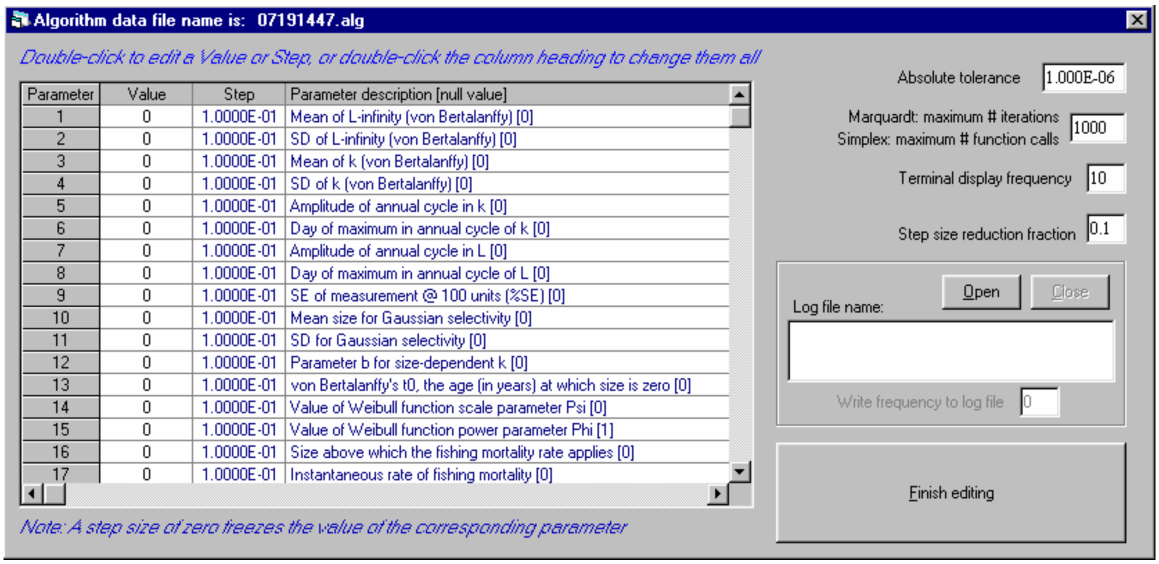
- Number of frequency series: 1
    - This allows you to evaluate up to 8 frequency series at once. Warning: more than one series can more than double the calculation time.
- First frequency cell evaluated: 1
    - If you do not consider the first few cells of data to be reliable, you can skip them with this option.
- Frequency values are zero beyond the extremes of the sampled distribution
    - If you believe you may not have sampled the population extensively enough, you can choose the truncation option. Usually it is best to choose the zero beyond extremes option
- Number of young identifiable age-classes in the series: 2
    - This is when you look at your size-frequency diagram. How many age modes can you identify before they become blended?
- Number of blended (unidentifiable) age-classes in the series: 99
    - Your options here are 0, 5, 10, 20, 50, and 99. The idea is to choose the amount which will take your population to zero by the end of the distribution.
- Choose age modes to be a gamma distribution
    - Combination of L and k is extremely slow; Gaussian and gamma are much faster, with the penalty of slightly less accuracy.

If you are using increment data, ignore the Seasonality in Growth box for the time being, and start by choosing your increment error distribution to be Gaussian, as it is less complicated. You can always come back later and switch to gamma. We will work with size-frequency data for now. The idea is the same for increment data, and it is possible to work with both at the same time.

Loading, saving, and editing algorithm data

Once you have loaded the size-frequency data, choose Load Algorithm Data from the main menu. When asked, choose "No" to load the default values (all zeroes). You will be told what the time-date based name of the algorithm file is. This is a temporary name, as we will be saving it under a more descriptive name later on.

Choose Edit Algorithm Data from the main menu. You will see the following dialog:



To start off, set the maximum number of function calls to 99 999. This will save you the trouble of interrupted calculations later on.

Now, double-click on the word "Value" and choose "Change all values and step sizes". This will allow you to enter your first guesses at the function parameters. See Appendix A for a list of the parameters. For this example, we will enter the following guesses:

Mean of L-infinity: 130
SD of L-infinity: 15
Mean of k: 0.15
SD of k: 0.015
Mean size for Gaussian selectivity: 10
SD for Gaussian selectivity: 10
von Bertalanffy's t0: 0.5
Weibull function parameter Phi: 0.05
Weibull function parameter Psi: 1
Proportion @ age for age-class 1 of frequency series: 0.10
Proportion @ age for age-class 2 of frequency series: 0.30
Day relative to 1 January when series 1 was sampled: 150

You will notice that as you set these parameters, their corresponding step-size is changed proportionately. For all the parameters that you do not set, and which you do not wish to estimate, set their step size to 0 by double-clicking on that cell. We will also set the step size for "Day relative to 1 January" to be 0, because we know that it was May 30, and thus 150 days after 1 January. Again, set the step size to be 0 for the Weibull function parameter Psi. We wish to estimate natural mortality, and when Psi is 1, the parameter Phi is the estimate we are looking for.

Choose the big "Finish editing" button. You will now be back at the main menu. Choose "Save Algorithm Data" and save your algorithm. It is recommended that you keep all the files of one analysis in the same folder or give them all the same filename, just with different extensions.

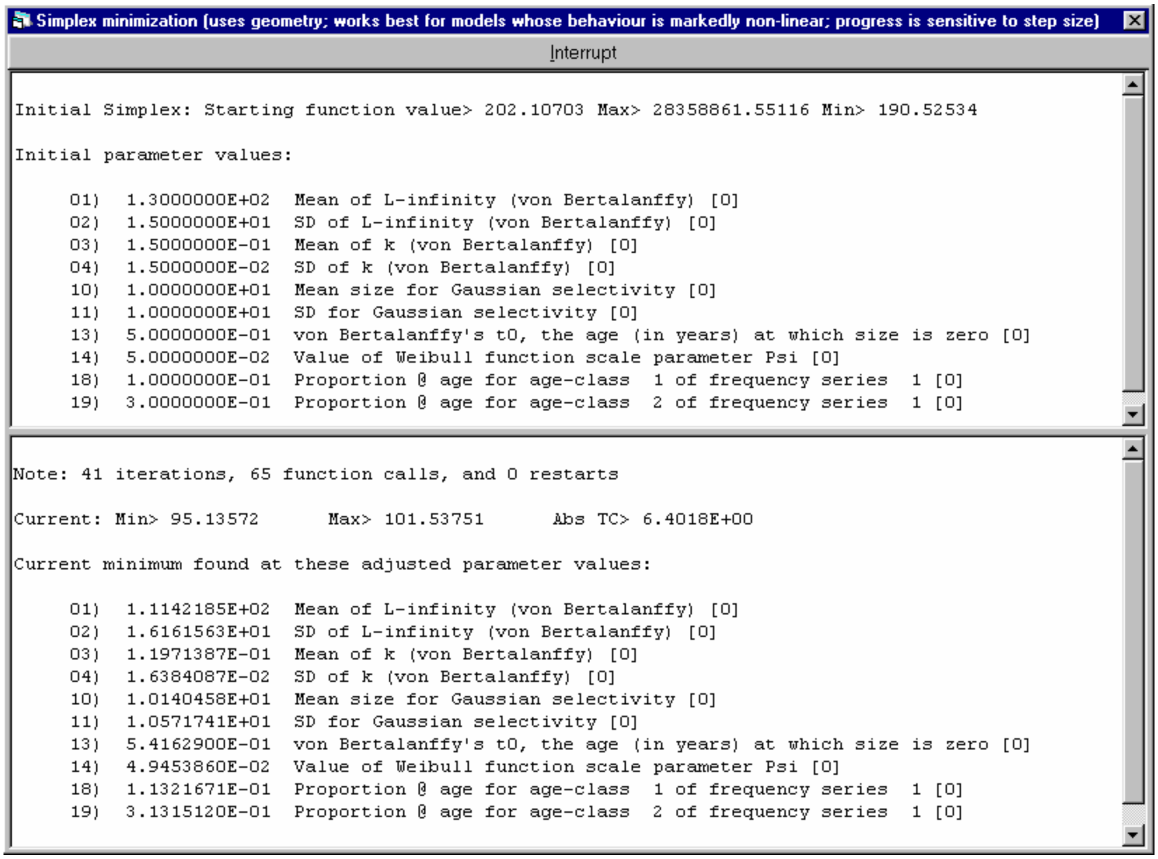
One function call

Now choose 1 Function Call from the main menu. This will give you an idea of how long it will take the program to find the 'best' parameters for your model. Hopefully your time will be under 1 second, as in this example:

```
Current minimum of 202.10703 found at the following parameter values:

     01)  1.3000000E+02  Mean of L-infinity (von Bertalanffy) [0]
     02)  1.5000000E+01  SD of L-infinity (von Bertalanffy) [0]
     03)  1.5000000E-01  Mean of k (von Bertalanffy) [0]
     04)  1.5000000E-02  SD of k (von Bertalanffy) [0]
     05)  0.0000000E+00  Amplitude of annual cycle in k [0]
     06)  0.0000000E+00  Day of maximum in annual cycle of k [0]
     07)  0.0000000E+00  Amplitude of annual cycle in L [0]
     08)  0.0000000E+00  Day of maximum in annual cycle of L [0]
     09)  0.0000000E+00  SE of measurement @ 100 units (%SE) [0]
     10)  1.0000000E+01  Mean size for Gaussian selectivity [0]
     11)  1.0000000E+01  SD for Gaussian selectivity [0]
     12)  0.0000000E+00  Parameter b for size-dependent k [0]
     13)  5.0000000E-01  von Bertalanffy's t0, the age (in years) at
     which size is zero [0]
     14)  5.0000000E-02  Value of Weibull function scale parameter Psi
     [0]
     15)  1.0000000E+00  Value of Weibull function power parameter Phi
     [1]
     16)  0.0000000E+00  Size above which the fishing mortality rate
     applies [0]
     17)  0.0000000E+00  Instantaneous rate of fishing mortality [0]
     17)  1.0000000E-01  Proportion @ age for age-class  1 of frequency
     series  1 [0]
     18)  3.0000000E-01  Proportion @ age for age-class  2 of frequency
     series  1 [0]
     19)  1.5000000E+02  Day relative to 1 January when series  1 was
     sampled [0]

Note: One function call requires 0.55 seconds
```

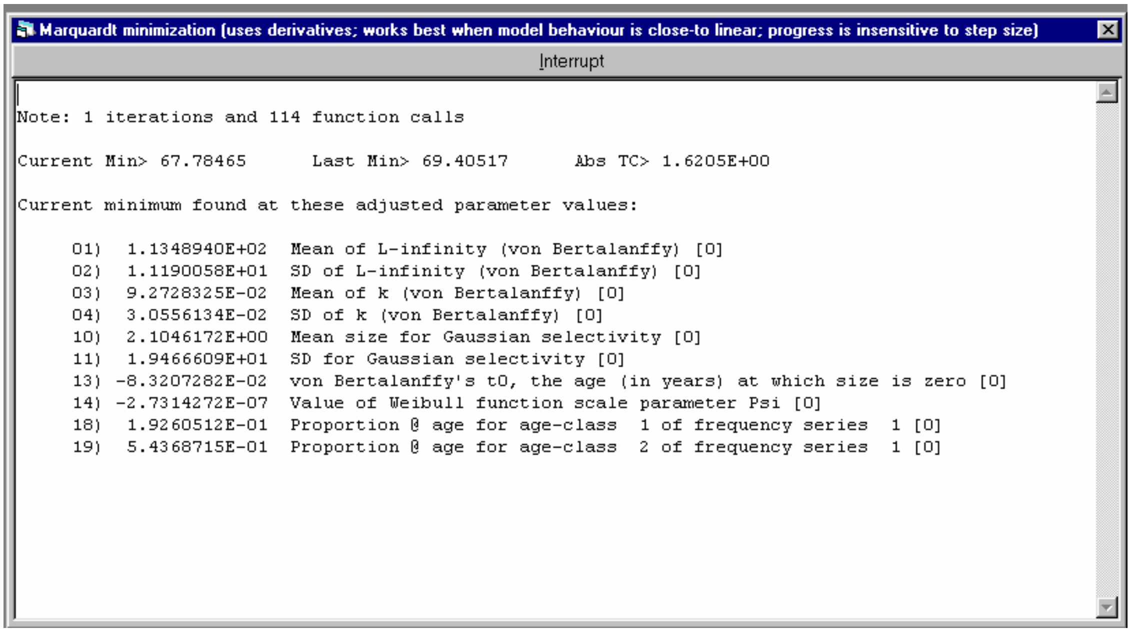
Minimization

There are two methods with which you can minimize your model: Simplex and Marquardt. It is recommended that you use both methods when analysing data. We will start by using Simplex. To run the minimization, click the “Minimize with Simplex” button on the main menu. While the Simplex minimization is running, you should see a window like the following:



Sometimes you will see one of the parameters has a negative value. Don’t worry, the program has the correct value stored.

Once the program has run through the Simplex minimization, it is a good idea to save your algorithm data. Next, try running the data through the Marquardt minimization, using numerical derivatives. Don't worry about the Bayesian Priors stuff. We won't be using it in this example. The Marquardt minimization window looks like the following:



Marquardt is slower than Simplex, but can obtain a lower minimum value.

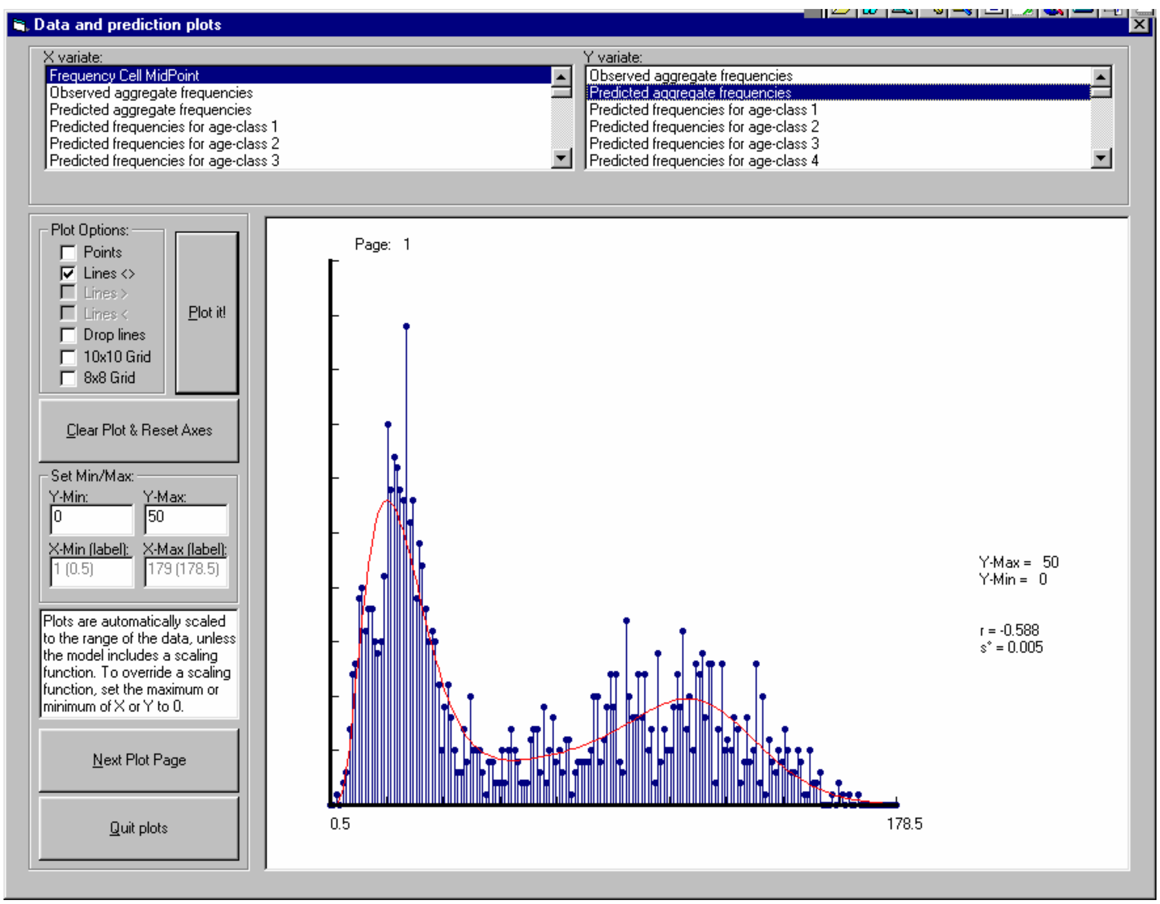
In the current example, I found that the mean size and standard deviation for Gaussian selectivity were not changing very much, so I froze those parameter values by interrupting the minimization and changing their step sizes to 0 in the Edit Algorithm Data window. I then set their values to be 5 and 2.5, respectively. You can interrupt the minimization process at any time by hitting the Interrupt bar at the top of the window.

Results Summary

When the program has identified a minimum, save your algorithm data. Next, choose Results Summary from the main menu. It will run One Function Call, then ask if you would like to see the plots as well as the summary. If you choose yes, a plotting window will come up.

Choose “drop lines” from the Plot Options box on the left, then hit “Clear Plot & Re-Set Axes”. Double-click on “Observed aggregate frequencies” in the top right-hand box. Your original size-frequency data will plot. Choose “lines” from the Plot Options box, then double-click on “Predicted aggregate frequencies” in the top right-hand box. You can double-click repeatedly to change the plot colour.

The following is a plot done before the minimization was completed for this data set:

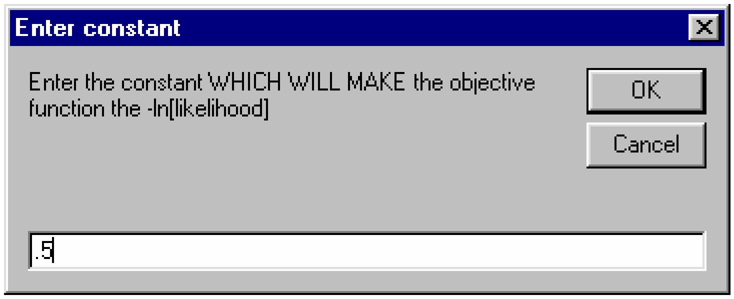


When you are done looking at your plots, click the Quit Plots button in the bottom left-hand corner. You will be asked how many parameters were estimated in this model. In the example data, there are nine. You will also be asked how many chi-square calculations you want done. 1000 is a good round number, and is usually fairly accurate. You will then be asked how many age-classes you want detailed reports for. For the example, choose 2. You will usually only be interested in the ones for which you have visible age modes in the data.

You can save your output as a **.sso** file.

Calculating Covariances

Once you have looked at your results, you can look at the covariances. Choose Calculate Covariances from the main menu. You will be asked about something called “grid factors”. The defaults are good, so I don’t recommend changing them. Just agree with the program until it asks you to:



Always enter 0.5, as the objective function is –2*ln[likelihood], as you will recall from the screen shot of the model initialization on page 5.

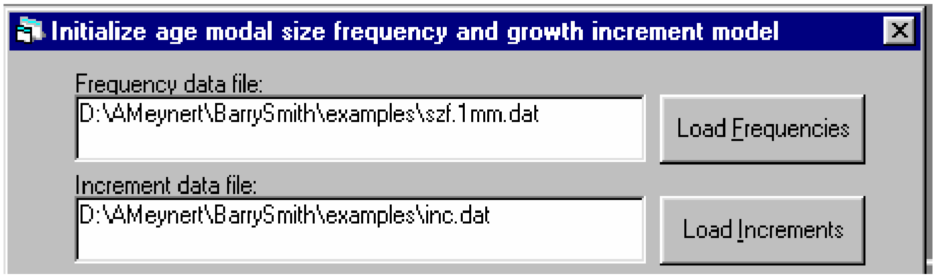
With a little luck, you will end up, after a moment of calculations by the program, with output similar to the following:

```
Covariance constant: 0.5

Factors:          1) 1.0000000E+00  2) 1.0000000E-01  3) 1.0000000E-02

Likelihood function value summary:
Mean:             1.4557005E+02     1.2819633E+02     1.2802132E+02
Std Dev:          2.5565022E+01     3.3888996E-01     2.1380607E-02
Coef of Var:      0.1756201         0.0026435         0.0001670
```

The closer the means are for each column, the better. You can save your covariance calculations as a **.cov** file.

Using both increment and size-frequency data

Increment data can help to offset variation in size-frequency data. By loading both increment and size-frequency data, as in the following example, your calculations will often be more accurate.



Everything else is done exactly as in the size-frequency-only example.

Appendix A: Parameter Descriptions

1	Mean of L-infinity (von Bertalanffy)
2	Standard deviation of L-infinity (von Bertalanffy)
3	Mean of k (von Bertalanffy)
4	Standard deviation of k (von Bertalanffy)
5	Amplitude of annual cycle in k
6	Day of maximum in annual cycle of k
7	Amplitude of annual cycle in L-infinity
8	Day of maximum in annual cycle of L-infinity
9	Standard error of measurement @ 100 units (percent standard error)
10	Mean size for Gaussian selectivity
11	Standard deviation for Gaussian selectivity
12	Parameter b for size-dependent k
13	Von Bertalanffy's t_0, the age (in years) at which size is zero
14	Value of Weibull function scale parameter Psi
15	Value of Weibull function power parameter Phi
16	Size above which the fishing mortality rate applies
17	Instantaneous rate of fishing mortality
18	Proportion @ age for age-class 1 of frequency series 1
19	Proportion @ age for age-class 2 of frequency series 1
20	Day relative to January 1 when series 1 was sampled