User's Guide

CALK — Combining Age-Length Keys: a C program to estimate age composition of a fish population using prior and current information.

Alberto G. Murta

IPIMAR - Institute of Fisheries and Sea Research, Avenida de Brasilia 1449-006 Lisboa, Portugal, e-mail: amurta@ipimar.pt.

1 Introduction

Most stock assessment techniques currently used in temperate waters fisheries have catch-at-age data as input. Age data collection is time and money consuming, so several methods have been developed which use prior information to convert length to age. Probably the most flexible of these methods is the one described by Hoenig et al. (1993, 1994, 2002), which can combine several age-length keys (ALK) and apply the combined ALK to several vectors with length frequencies. That is the method applied by this software. This method assumes that the proportions of each length-class in each age-class are the same in the combined ALK and in the populations from where the length data was obtained. Therefore, differences in age distributions at length, such as the ones caused by strong recruitments, do not invalidate the application of the method. However, differences in growth rate, or changes in the selection pattern of a fishery may

invalidate the application of this method. The probability that the assumption of homogeneous length distributions at age is violated increases as data collected further away in time is used together. Therefore, it is wise to only combine data sets collected within a relatively short period of time, which depends on the life span of the species being studied. A detailed description of the method is given in the original papers (Hoenig et al., 1993, 1994, 2002).

2 How to use the program

2.1 Format of the input data file

Suppose we have some catch data sets with and without age data (denoted respectively by k and z), for instance from different years. n_{ij}^k is a matrix with number of fish by length-class (i) and age-class (j), from year k. This matrix can be obtained by simple random sampling or length-stratified sampling from the catches of year k. f_i^k and f_i^z are vectors with number of fish catched by length-class in years k and z respectively. In the case that we have, for example, 4 data sets, 2 with age data ($n_{ij}^1, f_i^1, n_{ij}^2, f_i^2$) and 2 without age data (f_i^3, f_i^4), with 10 length-classes and 5 age classes, the input file would be made of a matrix with 10 rows (length-classes) and 14 columns, distributed in the following way: first 5 columns are from n_{ij}^1, next 5 columns from n_{ij}^2, and the last 4 columns are f_i^1, f_i^2, f_i^3, and f_i^4 (see example input file).

2.2   Input file check-list:

- Input data file has to be in text (ascii) format, without any titles.

- Input file may not have more than 10 data sets with age data and 20 data sets without age data.

- Data sets may not have more than 100 length-classes and 30 age-classes.

- Data have to be arranged in a single matrix, with space separated values.

- Given that input data are number of fish, decimal places are not necessary. However if they are used a dot must be used as decimal separator.

- All data sets must have the same number of length and age-classes.

- In a data set with age data, all length-classes present in the length frequencies vector (f_i^k) have to be also present in the corresponding matrix with number of fish by length and age-class (n_{ij}^k).

- Each length class must have at least one individual in all data sets with age data (f_i^k and n_{ij}^k).

- Each age class has to be present in all matrices with number of fish by length and age-class (n_{ij}^k).

- Each length-class in the length frequencies vectors of data sets without age data (f_i^z), has to be present in at least one of the matrices of the data sets with age data (n_{ij}^k).

REMEMBER: It is better to spend time checking the format of the data before running the program, than trying to spot possible mistakes afterwards.

2.3 Running the program

Presently there are compiled versions of the program to run in Linux (calk) and in DOS/MS Windows (calk.exe). The Linux version is started entering . /calk in the shell prompt, and the DOS/MS Windows version must be started entering calk in the DOS prompt. In MS Windows is better to open a DOS window and then start the program, rather than clicking with the mouse in the executable file.

When the program starts you'll see the following lines:

```
COMBINED AGE-LENGTH KEYS
(1999) Alberto G. Murta, IPIMAR - Lisbon
Name of data file?(max. 20 chars)
```

After entering the name of the input data file, you'll be asked for the number of length and age-classes:

```
Number of length classes?
```

```
Number of age classes?
```

Then you'll be asked for the number of data sets (e.g. number of years) for which there are age information (each data set is made of a matrix with number of fish with known age, by length and age-class, and the corresponding vector of length-class frequencies):

```
Number of data sets WITH age data?
```

Finally you have to enter the number of data sets with only length-class frequency data:

```
Number of data sets WITHOUT age data?
```

Then, the program will give some messages reporting what's going on, and if all goes well the following message appears:

```
Converged after (a number less than 3000) iterations.
Writing output file...
```

This means that the iterative process has stopped and the output file calk.out is being written. This output file has three sections. The 1st, with the title "Inverse Age-Length Key" corresponds to the matrix with the proportion of each length class (rows) in each age class (columns). The 2nd section has the title "Number of individuals by length and age-classes" and contains the same number of matrices as input data sets (with and without age data). These matrices are ordered in the same way as the length frequencies vectors in the input file, and are useful to calculate classic age-length keys, mean length-at-age or the standard deviation of length-at-age for each data set. Finally, the 3rd section has the proportions of each age-class in each data set (which can be also obtained from the matrices above).

In case there was no convergence, the following message appears:

```
Did not converge after 3000 iterations. Please check your data.
```

This method is based on the EM algorithm, which has proved in simulation studies to have good convergence properties. However sometimes convergence may not occur due to inconsistencies in the data. This is more likely to happen if many data sets are being combined. For example if age data from a large time span are combined, there is a great probability that the length distributions at age have changed during that time, therefore violating the assumption of the method. In this case is better to separate data sets in shorter time periods (e.g. around 5 years) and apply the method for each time period separately.

3   References

- Hoenig, J.M., Heisey, D.M. and Hanumara, R.C. (1993) Using prior and current information to estimate age composition: a new kind of age-length key. International Council for the Exploration of the Sea. C.M. 1993 / D: 52. 11 pp.

- Hoenig, J.M. Heisey, D.M. and Hanumara, R.C. (1994) A Computationally simple approach to using current and past data in an age-length key. International Council for the Exploration of the Sea. C.M. 1994 / D: 10. 5 pp.

- Hoenig, J.M., Hanumara, R.C., Heisey, D.M. 2002. Generalizing double and triple sampling for repeated surveys and partial verification. Biometrical Journal, 44: 603–618.