ALKr: A R package of methods based on age-length keys to estimate the age structure of fish populations

Alberto G. Murta[a]

José Francisco Loff[b]

Manuela Neves[c]

Laura Wise[a]

[a]Instituto Português do Mar e da Atmosfera (IPMA). Avenida de Brasília, 1449-006 Lisboa. Portugal; Tel.: +351 21 3027120; Fax: +351 21 3015948; email: amurta@ipma.pt.

[b]JFL - Consultoria Estatística. Rua Carvalhão Duarte, 8 - 4C, 1600-423 Lisboa. Portugal; Tel.: +351 91 7722012; Fax: +351 21 7578159; email: jfloff@phistat.com.

[c]Instituto Superior de Agronomia. Departamento de Matemática. Tapada da Ajuda. Lisboa. Portugal; email: manela@isa.utl.pt.

Abstract

In this work we review and compare the methods currently existing for estimating the age composition of a fish population, using a random sample stratified by length class, in which the age of each individual is determined, e.g. by otolith reading. The methods were implemented as an R package and compared based on the bias and variance of the estimators of the proportion of each age in the population, mean length-at-age and standard deviation of length-at-age. A simulated data set with known parameters was created, from which 1000 random samples, stratified by length class, were taken by a bootstrap procedure. Every method was applied to each bootstrap sample, and the deviation from the real value to each bootstrap estimate was calculated and used to compare the accuracy and precision of each parameter across methods. The advantages and disadvantages of each method are discussed, while the main results point out a method based on the Expectation-Maximisation algorithm as being the most flexible, precise and accurate.

Keywords: Age-length keys, bootstrap methodology, EM algorithm, maximum likelihood estimators.

Introduction

Many of the methods currently applied for fish stock assessment, such as cohort analysis and statistical catch-at-age models, use catch in numbers-at-age and catch-per-unit-effort (CPUE) in numbers-at-age as input data. Therefore, the estimation of the relative abundance of each age class in the data has a great influence on the final result of the stock assessments performed with those methods.

The most obvious method to estimate the empirical distribution of fish into age-classes would be to determine the ages of individuals from a simple random sample, and assume the proportion of each age-class in the sample to be representative of the whole population (population here is used in the statistical sense, not the biological one, and can refer to e.g. the fish caught by a given fleet at a given time). Although this method has some good asymptotic statistical properties (Kimura, 1977) it is not widely used due to practical reasons. Usually the bulk of the catches in a given fishery belong to just a few age-classes, making it difficult to cover all age-classes with a simple random sample of feasible size. Therefore, unless large samples are taken, this method may give biased results for the age-classes less frequent in the catches. Moreover, does not provide estimates for parameters such as mean length-at-age, which may be important

1

indicators regarding the population dynamics of the stock.

Given that length measurements are much easier to obtain than to determine the age of a fish, good estimates of length distributions are usually available from routine sampling. Thus, age distributions are usually estimated from length distributions using one of the following approaches:

- Length frequency analysis, which takes a length frequency distribution as a mixture of Gaussian distributions of length-at-age (e.g. Hasselblad, 1966; Bhattacharya, 1967; Schnute and Fournier, 1980; Macdonald, 1987; Fournier et al., 1990) The results from these methods are not very reliable (Hilborn and Walters, 1992), therefore their estimates must be taken with precaution.

- Fitting of a growth model to attribute ages to fish from given length-classes. This is also known as a deterministic age-length key (Hilborn and Walters, 1992) and is mainly used when there are insufficient age data to make a classic age-length key (Bartoo and Parker, 1983). Strong criticisms have been made to this procedure. The growth model most frequently used (von Bertalanffy, 1938) is based on unrealistic assumptions (Ricker, 1975) and some of its parameters have unclear biological meaning (Knight, 1968; Roff, 1980). Even models for individual growth made on sounder assumptions than von Bertalanffy's (e.g. Schnute, 1981) will produce biased estimates when applied to a population Mulligan and Leaman (1992). Stochastic models which account for individual variability in growth have been developed (Fabens, 1965; Sainsbury, 1980; Kirkwood and Somers, 1984; Hampton, 1991; Xiao, 1996), but Wang and Thomas (1995) showed that estimates produced by those models are far from satisfactory.

- Calculation of an age-length key (Friðriksson, 1934) also known as a distribution matrix (Hilborn and Walters, 1992), to distribute the number of individuals in a given length-class through age-classes. Each cell of the age-length key (ALK) gives the proportion of fish belonging to an age-class j, given it belongs to a length-class i: $Pr(j \mid i)$. This is the most used method to estimate an age distribution from a length distribution, when age data are available.

The classic ALK

The classic ALK (Friðriksson, 1934) is a matrix (Q) with elements given by

$$q_{ij} = a_{ij} \div \sum_{j} a_{ij} \tag{1}$$

where a_{ij} is the number of individuals with known age j in length-class i. The estimated number of individuals in a given length and age-class is

$$n_{ij} = q_{ij} \times f_i \qquad (2)$$

where f_i is the number of fish in length-class i. From each column of this N matrix, the calculation of the proportion of each age-class (p_j), of mean length-at-age $(\bar{l}_j)$ and of the standard-deviation (SD) of length-at-age (σ_j) is straightforward. A classic ALK can be calculated from an A matrix and a f_i vector using function `classic_ALK` in the `ALKr` package.

In the ideal case of application, a sample is taken from a given population, and the length of the fish in that sample is measured. Then, a subsample is taken from that sample (either by simple random sampling, or by random sampling stratified by length-class), and the ages of all fish in that subsample are determined, e.g. by analysing the otoliths or other rigid structures. Then, the ALK obtained from the subsample is used to estimate the age-frequencies of the fish in the initial sample. However, in fisheries research it is current practice to use ALKs obtained from a given population to length-frequency samples of other populations. For example, such a case would be to use an ALK made from a sample representing the catch of a given fleet in a given year, and use it with a length-frequency sample taken in another year. The two populations (the fish that were caught by that fleet in those two years) may have different characteristics, and in this case the use of the classic ALK relies on the basic assumption that $Pr\,(j \mid i)$ must be the same in the catch of both years (Kimura, 1977; Westrheim and Ricker, 1978).

Several factors, such as different survival rates or recruitments in the two populations may cause this assumption to fail. For example, if the recruitment in the year of the length-frequency sampling was higher than in the year when the ALK was built, it is expected that the proportion of recruits in some length classes will be different in the ALK and in the length-frequency sample. In that case, the estimates will be biased positively or negatively depending on the relative difference between the two recruitments (Westrheim and Ricker, 1978). This assumption of the classic ALK must be checked particularly when the sampling for the ALK was made at a given time or place and one wishes to use it for estimating age distributions from length-frequency samples collected at a different time or place. When this assumption is not met the estimates produced by the ALK reflect the age structure of the sample used to construct it, independently of the real age structure underlying the length-frequency samples (Clark, 1981). Given that some of the factors contributing to this possible bias can be very variable across years (e.g. recruitment), in many cases it is usual to calculate a new ALK

every year.

The fact that the $Pr\,(j \mid i)$ in a given ALK is different from those in the length-frequency data, does not imply that the inverse probabilities of being in a certain length-class, given being in a certain age-class ($Pr\,(i \mid j)$) are also different. Taking the example of a variable recruitment, a higher number of individuals at recruitment age in the ALK sample than in the length-frequency sample will cause differences in the proportion of each age in the length-classes corresponding to recruits. However, the proportion of each length-class at recruitment age may remain similar in both samples. This means that by using the inverse probabilities $Pr\,(i \mid j)$ instead of $Pr\,(j \mid i)$ one can, in certain cases, loosen the basic assumption of the classic ALK. This is done using methods based of inverse ALKs.

Method based on the classic ALK

A method based on the classic ALK was described by Martin and Cook (1990) with the objective of improving the estimates of the mean length-at-age and SD of length-at-age, in relation to those obtained using the classic ALK. In this method the parameters of interest are calculated by minimising the function

$$L = 2 \times \sum_i \left(f_i \times \log \frac{f_i}{\widehat{f}_i} \right) + 2 \times \sum_i \sum_j \left(a_{ij} \times \log \frac{a_{ij}}{\widehat{a}_{ij}} \right)$$

where $\widehat{f}_i$ is given by

$$\widehat{f}_i = \sum_i f_i \times \sum_j Pr(i,j)$$

and $\widehat{a}_{ij}$ is given by

$$\widehat{a}_{ij} = Pr(i,j) \times \sum_j a_{ij} \div \sum_j Pr(i,j).$$

$Pr(i,j)$, the proportion of fish in length class i and age class j, is calculated as

$$Pr(i,j) = p_j \times \left(\Phi\left(\frac{l_i + 0.5 - \bar{l}_j}{\sigma_j} \right) - \Phi\left(\frac{l_i - 0.5 - \bar{l}_j}{\sigma_j} \right) \right)$$

where l_i is the length of fish in length class i and Φ is the reduced Normal cumulative distribution function.

However, Murta (1998) has shown that this method performs poorly, with a high sensitivity to the initial values chosen for the optimization of the objective function.

These performance problems were later confirmed by one of the authors of the method (Martin, personal communication). Therefore, this method was excluded from the comparative review and from the software package described in this paper.

The inverse ALK

Age data collection is time and money consuming, so several methods have been developed which may be applied with less restrictions than the classic ALK. All these methods are based on inverse ALKs, in which each cell gives the proportion of fish in a length-class, given it belongs to a certain age-class: $Pr\,(i \mid j)$. All methods based on inverse ALKs are used under the assumption that $Pr\,(i \mid j)$ is the same in the sample used to obtain the inverse ALK and in the sample used to obtain the length-frequency distribution to which the ALK will be applied. With such methods it is possible to reduce the amount of new age data needed, using age data collected at a different time or place than the length-distribution sample. However, changes in the selectivity pattern of the gears used to obtain the different samples, or differences in growth rate between the populations from where the samples came would also invalidate the use of methods based on inverse ALKs, given that in those situations the probabilities $Pr\,(i \mid j)$ are not the same in the inverse ALK and in the length-frequency samples.

The inverse ALK (Clark, 1981; Bartoo and Parker, 1983; Hilborn and Walters, 1992) corresponds to a matrix $Q^{'}$ that is calculated from a N_1 matrix (see expression 2) as

$$q^{'}_{ij} = \widehat{n}_{ij1} \div \sum_i \widehat{n}_{ij1}, \tag{3}$$

where a_{ij1} and f_{i1} are the age and length data collected at time or place 1. This inverse ALK is then combined with the length distribution data f_{i2}, from time or place 2, from which age data is not available, in order to estimate a matrix N_2. The elements of N_2 are given by

$$n_{ij2} = Inv(q^{'}_{ij})^t \times f_{i2}$$

where $Inv(q^{'}_{ij})^t$ is element of the transposed generalized inverse matrix of $Q^{'}$, obtained by the method of singular value decomposition (Hilborn and Walters, 1992; Leon, 1994). The parameters of interest (p_{j2}, $\bar{l}_{j2}$ and σ_{j2}) are then calculated from N_2. This is the most basic way of applying an inverse ALK, which may be prone to numerical problems, especially in the calculation of the generalized inverse matrix. An ALK calculated by this method can be obtained from a matrix A and two vectors f_{i1} and f_{i2} by calling function `inverse_ALK` in package `ALKr`.

5

Methods based on the inverse ALK

Given that the simple application of an inverse ALK by calculating its generalized inverse matrix is prone to numerical instability, several iterative methods have been described, which allow the use of an inverse ALK in a way that is numerically more robust. The first of those methods, and probably still the most popular one in fisheries research, was described by Kimura and Chikuni (1987). In this approach, an inverse ALK is initially obtained as in expression 3 and initial values are given to p_{j2}. A classic ALK is then calculated:

$$q_{ij2} = q_{ij}^{'} \times p_{j2} \div \sum_{j} \left(q_{ij}^{'} \times p_{j2} \right)$$

and N_2 is estimated as

$$n_{ij2} = f_{i2} \times q_{ij2}.$$

From this matrix, new p_{j2} are obtained and the process is repeated iteratively until convergence is achieved. Kimura and Chikuni's method can be used to calculate an ALK by calling function `kimura_chikuni` in the `ALKr` package.

In a paper published later in the same year, Hoenig and Heisey (1987) describe a similar method, but with some refinements, and show that both Kimura and Chikuni's method and their own are similar applications of the Expectation-Maximization (EM) algorithm (Dempster et al., 1977). The EM algorithm is a family of methods sharing the same basic principles. Having an incomplete data set, a model to provide the expected values for the missing observations, and a way to calculate the maximum-likelihood estimates of the parameters of that model, a method based on the EM algorithm, such as the one by Kimura and Chikuni (1987), may include the following 4 steps (Schafer, 1997):

1. give initial values to the parameters;

2. complete the data set by calculating the expected values for the missing data using the model with the initial values for the parameters (E step);

3. using the complete data set, calculate the maximum-likelihood estimates of the parameters (M step);

4. with the new parameters' estimates, start a new iteration by recalculating the expected values of the missing data.

In their method, instead of using an inverse ALK that is kept fixed during the iterative process as Kimura and Chikuni (1987) did, Hoenig and Heisey (1987) calculate

the inverse ALK with all available data and make the inverse ALK (Q') to be updated, together with the p_{j2} estimates, until convergence is achieved. Thus, the method has the following steps:

1. give initial values to N_2 as

$$n_{ij2} = f_{i2} \times \frac{n_{ij1}}{\sum_j n_{ij1}};$$

2. calculate the inverse ALK using all available information:

$$q'_{ij} = \frac{n_{ij1} + n_{ij2}}{\sum_i (n_{ij1} + n_{ij2})};$$

3. calculate $p_{j2} = \sum_i n_{ij2} \div \sum_i \sum_j n_{ij2}$;

4. update N_2 as:

$$n_{ij2} = f_{i2} \times \frac{q'_{ij} \sum_i n_{ij2}}{\sum_j \left(q'_{ij} \sum_i n_{ij2} \right)};$$

5. repeat steps 2 to 4 until convergence is achieved.

This method is included in the `ALKr` package as function `hoenig_heisey`.

The method by Hoenig and Heisey (1987) was later extended to include any number of data sets with and without age data, in order to improve the quality of the estimates of the parameters of interest (Hoenig et al., 1993, 1994, 2002). Suppose we have some catch data sets with age data and others without age data (denoted respectively by k and z), for instance from different years or places. So a_{ijk} is an element of a matrix with the number of sampled 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_{ik} and f_{iz} are elements of vectors with the number of fish caught by length-class, and p_{jk} and p_{jz} are the proportions (to be estimated) of the catches in each age class j in year (or place) k and z, respectively. The first step of the algorithm is to calculate a classic ALK for each of the data sets with catch data

$$q_{ijk} = a_{ijk} \div \sum_j a_{ijk},$$

and the corresponding number of fish caught in year k, by length and age-class:

$$n_{ijk} = q_{ijk} \times f_{ik}.$$

Then, an inverse ALK is calculated using the data from all years with age data,

$$q'_{ij} = \sum_k n_{ijk} \div \sum_k \sum_i n_{ijk},$$

and initial values are given to the proportion of catches at age for the years (z) without age data (p_{jz}). With these, is then possible to give initial values for each matrix N_z, for the years (z) without age data:

$$n_{ijz} = q'_{ij} \times p_{jz} \times \sum_i f_{iz}.$$

Maximum-likelihood estimates (M step of the EM algorithm) for p_{jk} and p_{jz} are then obtained as

$$\begin{cases} p_{jk} & = \sum_i n_{jk} \div \sum_i \sum_j n_{ijk} \\ p_{jz} & = \sum_i n_{jz} \div \sum_i \sum_j n_{ijz} \end{cases} \tag{4}$$

and an updated inverse ALK can then be calculated using all data sets,

$$q'_{ij} = \frac{\sum_k n_{ijk} + \sum_z n_{ijz}}{\sum_k \sum_i n_{ijk} + \sum_z \sum_i n_{ijz}}. \tag{5}$$

The number of fish caught by length and age class in each year can be updated as

$$\begin{cases} n_{ijk} & = q_{ijk} \times \sum_j a_{ijk} + \frac{(f_{ik} - \sum_j a_{ijk}) \times q'_{ij} \ p_{jk}}{\sum_j (q'_{ij} \times p_{jk})} \\ n_{ijz} & = \frac{f_{iz} \times q'_{ij} \times p_{jz}}{\sum_j (q'_{ij} \times p_{jz})} \end{cases} \tag{6}$$

and the iterative process continues by repeating expressions 4 to 6 until convergence occurs. In the `ALKr` package, this method is implemented by function `hoenig`.

Gascuel (1994) described another extension to the Kimura and Chikuni (1987) method, combining it with a model for the distribution of lengths at each age-class. For the description of this method, Gascuel (1994) had in mind fish species for which the age structure of the populations were traditionally estimated using length-frequency analysis (e.g. albacore from the East-Atlantic). In this method, the distribution of length at each age-class (the inverse ALK) is assumed to follow a Normal distribution

$$q'_{ij} = \frac{1}{\sigma_j \sqrt{2\pi}} \times \exp\left(-\frac{(l_i - \bar{l}_j)^2}{2\sigma_j^2} \right),$$

where the SD of length at age, σ_j, is given by a linear model as a function of three

8

parameters, α, β, and γ:

$$\sigma_j = \alpha + \beta \times \bar{l}_j + \gamma \times \Delta\bar{l}_j,$$

where $\Delta\bar{l}_j$ is the difference between the mean lengths at age-class j and age-class $j-1$. For each combination of values of these three parameters, the inverse ALK is applied to the length frequency data using an iterative procedure similar to the one described by Kimura and Chikuni (1987):

1. initial values are given to the p_{j2};

2. an iterative process is initiated by calculating an estimate for f_{i2}:

$$\widehat{f}_{i2} = \sum_j \left(p_{j2} \times q'_{ij} \right) \times \sum_i f_{i2};$$

3. new estimates for p_{j2} are then obtained, to be used in the next iteration:

$$p_{j2}^{next} = \sum_i \left(p_{j2}^{previous} \times q'_{ij} \times \frac{f_{i2}}{\widehat{f}_{i2}} \right);$$

4. steps 2 and 3 alternate until convergence is reached.

The parameters of interest can then be calculated from the matrix N_{ij2},

$$n_{ij2} = p_{j2}^{final} \times q'_{ij} \times \sum_i f_{i2},$$

where p_{j2}^{final} is the p_{j2} from the last iteration. The best estimates for the parameters α, β, and γ are then obtained with an optimization process that minimizes the differences between f_{i2} and $\widehat{f}_{i2}$. The function `gascuel` included in the `ALKr` package implements this method, using Nelder and Mead (1965) ``simplex'' method for the optimization procedure.

Procedure for the comparison of methods

Given the availability of the different methods described above, based on classic and inverse ALKs, some of them are expected to perform better than others. Therefore, we use simulated data to compare those methods in terms of precision and accuracy. A total of 6 methods based on ALKs are compared:

9

- the classic ALK (Friðriksson, 1934);

- the inverse ALK (Clark, 1981; Bartoo and Parker, 1983; Hilborn and Walters, 1992);

- an iterative application of the inverse age-length key with the EM algorithm (Kimura and Chikuni, 1987);

- a refinement of Kimura and Chikuni (1987) method proposed by Hoenig and Heisey (1987);

- using prior and current information to estimate age composition with the EM algorithm (Hoenig et al., 1993, 1994, 2002);

- applying inverse ALKs as described by Gascuel (1994).

The first method was applied to data sets that were simulated as if both the length frequency data and the age data were obtained from the same population, which is in accordance with the classic ALK assumption. The other five methods were applied to data sets simulated as if there were length frequency and age data from one population, and just length frequency data from another population with a slightly different demographic structure, to which we wish to apply an inverse ALK.

Data simulation

The generation of the input data used for all methods is described in detail by Murta (1998). The simulated data sets correspond to two *Trachurus trachurus* (Horse mackerel) populations sampled in 1992 and 1993 during bottom-trawl research surveys off the Portuguese coast. In this context we use the term population in its statistical meaning, not in the biological one. The length distributions were taken from the survey's data base, and the distribution of those individuals through age classes was done as follows:

1. Mean-lengths-at-age were defined based on the existing data for the species, and were assumed to remain constant in time (therefore, the fish caught in 1992 and 1993 had the same growth rate).

2. Standard deviations (SD) of length-at-age for age 0 was defined as $\sigma_0 = 2$ and for age 15+ as $\sigma_{15+} = 4$. The SD for the other age classes were defined as

$$\sigma_j = \sigma_0 + \frac{(\sigma_{15+} - \sigma_0) \times (\bar{l}_j - \bar{l}_0)}{\bar{l}_{15+} - \bar{l}_0}$$

where $\bar{l}_j$ is the mean length at age j.

3. The proportion of each age-class in each population were calculated by applying the classic age-length keys made in each year to the length distributions of the corresponding year.

4. An inverse age-length key, giving the probability on an individual belonging to length-class l given it belongs to age-class j ($\Pr(i \mid j)$), which is common to both populations, was calculated assuming a Normal distribution of lengths at age:

$$\Pr(i \mid j) = \Phi\left(\frac{l_{i+0.5,j} - \bar{l}_j}{\sigma_j}\right) - \Phi\left(\frac{l_{i-0.5,j} - \bar{l}_j}{\sigma_j}\right),$$

where Φ is the cumulative distribution function of the Normal distribution.

5. Finally, the probability of an individual being in age-class j and length-class i in data set d was given by $\Pr(i,j)_d = \Pr(i \mid j) \cdot_{j,d}$.

The simulated data are included in the `ALKr` package as data set `hom`, in which items `N1992` and `N1993` are matrices containing the number of individuals of a given age (columns) in each length class (rows) collected in 1992 and 1993. Items `F1992` and `F1993` are the row sums of `N1992` and `N1993`, that is, they contain the number of fish in each length class.

```
> library(ALKr)
> data(hom)
> hom$N1992
```

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
7	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	8	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	42	53	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	173	235	1	0	0	0	0	0	0	0	0	0	0	0	0	0
11	562	836	4	0	0	0	0	0	0	0	0	0	0	0	0	0
12	1424	2380	14	0	0	0	0	0	0	0	0	0	0	0	0	0
13	2828	5422	50	1	0	0	0	0	0	0	0	0	0	0	0	0
14	4394	9887	145	5	1	0	0	0	0	0	0	0	0	0	0	0
15	5344	14433	356	15	2	0	0	0	0	0	0	0	0	0	0	0
16	5089	16865	739	42	7	1	0	0	0	0	0	0	0	0	0	0
17	3793	15776	1293	99	18	2	0	0	0	0	0	0	0	0	0	0
18	2213	11814	1911	203	43	5	1	0	0	0	0	0	0	0	0	0
19	1011	7081	2382	364	89	13	2	1	0	0	0	0	0	0	0	0
20	361	3398	2507	567	161	28	4	2	1	0	0	0	0	0	0	0
21	101	1305	2226	769	258	53	10	4	1	1	1	0	0	0	0	0
22	22	401	1668	908	366	89	19	8	3	2	1	0	0	0	0	0

```
23     4     99 1055 934 457 136   33 16   6   3    3   1   0   0   0   0
24     1     19  563 836 505 186   53 27  11   7    6   1   0   0   0   0
25     0      3  254 651 493 227   77 42  19  12   12   3   1   0   0   0
26     0      0   96 442 425 248  102 60  30  20   21   5   2   1   0   0
27     0      0   31 261 323 242  121 77  43  30   35   9   4   1   0   0
28     0      0    8 134 217 212  131 91  56  42   54  14   6   2   1   0
29     0      0    2  60 129 166  128 98  68  55   78  21  10   3   1   0
30     0      0    0  23  67 117  114 96  75  65  103  29  14   4   1   0
31     0      0    0   8  31  73   92 86  77  71  126  37  20   6   2   0
32     0      0    0   2  13  41   67 70  72  72  143  43  25   7   3   1
33     0      0    0   1   5  21   45 52  61  67  151  48  30   9   3   1
34     0      0    0   0   1   9   27 35  48  57  148  49  34  10   4   1
35     0      0    0   0   0   4   15 22  35  45  134  46  35  11   5   1
36     0      0    0   0   0   1    7 12  23  33  113  41  34  11   5   2
37     0      0    0   0   0   0    3  6  14  22   88  34  31  10   5   2
38     0      0    0   0   0   0    1  3   8  13   63  26  26   9   5   2
39     0      0    0   0   0   0    0  1   4   8   42  18  21   7   4   1
40     0      0    0   0   0   0    0  0   2   4   26  12  15   6   3   1
41     0      0    0   0   0   0    0  0   1   2   15   7  10   4   2   1

> hom$F1992
```

7	8	9	10	11	12	13	14	15	16	17
2	18	95	409	1402	3818	8301	14432	20150	22743	20981

18	19	20	21	22	23	24	25	26	27	28
16190	10943	7029	4729	3487	2747	2215	1794	1452	1177	968

29	30	31	32	33	34	35	36	37	38	39
819	708	629	559	494	423	353	282	215	156	106

40	41
69	42

A total of 1000 length-stratified replicates were drawn from the data simulated using the 1992 length distribution, with a fixed sample size of 10 individuals per length class. If there were not enough individuals on a given length class, then all individuals were sampled. Each of these samples mimics the process of sampling a number of fish from each length class and determining their age by otolith reading. These 1000 samples are also a part of the hom data set, as a list named otoliths.

```
> hom$otoliths[[1]]

    0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
7   1 1 0 0 0 0 0 0 0 0  0  0  0  0  0  0
8   5 5 0 0 0 0 0 0 0 0  0  0  0  0  0  0
9   5 5 0 0 0 0 0 0 0 0  0  0  0  0  0  0
10  5 5 0 0 0 0 0 0 0 0  0  0  0  0  0  0
11  6 4 0 0 0 0 0 0 0 0  0  0  0  0  0  0
12  5 5 0 0 0 0 0 0 0 0  0  0  0  0  0  0
```

```
13 4 6 0 0 0 0 0 0 0 0  0  0  0  0  0  0
14 3 7 0 0 0 0 0 0 0 0  0  0  0  0  0  0
15 5 5 0 0 0 0 0 0 0 0  0  0  0  0  0  0
16 2 8 0 0 0 0 0 0 0 0  0  0  0  0  0  0
17 3 6 1 0 0 0 0 0 0 0  0  0  0  0  0  0
18 0 9 1 0 0 0 0 0 0 0  0  0  0  0  0  0
19 1 4 4 1 0 0 0 0 0 0  0  0  0  0  0  0
20 0 5 5 0 0 0 0 0 0 0  0  0  0  0  0  0
21 0 2 5 2 1 0 0 0 0 0  0  0  0  0  0  0
22 0 2 5 2 1 0 0 0 0 0  0  0  0  0  0  0
23 0 0 3 3 2 0 2 0 0 0  0  0  0  0  0  0
24 0 0 0 4 3 2 0 1 0 0  0  0  0  0  0  0
25 0 0 2 4 3 1 0 0 0 0  0  0  0  0  0  0
26 0 0 0 3 4 2 0 0 0 1  0  0  0  0  0  0
27 0 0 0 4 4 1 0 0 1 0  0  0  0  0  0  0
28 0 0 0 1 1 5 1 2 0 0  0  0  0  0  0  0
29 0 0 0 1 1 2 0 1 0 2  3  0  0  0  0  0
30 0 0 0 0 1 1 2 1 1 2  1  0  1  0  0  0
31 0 0 0 0 0 0 1 0 4 2  2  1  0  0  0  0
32 0 0 0 0 0 0 1 5 1 1  0  1  1  0  0  0
33 0 0 0 0 0 1 0 1 3 2  2  0  1  0  0  0
34 0 0 0 0 0 0 1 0 3 0  4  2  0  0  0  0
35 0 0 0 0 0 0 0 1 2 1  4  2  0  0  0  0
36 0 0 0 0 0 0 0 0 0 0  4  2  2  1  1  0
37 0 0 0 0 0 0 0 0 2 0  4  3  0  1  0  0
38 0 0 0 0 0 0 0 0 1 0  6  0  0  1  2  0
39 0 0 0 0 0 0 0 0 0 1  4  1  4  0  0  0
40 0 0 0 0 0 0 0 0 2 1  3  2  1  1  0  0
41 0 0 0 0 0 0 0 0 0 0  3  1  3  1  1  1
```

ALK generation

The classic age-length key was applied to each bootstrap replicate and to the 1992 length distribution, while each of the other methods was applied using the 1992 length distribution and each bootstrap replicate as the data set with age data, and the 1993 length distribution as the data set without age data. The proportion of each age in the catch (p_j), mean length at age ($\bar{l}_j$) and the standard deviation of length at age (σ_j) were calculated for each resulting ALK, resulting in a total of 48 parameters per method (3 parameters $\times$ 16 age-classes).

```
> attach(hom)
> results <- list(
+   cALK = lapply(otoliths,
+                 function(x) summary(classic_ALK(x, F1992))),
+   invALK = lapply(otoliths,
+                   function(x) summary(inverse_ALK(x, F1992, F1993))),
+   kc = lapply(otoliths,
```

```
+                 function(x) summary(kimura_chikuni(x, F1992, F1993))),
+   hh = lapply(otoliths,
+                 function(x) summary(hoenig_heisey(x, F1992, F1993))),
+   hoenig = lapply(otoliths,
+                    function(x) summary(hoenig(list(x), list(F1992),
+                                                 list(F1993))[[1]])),
+   gascuel = lapply(otoliths,
+                     function(x) summary(gascuel(x, F1992, F1993,
+                                          initial_values = c(0.1, 0.07, 0.06)))))
> detach(hom)
```

The relative performance of each method was assessed by calculating the bias and
variance of the estimates of these parameters. The difference between the estimate of
each parameter and its corresponding value on the original data was calculated for each
sample, and from these values the mean-squared error of the parameter.

```
> MSE <- list(
+   mean_lj = sapply(results, function(x) apply(sapply(x,
+                 function(y) (y$mean_lj - hom$lmed)^2), 1,
+                                               mean, na.rm = TRUE)),
+   var_lj = sapply(results, function(x) apply(sapply(x,
+                 function(y) (y$var_lj - hom$stdv)^2), 1,
+                                               mean, na.rm = TRUE)),
+   pj = sapply(results[-1], function(x) apply(sapply(x,
+                 function(y) (y$pj - hom$pj_1993)^2), 1,
+                                               mean, na.rm = TRUE)))
> MSE[["pj"]] <- cbind(MSE[["pj"]],
+                     cALK = apply(sapply(results[["cALK"]],
+                                  function(y) (y$pj - hom$pj_1992)^2),
+                                  1, mean, na.rm = TRUE))
```

Age slicing

Age slicing is a commonly used method of estimating age distributions from length
distributions based on a given growth model (see Introduction), often von Bertalanffy's
growth curve. The performance of this approach was also compared with the ALK
models implemented by the ALKr package, namely its estimates of the proportion of
age in the population (p_j).

Von Bertalanffy's growth model defines length l as a function of age a, an assymp-
totic maximum length L_∞, a growth rate K and a theoretical length at age 0, t_0:

$$l = L_\infty \times (1 - \exp(-K \times (a - t_0)))$$

.

As these parameters were unavailable for the horse mackerel population being studied, the length/age data was used to estimate them by non-linear regression.

```
> library(reshape)
> mN1992 <- melt(hom$N1992)
> mN1992 <- data.frame(l = rep(mN1992[, 1], mN1992[, 3]),
+                      a = rep(mN1992[, 2], mN1992[, 3]))
> vb_params <- nls(l ~ Linf * (1 - exp(-K * (a - t0))),
+                  data = mN1992,
+                  start = list(Linf = 42, K = 0.18, t0 = -0.5))$m$getPars()
> vb_params

       Linf          K         t0
54.98434344 0.06403416 -4.68226399
```

Figure 1 shows the resulting growth curve against the length-at-age distributions.

```
> library(ggplot2)
> vb_curve <- function(a, vb) {
+               vb["Linf"] * (1 - exp(-vb["K"] * (a - vb["t0"])))
+             }
> vb_plot <- ggplot(mN1992) +
+   theme_bw() +
+   geom_violin(aes(x = a, y = l, group = a), adjust = 2.5) +
+   stat_function(fun = vb_curve, args = list(vb = vb_params),
+                 colour = "red", geom = "line") +
+   labs(x = "Age (years)", y = "Length (cm)")
```

Age slicing was then applied to estimate the proportion of each age class in the population.

```
> age_slicing <- function(fi, li = as.numeric(names(fi)), vb,
+                          age_limits, timing = 0.5) {
+
+   age <- floor(vb["t0"] - log(1 - pmin(li / vb["Linf"], 1 - 1e-16)) /
+                  vb["K"] + timing)
+   age <- pmax(pmin(age, age_limits[2]), age_limits[1])
+   age <- factor(age, levels = age_limits[1]:age_limits[2])
+
+   tapply(fi, age, sum) / sum(fi)
+ }
> pj_age <- age_slicing(hom$F1992,
+                       vb = vb_params,
+                       age_limits = c(0, 15))
> pj_age

          0           1           2           3           4
0.324316213 0.291615812 0.227842360 0.054796348 0.033093899
```

Figure 1: Length-at-age distributions for the 1992 horse mackerel population and the estimated von Bertalanffy's growth curve
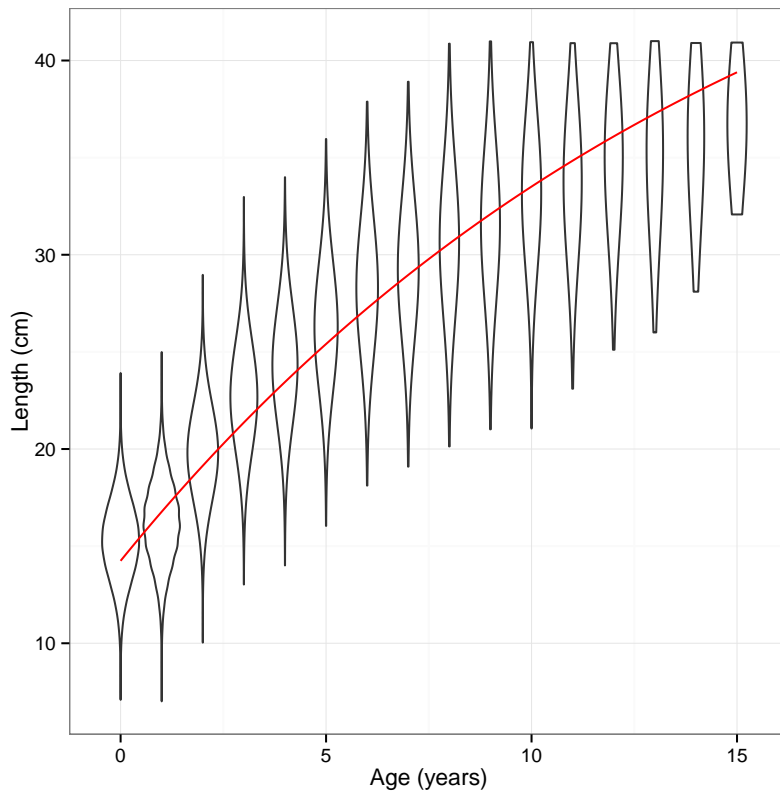
```
         5              6              7              8              9
0.021649093  0.014306009  0.005462294  0.008917079  0.003728233
        10             11             12             13             14
0.006115902  0.002354322  0.001880790  0.001433936  0.001040437
        15
0.001447275
```

The squared error of these estimates was then appended to the previous results.

```
> MSE[["pj"]] <- cbind(MSE[["pj"]], age_slicing = (pj_age - hom$pj_1992)^2)
```

Results

Figure 2 and tables 1 to 3 show the mean squared error (MSE) for the mean length-at-age, variance of length-at-age and proportion of age in the population for each of the ALK methods evaluated, as well as the squared error of the proportion of age in the population obtained by age slicing.

```
> mMSE <- melt(MSE)
> names(mMSE) <- c("Age", "Method", "MSE", "Variable")
> mMSE$Method <- factor(mMSE$Method,
+   levels = c("cALK", "invALK", "kc", "hh",
+             "hoenig", "gascuel", "age_slicing"),
+   labels = c("Classic ALK", "Inverse ALK", "K & C", "H & H",
+             "Hoenig et al.","Gascuel", "Age slicing"))
> mMSE$Variable <- factor(mMSE$Variable,
+   levels = c("mean_lj", "var_lj", "pj"),
+   labels = c("Mean length-at-age", "Variance of length-at-age",
+             "Proportion of age"))
> mse_plot <- ggplot(mMSE) +
+   theme_bw() +
+   geom_line(aes(x = Age, y = MSE, colour = Method)) +
+   facet_wrap(~Variable, ncol = 1, scale = "free_y") +
+   scale_y_log10() +
+   scale_colour_manual(values = c("#F92713", "#5A7A83", "#6CA11D", "#BD1C5B",
+                                  "#A7D8C8", "#E0AA08", "#295A00"))
```

The results clearly illustrate the problems of the Inverse ALK method, which shows very large MSE of its estimates of mean length-at-age ($\bar{l}_j$) and variance of length-at-age (σ_j), and also has the worst proportion of age in the population (p_j) estimates. All other methods present similar MSE of their $\bar{l}_j$ estimates, although the Hoenig & Heisey method shows a slightly worse performance on the higher age classes.

With respect to p_j estimates, the Classic ALK shows good performance on lower age classes, but average results on higher classes, where both Kimura & Chikuni and

17

Figure 2: Mean squared errors (MSE) of the mean length-at-age, variance of length-at-age and proportion of age in the population for each method
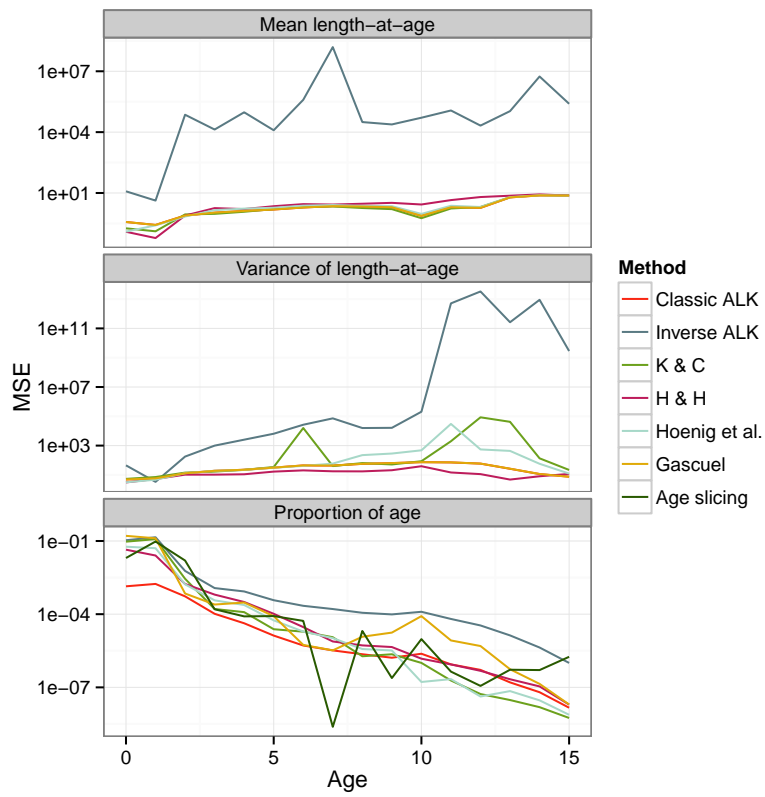
Hoenig *et al.* methods perform best. Gascuel's method shows relatively large MSE on the upper half of the age range.

Conclusions

The classic ALK (Friðriksson, 1934) is a good method for estimating age distributions if populations (in the biological sense) are well sampled and if there is some age/length data already avaliable. It produces maximum likelihood estimates and is a computationally simple algorithm. However, to apply it to populations for which no age/length data is known, one must make strong assumptions, that are unlikely to be met.

The inverse ALK method (Clark, 1981; Bartoo and Parker, 1983; Hilborn and Walters, 1992) performs poorly mostly becaus of its usage of a generalized inverse matrix whose computation may be problematic, but all methods based on it perform similarly. However, Hoenig *et al.*'s method (Hoenig et al., 1993, 1994, 2002) shows good results on the highest age classes (in which length distributions on the original data have large overlaps), also yields maximum likelihood estimates and is the only method that can combine several complementary data sets, either with or without age information.

When compared to the ALK methods, age slicing performed surprisingly well. However, it must be noted that the analysis was done with simulated data, instead of real-world samples. Moreover, the input data was generated assuming that the mean lengths-at-age had normal distributions and relatively low variances, which are both desirable characteristics for age slicing.

Acknowledgments

Table 1: Mean squared error of mean length-at-age estimates for each method, per age class

	Classic ALK	Inverse ALK	Kimura & Chikuni	Hoenig & Heisey	Hoenig et al.	Gascuel
0	0.37	12.09	0.18	0.12	0.12	0.37
1	0.27	4.23	0.13	0.06	0.25	0.27
2	0.77	71972.19	0.88	0.77	0.69	0.77
3	1.07	13422.56	0.94	1.79	1.43	1.07
4	1.30	95071.07	1.19	1.61	1.70	1.30
5	1.53	12372.09	1.61	2.23	1.87	1.53
6	1.93	390957.50	2.02	2.82	2.37	1.93
7	2.24	155360440.90	2.17	2.69	2.54	2.24
8	2.15	31281.91	1.82	2.95	2.31	2.15
9	1.99	23940.90	1.60	3.26	2.24	1.99
10	0.75	51278.42	0.57	2.71	0.93	0.75
11	1.98	116805.32	1.72	4.46	2.34	1.98
12	1.85	20966.49	2.05	6.32	2.04	1.85
13	6.07	108315.11	6.20	7.34	6.25	6.07
14	7.60	5514873.64	7.88	8.51	7.73	7.60
15	7.43	249800.49	7.50	7.43	7.34	7.43

Table 2: Mean squared error of variance of length-at-age estimates for each method, per age class

	Classic ALK	Inverse ALK	Kimura & Chikuni	Hoenig & Heisey	Hoenig et al.	Gascuel
0	4.88	44.80	5.35	3.07	3.12	4.89
1	5.78	3.35	7.15	5.26	4.71	5.78
2	12.90	178.33	14.30	10.38	13.10	12.90
3	19.00	994.25	16.83	10.40	19.18	19.00
4	22.40	2537.25	22.16	11.01	21.74	22.42
5	31.34	6453.16	34.43	16.81	30.97	31.38
6	44.23	26521.84	15822.81	20.32	41.10	44.33
7	43.56	72398.71	39.98	17.63	59.09	43.70
8	56.67	15832.66	64.08	17.43	224.08	56.89
9	63.79	16340.26	51.66	20.83	287.50	64.04
10	74.12	205129.36	85.90	39.18	477.69	74.26
11	71.17	51206030812779.57	1984.26	14.64	30228.91	71.56
12	58.84	33221578090057.21	85307.65	11.31	548.24	59.28
13	26.16	264126953163.96	41966.58	4.80	434.78	26.59
14	11.41	8960313118190.76	135.31	8.07	57.23	11.53
15	7.56	2851269173.73	21.73	11.63	12.30	7.44

Table 3: Mean squared error of the propotion of age in the population estimates for each method, per age class

	Classic ALK	Inverse ALK	Kimura & Chikuni	Hoenig & Heisey	Hoenig et al.	Gascuel	Age slicing
0	1.07e-01	9.40e-02	4.40e-02	5.83e-02	1.63e-01	1.39e-03	2.01e-02
1	1.42e-01	1.19e-01	2.54e-02	5.06e-02	1.29e-01	1.74e-03	9.52e-02
2	5.93e-03	2.83e-03	1.73e-03	1.67e-03	7.03e-04	5.37e-04	1.58e-02
3	1.18e-03	1.63e-04	6.35e-04	3.69e-04	2.49e-04	1.04e-04	1.59e-04
4	8.55e-04	1.21e-04	3.16e-04	2.38e-04	3.00e-04	4.24e-05	8.14e-05
5	3.74e-04	2.42e-05	1.04e-04	5.58e-05	8.32e-05	1.33e-05	8.37e-05
6	2.20e-04	1.92e-05	2.96e-05	1.94e-05	5.50e-06	5.21e-06	5.30e-05
7	1.62e-04	1.14e-05	7.61e-06	1.04e-05	3.28e-06	3.27e-06	2.42e-09
8	1.14e-04	1.88e-06	5.26e-06	3.73e-06	1.19e-05	2.28e-06	2.05e-05
9	9.79e-05	2.27e-06	4.50e-06	3.25e-06	1.76e-05	1.64e-06	2.41e-07
10	1.25e-04	1.01e-06	1.50e-06	1.66e-07	8.33e-05	2.39e-06	9.48e-06
11	6.37e-05	1.91e-07	8.70e-07	2.18e-07	8.38e-06	8.75e-07	4.42e-07
12	3.45e-05	5.32e-08	4.76e-07	4.21e-08	4.95e-06	5.20e-07	1.15e-07
13	1.35e-05	3.03e-08	2.15e-07	7.03e-08	5.53e-07	1.60e-07	5.29e-07
14	4.30e-06	1.54e-08	1.09e-07	2.93e-08	1.39e-07	6.33e-08	5.12e-07
15	1.00e-06	5.56e-09	1.96e-08	7.51e-09	1.98e-08	1.46e-08	1.79e-06

References

N. W. Bartoo and K. R. Parker. Stochastic age-frequency estimation using the von bertalanffy growth equation. *Fishery Bulletin*, 81(1):91--96, 1983.

C. G. Bhattacharya. A simple method of resolution of a distribution into gaussian components. *Biometrics*, 23(1):115, March 1967. ISSN 0006341X. doi: 10.2307/2528285.

W. G. Clark. Restricted least-squares estimates of age composition from length composition. *Canadian Journal of Fisheries and Aquatic Sciences*, 38(3):297--307, March 1981. ISSN 0706-652X, 1205-7533. doi: 10.1139/f81-041.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1--38, 1977. ISSN 00359246. doi: 10.2307/2984875. ArticleType: research-article / Full publication date: 1977 / Copyright © 1977 Royal Statistical Society.

A. J. Fabens. Properties and fitting of the von bertalanffy growth curve. *Growth*, 29 (3):265--289, September 1965. ISSN 0017-4793. PMID: 5865688.

D. A. Fournier, John R. Sibert, Jacek Majkowski, and John Hampton. MULTIFAN a likelihood-based method for estimating growth parameters and age composition from multiple length frequency data sets illustrated using data for southern bluefin tuna (Thunnus maccoyii). *Canadian Journal of Fisheries and Aquatic Sciences*, 47 (2):301--317, February 1990. ISSN 0706-652X, 1205-7533. doi: 10.1139/f90-032.

Á. Friðriksson. On the calculation of age-distribution within a stock of cod by means of relatively few age determinations as a key to measurements on a large scale. *Rapp. P.-V. CIEM*, 86:1--5, 1934.

Didier Gascuel. Une méthode simple d'ajustement des clés taille/âge : application aux captures d'albacores (thunnus albacares) de l'Atlantique est. *Canadian Journal of Fisheries and Aquatic Sciences*, 51(3):723--733, March 1994. ISSN 0706-652X, 1205-7533. doi: 10.1139/f94-072.

John Hampton. Estimation of southern bluefin tuna thunnus maccoyii growth parameters from tagging data, using von bertalanffy models incorporating individual variation. *Fishery Bulletin*, 89(4):577--590, 1991.

Victor Hasselblad. Estimation of parameters for a mixture of normal distributions. *Technometrics*, 8(3):431--444, August 1966. ISSN 0040-1706, 1537-2723. doi: 10.1080/00401706.1966.10490375.

R. Hilborn and C. J. Walters. Quantitative fisheries stock assessment: Choice, dynamics and uncertainty. *Reviews in Fish Biology and Fisheries*, 2(2):177--178, June 1992. ISSN 0960-3166. doi: 10.1007/BF00042883.

John M. Hoenig, R. Choudary Hanumara, and Dennis M. Heisey. Generalizing double and triple sampling for repeated surveys and partial verification. *Biometrical Journal*, 44(5):603--618, 2002. ISSN 1521-4036. doi: 10.1002/1521-4036(200207)44:5<603::AID-BIMJ603>3.0.CO;2-4.

John Maurice Hoenig and Dennis M. Heisey. Use of a log-linear model with the EM algorithm to correct estimates of stock composition and to convert length to age. *Transactions of the American Fisheries Society*, 116(2):232--243, March 1987. ISSN 0002-8487, 1548-8659. doi: 10.1577/1548-8659(1987)116<232:UOALMW>2.0.CO;2.

John Maurice Hoenig, Dennis M. Heisey, and R. C. Hanumara. Using prior and current information to estimate age composition: a new kind of age-length key. In *ICES CM Documents 1993*, ICES Conference and Meeting Documents, page 10. International Council for the Exploration of the Sea, 1993.

John Maurice Hoenig, Dennis M. Heisey, and R. C. Hanumara. A computationally simple approach to using current and past data in age-length key. In *ICES CM Documents 1994*, ICES Conference and Meeting Documents, page 5, St. John's, ND, Canada, 1994. International Council for the Exploration of the Sea.

Daniel K. Kimura. Statistical assessment of the Age–Length key. *Journal of the Fisheries Research Board of Canada*, 34(3):317--324, March 1977. ISSN 0015-296X. doi: 10.1139/f77-052.

Daniel K. Kimura and Shiro Chikuni. Mixtures of empirical distributions: An iterative application of the age- length key. *Biometrics*, 43(1):23, March 1987. ISSN 0006341X. doi: 10.2307/2531945.

G. P. Kirkwood and I. F. Somers. Growth of two species of tiger prawn, penaeus esculentus and p. semisulcatus, in the western gulf of carpentaria. *Marine and Freshwater Research*, 35(6):703, 1984. ISSN 1323-1650. doi: 10.1071/MF9840703.

William Knight. Asymptotic growth: An example of nonsense disguised as mathematics. *Journal of the Fisheries Research Board of Canada*, 25(6):1303--1307, June 1968. ISSN 0015-296X. doi: 10.1139/f68-114.

S. J. Leon. *Linear algebra with applications*. Prentice Hall, 5th edition, 1994.

P. D. M. Macdonald. *The age and growth of fish*, chapter Analysis of length-frequency distributions, pages 371--383. The Iowa State University Press, 1987.

I. Martin and R. M. Cook. Combined analysis of length and age-at-length data. *Journal du Conseil: ICES Journal of Marine Science*, 46(2):178--186, 1990. doi: 10.1093/icesjms/46.2.178.

T. J. Mulligan and B. M. Leaman. Length-at-age analysis: Can you get what you see? *Canadian Journal of Fisheries and Aquatic Sciences*, 49(4):632--643, April 1992. ISSN 0706-652X, 1205-7533. doi: 10.1139/f92-072.

Alberto Murta. Methods to estimate the age structure of a fish population based on the determination of ages. Master's thesis, Instituto Superior de Agronomia da Universidade Técnica de Lisboa, 1998.

J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308--313, 1965. ISSN 0010-4620, 1460-2067. doi: 10.1093/comjnl/7.4.308.

W. E. Ricker. *Computation and interpretation of biological statistics of fish populations*. Number 191 in Bulletin. Fisheries Research Board of Canada, 1975.

Derek A. Roff. A motion for the retirement of the von bertalanffy function. *Canadian Journal of Fisheries and Aquatic Sciences*, 37(1):127--129, 1980. doi: 10.1139/f80-016.

K. J. Sainsbury. Effect of individual variability on the von bertalanffy growth equation. *Canadian Journal of Fisheries and Aquatic Sciences*, 37(2):241--247, February 1980. ISSN 0706-652X, 1205-7533. doi: 10.1139/f80-031.

J. L. Schafer. *Analysis of incomplete multivariate data*. Number 72 in Monographs on statistics and applied probability. Chapman & Hall, 1997.

Jon Schnute. A versatile growth model with statistically stable parameters. *Canadian Journal of Fisheries and Aquatic Sciences*, 38(9):1128--1140, September 1981. ISSN 0706-652X, 1205-7533. doi: 10.1139/f81-153.

Jon Schnute and David Fournier. A new approach to Length–Frequency analysis: Growth structure. *Canadian Journal of Fisheries and Aquatic Sciences*, 37(9):1337--1351, September 1980. ISSN 0706-652X, 1205-7533. doi: 10.1139/f80-172.

L. von Bertalanffy. A quantitative theory of organic growth (inquiries on growth laws, II). *Human Biology, a record of research*, 10(2):181--213, 1938.

Yoy-Gan Wang and Mervyn R. Thomas. Accounting for individual variability in the von bertalanffy growth model. *Canadian Journal of Fisheries and Aquatic Sciences*, 52(7):1368--1375, July 1995. ISSN 0706-652X, 1205-7533. doi: 10.1139/f95-132.

S. J. Westrheim and W. E. Ricker. Bias in using an Age–Length key to estimate age-frequency distributions. *Journal of the Fisheries Research Board of Canada*, 35(2):184--189, February 1978. ISSN 0015-296X. doi: 10.1139/f78-030.

Yongshun Xiao. Estimation of southern bluefin tuna thunnus maccoyii growth parameters from tagging data, using von bertalanffy models incorporating individual variation. *Canadian Journal of Fisheries and Aquatic Sciences*, 53(7):1675--1682, 1996. doi: 10.1139/f96-074.