



## **ESTIMATION OF AGE PROFILES OF SOUTHERN BLUEFIN TUNA**

---

**Richard Morton\***  
**Mark Bravington\***

**\*CSIRO Maths and Information Science**

**CCSBT-ESC/0309/32**

**Table of Contents**

ABSTRACT ..... 1

1 Introduction ..... 1

2 A likelihood framework for length and age-at-length data ..... 2

    2.1 Estimating proportions-at-age ..... 4

        2.1.1 Age-length key ..... 4

        2.1.2 Iterated age-length key ..... 5

        2.1.3 Parametric estimator: known growth..... 5

        2.1.4 Parametric estimator: unknown growth..... 6

    2.2 Example: application to Greenland turbot data ..... 6

3 Application to sampling design for sbt..... 7

    3.1 Results by method ..... 8

    3.2 Results by fishery ..... 9

        3.2.2 Effect of varying the subsampling pattern..... 10

4 Conclusions and DISCUSSION ..... 10

    4.1 Discussion points ..... 12

        4.1.1 Within-season growth ..... 12

        4.1.2 How should precision vary with age?..... 12

        4.1.3 Effective sample size for the length frequency data ..... 13

        4.1.4 Errors in age and length measurements..... 14

        4.1.5 Further methodological work..... 14

    4.2 Summary ..... 15

5 REFERENCES..... 16

6 APPENDIX: THEORETICAL DETAILS ..... 17

    Age-Length Key ..... 17

    Iterated Age-Length Key ..... 18

    Random sampling ..... 18

    6.1 Parametric Methods and Mixture Distributions..... 18

        6.1.1 Estimation ..... 18

        6.1.2 No length data..... 20

    6.2 Information for the estimates..... 20

        6.2.1 Two age classes, length-at-age distributions known ..... 21

        6.2.2  $K$  age classes, length-at-age parameters to be estimated..... 21

        6.2.3 ALK ..... 24

    6.3 Likelihood Under Subsampling ..... 24

7 TABLES ..... 26

**List of Tables**

Table 1. Comparative estimates of proportions-at-age in 1983 for Greenland turbot data ..... 26

Table 2. Precision results for uniform-intensity sampling:  $n$  otoliths and  $N$  length samples. "p[age]" is proportion at age. "1" and "2" refer to the parametric known-growth and unknown-growth methods respectively. "CV" is the coefficient of variation. "neq" is the size of a randomly-selected sample (random with respect to age and length) that would deliver equal precision to a particular estimator. The ARE columns show the efficiency of the ALK relative to each parametric method. .... 27

Table 3. Summary of CVs in Table 2 for the parametric unknown-growth estimator and the Japanese fishery ..... 38

Table 4. Precision results for unbalanced subsampling (relatively more big fish): Japanese long-lines. Column headings are as for Table 2. .... 39

## ABSTRACT

This paper discusses various possible sampling and statistical analysis schemes for estimating proportions-at-age of SBT in catches, based on the collection of otoliths and length frequency samples. We introduce two hybrid estimators that efficiently use the length sample as well as the age subsamples to estimate proportions-at-age. The new methods are shown to give more precise estimates than age-length keys for SBT. The gain in precision is large for the GAB fishery, intermediate for the Japanese long-line fishery, and minimal for the Indonesian fishery where fish length carries almost no information about likely age. We consider what gains in efficiency might be achievable using non-proportional subsampling of length groups. Although subsampling design and size of length frequency sample have some impact on precision, the main determinant is number of otoliths sampled. Tables are included showing what CVs could be obtained at different sample sizes of otoliths and length frequencies.

## 1 INTRODUCTION

This paper discusses various possible sampling and statistical analysis schemes for estimating proportions-at-age of SBT in catches, based on the collection of otoliths and length frequency samples.

The usual way of estimating proportions-at-age in a given year, using age-at-length samples and a length distribution sample in that same year, is via an age-length key (ALK). The length distribution is only used to weight the sample distributions of age-at-length; no use is made of the information content of the length distribution itself. This is potentially inefficient because—at least for a small fish—fish length alone does provide some information on fish age, given that we already know something about growth. For large fish, the information provided about age is very imprecise, so it is impossible to manage without at least some otolith sampling. However, otolith sampling is expensive, especially compared to the collection of length frequency data. In designing a sampling scheme, it is therefore important to assess how much extra precision can be obtained by efficient use of length samples as well as age samples.

The paper is structured as follows. In section 2, we develop a common framework for estimating age proportions, showing how the ALK and iterated age-length key (IALK) fit in. We then develop a hybrid maximum-likelihood estimator, which combines (i) the prior information on age provided by fish length distributions alone, and (ii) the data on age-at-length distributions from otolith sub-samples. As an example, we show how the hybrid methods can be applied to published data on Greenland turbot. Section 3 shows what CVs on proportions-at-age could be achieved under different sampling schemes, in three major SBT fisheries with very different age distributions: purse seiners in the GAB, Japanese long-liners, and Indonesian long-liners. We show how CV would vary for different sample sizes of (i) length distribution, and (ii) age-at-length distributions. We compare the precision of three different estimators: (a) the hybrid estimator with known growth parameters, (b) the hybrid estimator where growth parameters need to be estimated from the same data, and (c) a conventional ALK. We also show some simple examples of how precision

would be affected by different length-subsampling protocols within a fishery (i.e. taking otoliths from relatively more small fish, or from relatively more big fish). Finally, section 4 contains the conclusions and discussion.

It is important to note that we are not trying to fully specify a sampling and analysis scheme in this paper. Rather, the aim is to establish broadly what precision could be expected from various levels of sampling coverage in the main SBT fisheries. Appropriate target levels for precision depend on the assessment and management process, and are not explicitly specified as yet. We have not, therefore, tried to devise an "optimal" sampling design tuned to achieve some pre-specified precision at minimum cost. However, based on the achievable CVs in section 3, it should not be difficult to evaluate the general implications of different designs for assessment and management; for example, Pope (CCSBT-CPUE/0203/10) shows how "management precision" can be estimated for a simple "management rule" (a status quo TAC in his example).

We have deliberately, but only temporarily, overlooked several details: in particular, within-season growth, errors in age or length measurements, and estimation of overdispersion in the length distribution (e.g. when fish from the same set tend to be of similar size). Although it would be necessary to give these at least some consideration when faced with real data, they are essentially technical details which could be addressed in future methodological work. We do not expect the overlooked details to substantially affect the precision that could be achieved (except perhaps in the case of measurement error). Further discussion may be found in section 4.

## 2 A LIKELIHOOD FRAMEWORK FOR LENGTH AND AGE-AT-LENGTH DATA

Let  $a$  be the measured age of a fish, and  $s$  be the size (i.e. the length; the terms "size" and "length" are interchangeable here). Suppose we have  $n$  pairs of measurements  $(a_i, s_i)$  from the same fishery, condensed into frequencies  $n_{as}$  giving the number of fish aged  $a$  and of size  $s$ . We assume that these data have been obtained through random subsampling within size classes, so that the set  $\{n_{as} : a_{\min} \leq a \leq a_{\max}\}$  is distributed multinomially with fixed total  $n_s = \sum_a n_{as}$  and probabilities of age-at-length  $p_{a|s}$  (so that  $\sum_a p_{a|s} = 1$  for all  $s$ ). The form of  $p_{a|s}$  is discussed below. The log-likelihood  $\Lambda_{AS}$  from the age subsamples is therefore

$$\Lambda_{AS} = \sum_s \left( \sum_a n_{as} \log p_{a|s} \right) \quad (1)$$

We also have available  $N$  measurements of size  $s_j$ , condensed into frequencies  $N_s$  giving the number of fish of size  $s$ . We assume that each of these  $N$  fish is an independent random sample from the true length-frequency distribution. The set  $\{N_s : s_{\min} \leq s \leq s_{\max}\}$  is therefore distributed multinomially with total  $N$  and probabilities of length  $p_s$  (the form of which is discussed below). The log-likelihood  $\Lambda_S$  from the length samples is therefore

$$\Lambda_S = \sum_s N_s \log p_s \quad (2)$$

The joint log-likelihood is  $\Lambda_S + \Lambda_{AS}$ . It is shown in the Appendix that this is valid regardless of whether the  $n$  age subsamples are, or are not, taken from the same actual fish used in the  $N$  length samples (provided that all the fish do come from the same underlying population).

The goal of analysis is to estimate the proportions at age  $p_a$ . These do not appear directly in the log-likelihoods, but do affect the distribution of the data through their influence on  $p_{a|s}$  and  $p_s$ , as explained below.

Section 2.1 discusses four approaches, but first there we make some general points about the above framework.

**Note 1:** Our definition of  $p_a$  is the expected proportion at age in a random sample caught from the population by the fishery. This means that  $p_a$  is a product of two terms: the true proportion-at-age in the population, multiplied by the selectivity-at-age in the particular fishery being sampled. Similar definitions apply to  $p_s$ ,  $p_{a|s}$  and  $p_{s|a}$  (the distribution of length-at-age); they are to be interpreted as applying to a "selectivity-weighted population". This is entirely appropriate for estimating age composition in the catch, where selectivity *per se* is irrelevant. We therefore do not explicitly incorporate selectivity parameters in our framework. However, it is important not to forget about selectivity when incorporating prior information pertaining to the whole population (e.g. on size-at-age) into analyses pertaining to one fishery. Selectivity is also crucial when making inferences about the true population age composition, but that depends on details of the stock assessment which are beyond the scope of this paper.

**Note 2:** In practice, length frequency samples are unlikely to be independent; two fish caught in the same set are likely to be more similar in size than two fish caught in different sets or in a different part of the fishery area. The above formulation is nevertheless reasonable if  $N$  is regarded as the "effective independent sample size", i.e. the size of a sample of truly independent length measurements that would provide the same information about the length frequency in the population (i.e. in the selectivity-weighted population). Depending on the nature of the fishery and the sampling design,  $N$  may be smaller than the number of fish measured. This can be true even if the *entire* catch is measured for length, and even when we are mainly concerned with estimating actual numbers-at-age in the catch rather than proportions-at-age in the population. Practical ways of estimating effective sample size are discussed in section 4.

**Note 3:** We have assumed instantaneous sampling at one moment in time, neglecting any growth within the season. In reality, the quantities  $n_{as}$ ,  $N_s$ ,  $p_{a|s}$  and  $p_s$  (but not  $p_a$ ) all ought to have an extra subscript  $t$  to denote time-of-year. See section 4 for further discussion.

## 2.1 Estimating proportions-at-age

The estimators discussed in this section can all be viewed as maximum likelihood estimators of  $p_a$ , placing different restrictions on  $p_{a|s}$  and/or using only part of the data. The derivations make use of three standard probability formulae:

$$\begin{aligned} p_s &= \sum_a p_a p_{s|a} \\ p_a &= \sum_s p_s p_{a|s} \\ p_{a|s} &= \frac{p_{s|a} p_a}{p_s} \end{aligned} \tag{3}$$

### 2.1.1 Age-length key

In the ALK, the proportions of age-at-length  $p_{a|s}$  are estimated nonparametrically. It is easy to show that the nonparametric MLE is just  $\hat{p}_{a|s} = n_{as} / n_s$ . Proportions-at-length  $p_s$  are also estimated nonparametrically, via  $\hat{p}_s = N_s / N$ . Then the MLE of  $p_a$  is given by the above formulae as

$$\hat{p}_a = \sum_s \hat{p}_s \hat{p}_{a|s} \tag{4}$$

Although the ALK is a full-likelihood method using both  $\Lambda_S$  and  $\Lambda_{AS}$ , the  $\Lambda_S$  part contains no extra information about age composition, so in effect ALK relies only on the age-subsample log-likelihood  $\Lambda_{AS}$ . The underlying reason is that the length-at-age proportions  $p_{s|a}$  are completely free parameters in the model; if  $p_s$  is changed, then  $p_{s|a}$  can be adjusted accordingly so as to keep  $p_{a|s}$  the same. Consequently the fit to the length frequency can be adjusted (via  $p_s$ ) without affecting the fit to the age subsample at all, and so  $\Lambda_S$  carries no information about age.

Variance estimates for the ALK can be obtained very easily because the estimator has a closed form. In most fishery applications, estimates of  $\text{cov}[\hat{p}_a, \hat{p}_{a'}]$  for the ALK treat  $\hat{p}_s$  as known exactly, without error. However, if the effective sample size  $N$  is very small, then uncertainty about the true length composition adds an extra source of variation. The modification to the variance estimates is straightforward; see the Appendix for further details.

Note that the above formulae imply a corresponding nonparametric estimate of  $p_{s|a}$ , namely

$$\hat{p}_{s|a} = \frac{\hat{p}_{a|s} \hat{p}_s}{\hat{p}_a} \tag{5}$$

This is not directly needed in the ALK, but is usually a key ingredient of the iterated age-length key, described next.

### 2.1.2 Iterated age-length key

This method was introduced by Kimura and Chikuni (1987) as a way of using pre-existing estimates of length-at-age distribution, together with length samples but no age samples from the current year, to make inferences about age distribution in the current year. It does not use age subsamples from the current year, and therefore only uses the  $\Lambda_S$  part of the log-likelihood. In the IALK,  $\Lambda_S$  is maximized over the unknown parameters  $p_a$ , using the formula  $p_s = \sum_a p_a p_{s|a}$  and treating  $p_{s|a}$  is known exactly. In the original specification of the IALK,  $p_{s|a}$  is replaced by an estimate derived from applications of the ALK to previous data, but it would also be possible to use a parametric estimate (see 2.1.3). Because  $p_{s|a}$  describes the distribution of length-at-age, there is some reason to expect that it will be fairly consistent over time (i.e. over cohorts). However, inaccuracies in  $p_{s|a}$  will lead to inaccuracies in the IALK. In fact, the IALK is known to be very sensitive to variations in growth between cohorts.

The name "iterated age-length key" (IALK) arises from Kimura & Chikuni's particular choice of algorithm; other algorithms could be used to give the same MLE. The name is in one sense misleading, because IALK actually extracts its age information from the  $\Lambda_S$  part of the log-likelihood, whereas ALK extracts its age information from the  $\Lambda_{AS}$  part; thus the two methods are more different than the names suggest.

The IALK is not of much direct relevance to SBT, because there will presumably always be some direct age data available in any given year. However, it is instructive in showing "what the ALK is missing": i.e., that it is possible to make some inferences about age distribution based purely on length samples and prior knowledge about growth.

### 2.1.3 Parametric estimator: known growth

We can re-write the combined log-likelihood as

$$\begin{aligned} \Lambda &= \sum_s \left\{ N_s \log\left(\sum_a p_a p_{s|a}\right) + \sum_a n_{as} \log\left(\frac{p_{s|a} p_a}{\sum_{a'} p_{s|a'} p_{a'}}\right) \right\} \\ &= \sum_s \left\{ (N_s - n_s) \log\left(\sum_a p_a p_{s|a}\right) + \sum_a n_{as} \log(p_a p_{s|a}) \right\} \end{aligned} \quad (6)$$

If we have good prior information on length-at-age, we can take  $p_{s|a}$  as known and maximize (6) with respect to  $p_a$ . For SBT, we follow Polacheck *et al.* (2003) in taking a Normal distribution for length-at-age, with known mean and variance at age  $a$  of  $\mu_a$  and  $\sigma_a^2$ . For most ages commonly caught in a particular fishery, it will be acceptable to use estimates of  $\mu_a$  and  $\sigma_a^2$  that apply to the whole population at that age, rather than to the specific fishery. However, for fish near the ends of the length range caught by a fishery, selectivity may affect the sampling distribution of length-at-age compared to the population distribution, e.g. in the Indonesian fishery where the only 8-year-olds to be seen tend to be relatively larger than 8-year-olds elsewhere. This

would change the mean and variance in the exploitable 8-year-olds away from the values for 8-year-olds in the whole population. If selectivity-at-length is known, this can be allowed for. Note that, even if mean and variance of length-at-age for some age classes is affected by selectivity, the adequacy of a Normal approximation per se to length-at-age distribution should not be much affected, assuming the window of selectivity in length is reasonably large compared to the variation in length-at-age within a cohort.

The parametric method uses both parts of the log-likelihood to make inferences about age distribution, and is therefore in some sense a hybrid between ALK and IALK. The ability to extract information from the length frequency alone, comes at the price of having to make a parametric restriction on the form of  $p_{s|a}$ . However, the extra information used makes the parametric method more efficient (i.e. higher precision for the same sample size), provided its parametric assumptions are met.

Details of the estimation algorithm are given in Appendix 1, which also gives expressions for parameter (co)variances.

#### 2.1.4 Parametric estimator: unknown growth

This follows 2.1.3 exactly, except that  $(\mu_a, \sigma_a^2)$  are treated as additional unknown parameters that have to be estimated. This is a rather drastic way of dealing with concerns about possible variations in size-at-age between cohorts, since it ignores all prior information from that cohort in previous years, and from previous cohorts at that age. It is possible to envisage more elaborate hybrids between 2.1.3 and 2.1.4, in which the parameters size-at-age are treated as random effects for which historical growth studies provide a prior distribution. This might, in principle, combine the efficiency of 2.1.3 (when its assumption of consistency across cohorts is valid) with the presumed robustness of 2.1.4. In this paper, though, we have used only the extreme cases, of length-at-age completely known or completely unknown (except for being Normally distributed).

Details of estimation are again given in the Appendix, where the influence of the extra mean and variance parameters on the precision of  $\hat{p}_a$  is also discussed.

## 2.2 Example: application to Greenland turbot data

We first consider the Greenland Turbot data analysed in Kimura & Chikuni (1987). Length-and-age measurements from randomly-selected fish can be used to construct length-at-age distributions directly. The IALK is used to apply these length-at-age distributions to length samples from other years, to estimate age compositions in those years. IALK per se will not be necessary for SBT if some new age samples are available in every year, since it does not use the new age samples. However, it is of interest to see how the parametric methods compare with the IALK method even in the absence of age data.

An example of the possible capriciousness of the IALK method can be seen for age 14 where only one fish was found, of length 94cm. Thus all 94cm fish in subsequent length samples would be assigned to age 14. Intuitively, it seems likely that this could seriously increase the standard errors associated with the  $p_a$ . Thus, it could be better to estimate  $p_{s|a}$  in a way that produces more realistic densities. One solution would be to widen the length classes. Another would be to estimate  $p_{s|a}$  from an appropriate parametric family.

For illustration, we considered the 1983 catch. To estimate  $p_a$ , we assumed that the  $p_{s|a}$  were known Normal with means given by the length-at-age table. For modest sample sizes, the representation of the variances should involve relatively few parameters. The variances were taken to be 7.45 (ages 4-10) and 26.37 (ages 11-20), which represented the individual variance for the ages quite closely.

The estimation procedure described in the Appendix converged successfully from starting values of  $p_a = \text{constant}$ . Estimated proportions are given in Table 1. The proportions for ages 12-15 converged to zero, their share having been allocated roughly to ages 11 and 16. Kimura & Chikuni (1987) bulked the ages 11+, and possibly this was their reason. This feature suggests that there is an inherent problem with estimating small proportions from length data based on limited data for constructing length-at-age distributions. That is, we cannot easily detect small components in a mixture distribution unless the separation between modes of length-at-age is large.

### 3 APPLICATION TO SAMPLING DESIGN FOR SBT

The formulae in the Appendix can be used to calculate precision and relative efficiency without needing simulation. Results depend only on the true proportions at age, the growth parameters, the number of samples ( $n$  and  $N$ ), and the subsampling protocol. We used the following setup:

1. The three fisheries we considered were: purse seiners in the GAB (young fish), Japanese long-liners (mostly fish of intermediate age), and Indonesian long-liners (older, mature fish). We truncated the age ranges to: 1-5 for the GAB, 2-16 with a plus-group at 17 for the Japanese long-liners, and 8-24 with a plus-group at 25 for the Indonesian fishery. These ranges contained at least 95% of the catch for the year considered in the example.
2. For the true values of  $p_a$ , we used one set of estimated numbers-at-age for the whole population in 1997, multiplied by selectivity-at-age by fishery. The values were taken from the conditioning results in the current Management Procedure development (Table 1). The run chosen had steepness set to 0.55 and natural mortality vector 2. Note that the conditioning model assumes an age-based rather than a length-based selectivity.
3. We used estimated means and variance of length-at-age from Polacheck et al (*ibid.*). In order to avoid the complication of within-year growth, we used means and variances for January only, rather than trying to integrate growth across a whole fishing season (see section 4).

4. We used three different values of  $n$  (100, 200, and 500) and four different values of  $N$  (200, 500, 1000, 2000), excluding the case  $n=500$  and  $N=200$ .
5. For the first set of results, the subsampling design was uniform sampling intensity across all lengths, with the number subsampled at length being proportional to numbers-at-length in the catch.
6. To ensure a sufficient number of samples within each length group for the ALK, we assumed that lengths would be aggregated into 5 cm groups before applying the ALK.

Table 2 shows the results for the parametric methods with known and unknown growth, and for the ALK. The "neq1" columns show what sample size would be required to give the same CV based on a random sample of ages with no length data. ARE1 and ARE2 show asymptotic relative efficiency of the ALK relative to the parametric known- and unknown-growth methods, i.e. the inverse ratio of the  $n$ -equivalents.

A general caveat on Table 2 is that very high CVs should not be taken too literally; they should rather be interpreted as "no useful information on this  $p_a$ ". The table shows *average* CV across all datasets, which reflects the *average* number of fish aged  $a$  in a sample. For an average  $n_a$  of 1, the CV will be about 100% (ignoring information from the length frequency). The average value of  $n_a$  is given by  $E[n_a] = np_a$ , and if  $E[n_a] < 1$ , the average CV will exceed 100%. However,  $n_a$  can only take integer values in real data, so real datasets will often have  $n_a=0$  for some ages. When  $n_a=0$ , the estimated CV for  $p_a$  is effectively infinite (unless a prior on  $p_a$  is assumed), although an upper confidence limit could still be found. When  $n_a=1$ , the estimated CV will suddenly drop to 100%. The general point is that the estimated CV will be highly variable for real datasets. This is a common phenomenon in statistics; the estimated uncertainty about a quantity is generally more uncertain than the quantity itself.

### 3.1 Results by method

In general, the parametric methods give substantially higher precision than the ALK; they are considerably more efficient because they use length frequency information. Also, the ALK requires that length samples be grouped more coarsely. For the sample sizes we considered here, the smallest feasible grouping (into 5cm classes) is similar to the average standard deviation of length-at-age, and so the overlap between ages within length classes is increased.

The difference in efficiency between ALK and parametric methods is greatest for younger fish, where length is a more precise predictor of age. In summarizing results by fishery below, we concentrate on results from the parametric methods, because these best show the limits to attainable efficiency.

The  $n$ -equivalent for ALK can be less than  $n$ . That is, the estimate can sometimes be less efficient than a purely random sample of size  $n$  with no length measurements. From a purely random sample, we would estimate  $p_a$  from  $n_a/n$ , the observed proportion in age class  $a$ . Because the ALK does not assume random sampling, we cannot use this estimate, but must scale the proportions in each length class by  $N_s/N$ . This introduces an extra component of error, larger if  $N$  is smaller, which can more than compensate the gain due to breaking the total sample into the length classes.

There is surprisingly little difference between the known-growth and unknown-growth estimators. The greatest differences occur when  $n$  is small (so that there is little data available for estimating length-at-age within each age class) but  $N$  is large (so that any errors in  $\hat{p}_{s|a}$  are magnified by the large information content of the length-frequency data).. A heuristic explanation for the general lack of difference might be as follows: for older fish, the length data (and thus the Normal parameters) are of little value, while for younger fish, there is little overlap in the distributions between adjacent ages, so estimation of the Normal parameters is not influential.

### 3.2 Results by fishery

#### GAB

For such young fish, the length data alone are very informative; equivalent fully-aged sampled sizes (neq1 and neq2) suggest that each age sample is worth between 2 and 4 length samples in terms of information content about age. The sample size required in the GAB is small; CVs under 20% can be obtained even with  $n=100$  and  $N=200$ , at least for the ages 2-4 which are by far the most important.

#### Indonesian long-liners

The length frequency data provide almost no information about age. The ALK performs almost as well as the parametric methods, at least for ages 12 up, except in the case  $n=100$  and  $N=200$ . The parametric methods work substantially better than ALK for ages 8-11, presumably because of the increased efficiency obtained by imposing a parametric form on  $p_{s|a}$ . When  $n=100$ , CVs of  $\hat{p}_a$  are very high (at least 33% for all age classes), and there is a substantial risk that several age classes may simply fail to show up in the  $n$ -sample at all.

#### Japanese long-liners

The bulk of the catch is aged 3-9, and the commonest age was 6 in 1997. The length frequency data are certainly useful, at least for smaller fish, because equivalent- $n$  is substantially higher than  $n$ . Table 3 shows comparative CVs for 3, 6 and 9-year-olds, using the parametric-unknown-growth estimator. The law of diminishing returns seems to set in for  $N$  above about 500. There are clear gains to be made for  $n=500$  compared to  $n=100$ , especially for fish aged 6 and up.

#### Comparison between fisheries

It is not surprising that precision is much better for the GAB than elsewhere. In the GAB, 95% of the catch is split between 3 age classes, so the number of fish sampled per age class will be rather high. For the other fisheries, there are more possible age classes, so the number of fish per age class will be lower. As with any sampling

problem that is fundamentally multinomial, fewer samples per class means lower relative precision per class. Also, the greater separation between length-at-age distributions in the GAB means that the length frequency data is relatively more informative than elsewhere.

### 3.2.2 Effect of varying the subsampling pattern

Table 2 also shows that uniform intensity subsampling (i.e. in proportion to numbers-at-length) usually leads to relatively higher precision for younger fish compared to older fish. This may not be desirable (though see section 4). To show what improvements might be possible, we also investigated a non-uniform subsampling within limited stratification by length for the Japanese long-line fishery only. We chose this fishery because it has the most to gain from subsampling; for the GAB fishery, length alone is very informative, while for the Indonesian fishery, there is little difference in age composition across the length frequencies, and stratification may not be logistically feasible.

We investigated the following subsampling scheme: length was split into 3 classes (<120cm; 120-130cm; >130cm), and subsamples were taken within each group in proportion to weighted numbers within the length classes (rather than unweighted as before). The weights were in the ratio 1:5:10, so that fish >130 cm were over-represented relative to fish <120cm by a factor of 10. Results are shown in Table 4.

In general, this subsampling pattern led to a relative improvement in CV of about 25% for fish aged 9 (i.e. the stratified CV was about 3/4 of the unstratified CV). There was little change for fish aged 6. Under the known-growth model, the CV for age-3 fish increased only very slightly even though fewer were sampled, provided  $N$  was at least 1000. This is because the length frequency data alone is very informative about 3 year olds, and the length frequency data is unaffected by subsampling. In the unknown-growth model, however, the reduced sampling of smaller fish means that length-at-age parameters were estimated worse, and precision dropped dramatically. For example, with  $n=200$  and  $N=1000$ , the CV for 3-year-olds under the unknown-growth model changed from 11% without stratification to 20% with stratification.

## 4 CONCLUSIONS AND DISCUSSION

1. The parametric methods can often be used effectively with considerably smaller age samples than can ALK. Using the length distribution greatly increases the information for the younger ages. It has less advantage for the older fish, for which length-at-age distributions overlap more.
2. The *potential* gain in robustness of ALK compared to parametric methods, basically stems from ALK not having to assume a Normal (or some other) distribution for length-at-age. However, in their study of SBT growth, Polacheck *et al.* (*ibid.*) show that length-at-age is reasonably well described by a Normal distribution, at least for younger ages where the length data is of

most importance. Consequently, there seems no benefit from staying with ALK for SBT.

3. The parametric-unknown-growth estimator is almost as efficient as the known-growth estimator when age-subsampling is uniform intensity (i.e. numbers subsampled at length are proportional to numbers caught at length), and is presumably more robust since it does not assume identical growth patterns across cohorts. If age-subsampling in the Japanese long-line fishery was heavily skewed towards larger fish, though, the unknown-growth estimator would become substantially less efficient.
4. Judicious subsampling design (i.e. subsampling relatively more big fish) can somewhat improve precision in estimating the proportions of ages of older fish, but only by a modest amount (say 25%). If length-at-age distributions can be assumed known for younger fish, so that the known-growth estimator can be used, then such subsampling will not lead to a corresponding drop in precision for younger fish, for which there is extra information from the length data alone. However, if there are concerns about variations in length-at-age between cohorts, then the unknown-growth estimator should be used instead. In that case, subsampling relatively more big fish will reduce precision on younger age classes.
5. The main determinant of precision is the number of otoliths collected, plus the effective size of the length sample in the GAB fishery and (to a lesser extent) in the Japanese long-line fishery. Subsampling design has less impact. The minimum viable sample size depends on fishery, as well as on target precision. As a guideline, the results suggest CVs of no more than 25% for the commonest ages (those constituting at least 5% of the catch) might in principle be achieved with  $n=100$  and  $N=200$  in the GAB (ages 2-4), but would need  $n=200$  and  $N=500$  in the Japanese long-line fishery (ages 3-9); but see section 4.1.4 for a caveat. Note that  $N$  is effective length sample size, so the actual number of fish measured would need to be greater (see section 4.1.3). The Indonesian fishery has more age classes and the length data is not informative about age, so the sample size needs to be larger. Based on 1997 proportions, the 25% CV ceiling for ages 13+ is achieved for  $n=500$ , but the CV remains above 30% for ages 8-12. If the proportion of young fish increased, their CV would drop.

## 4.1 Discussion points

### 4.1.1 Within-season growth

Fishing seasons for SBT are quite long and individual fish can show substantial growth within a single season, especially when young. The results in this paper ignore this, and therefore pertain to an idealized situation of "point sampling". As explained below, we do not think this deliberate oversight has any major implications for our results on precision, but it is nevertheless important to consider how within-season growth might be dealt with statistically.

One way to handle within-year growth, is simply to aggregate all data within a fishing season without regard to time of collection. If prior information on  $p_{s|a}$  is available, then this may already be aggregated, or may be a function of time within the season. If the latter, it is possible to "smear out" growth by integrating across the season, e.g. to approximate the time-aggregated mean and variance.

However, this kind of aggregation is not a good idea. Smearing growth across a season increases the overlap between the length distributions of successive cohorts, and therefore makes it harder to separate cohorts using length data. A much better approach is to extend the estimators in section 2 so that they can be applied at different times during the season, keeping  $p_a$  constant across times even though  $p_{s|a}$  and  $p_s$  might vary. Then the intervals of aggregation can be much shorter (one month, say), and there is no serious loss in efficiency, at least in principle. For the parametric estimator with known  $p_{s|a}$ , the modifications should not be difficult and there should be no major effect on precision since there are no extra parameters. For the parametric estimator with unknown growth, the modifications are easy in principle, but extra parameters are needed to describe length-at-age at different times during the season. It might be desirable to impose some extra constraints over the season, such as constant  $\sigma^2$  and linearly-increasing  $\mu$ , to prevent "shrinking fish syndrome".

The ALK would require more extensive overhaul to handle within-season growth, both in terms of the estimator itself and how to avoid implying bizarre changes in  $p_{s|a,t}$  from one time  $t$  to the next. This would be difficult and provides another reason to use a parametric method, which are simpler to modify.

#### 4.1.2 How should precision vary with age?

This is not obvious. The results in Table 2 show that precision is much lower for older fish, especially in the Indonesian fishery, unless there is a deliberate effort to subsample relatively more big fish (Table 3). At first sight, this lower precision looks worrying. However, equal precision on all age classes may not actually be desirable. As a cohort progresses through the fishery, information on its relative abundance will accrue year-by-year; thus the eventual uncertainty about the cohort's abundance at age 20 may be low even if the CV of a single estimate of  $p_{a=20}$  is high. It *may* be more important to aim for higher precision for younger cohorts, so that precise information about recent recruitment can be obtained relatively fast; this might help with short-term forecasts and quota setting. On the other hand, it *may* be more important to get higher precision on older fish, e.g. because these form the bulk of the spawning stock and thus influence medium- to long-term yields. The ideal balance between subsamples of big and small fish could be studied in future in an assessment context. However, it is not necessary to get the balance perfect before starting data collection. Design of subsampling has only a limited effect on precision, and the main determinant of precision is the overall level of coverage.

#### 4.1.3 Effective sample size for the length frequency data

Non-independence of length measurements can be thought of on two quite different space/time scales. First, if SBT schools are somewhat length-segregated, then there

may be set-to-set variability because some sets will encounter schools of small fish, while other sets will mostly encounter schools of bigger fish. This is obviously true for purse-seiners but may apply to long-liners, too. Also, differences in fishing practice between boats might lead to systematic variations in length frequency samples at the boat level (i.e. selectivity differences). Logistics demand that multiple samples be taken from the same set and the same boat, so small-scale differences of this nature increase the uncertainty of estimated proportions-at-length relative to what could be achieved from truly independent length samples.

Second, there may be systematic variations in length composition across the spatial extent of a fishery. This can lead not just to increased variance, but also to bias if some sub-areas are relatively over- or under-sampled. This potential problem can be diagnosed, and largely corrected post hoc if necessary, provided that length data are available disaggregated by sub-area, along with records of total catch by sub-area. As discussed above, it is also necessary to avoid aggregating the data over an entire fishing season.

The effective sample size can be estimated statistically by using a bootstrap at a sufficiently coarse unit of sampling to get around the small-scale variability problem: e.g. by set if fishing practices between boats are very similar, or otherwise by boat. If large-scale variations have been identified, a stratified bootstrap would be better (i.e. separate bootstraps within sub-areas). The bootstrap replicates provide an estimate  $v_s$  of  $\mathbf{V}[\hat{p}_s]$  for each length class  $s$ . For a truly independent sample of size  $N$ , this variance would be  $\hat{p}_s(1 - \hat{p}_s)/N$ , so  $N$  can be estimated by equating this expression to  $v_s$ . This gives one estimate of  $N$  per length class. These could be combined into a weighted average or a median to give an overall estimate of  $N$ . Extreme accuracy is not crucial here. It is also not necessary to have set-by-set data to do this calculation; it would be sufficient to report the bootstrap variances and proportions-at-age.

Until these or similar appropriate statistics from disaggregated data are available, it would not be possible to make use of the length data in our parametric methods, so the precision of age composition estimates would be reduced compared to the Tables (or more otoliths would be needed). Pending the availability of data to allow effective sample size to be estimated, it makes sense to collect as much length data as possible, because it is cheap.

#### 4.1.4 Errors in age and length measurements

Polacheck *et al.* (*ibid.*) discuss problems associated with age determination from SBT otoliths. The timing of ring formation is not fixed, generally occurring at some point between May and September but not at a consistent date across fish. Also, the interpretation of otoliths during this period is difficult because it is hard to distinguish new rings. And although readings of matched otolith pairs from the same individual are generally consistent, there are exceptions, even among experienced readers and even for otoliths collected outside the difficult May-September season.

It is possible to envisage various possible ways that these issues might be addressed, both statistical and technical. For example, if otoliths in May are reliable, and otoliths

in September are reliable, and the proportion-at-true-age remains constant through the fishing season, then a missing-data model might be constructed with recorded age assumed to follow the (unobservable) true age but possibly subject to errors that have a date-dependent distribution. Further investigation of such issues would require reasonably large samples of otoliths collected at different times during a season, and would therefore have to wait until after an otolith collection scheme has begun.

There is also the possibility of length measurement errors, mainly individual biases due to differences in measurement technique. Such errors are presumably of less concern than ageing errors, but it is difficult to assess the likely impact on precision/bias without more data to compare observer and non-observer measurements; again, this can only be done after an observation scheme has begun.

#### 4.1.5 Further methodological work

1. The most important development for practical application is to adapt the parametric estimators to handle within-season growth. This would be difficult for an ALK, but should be fairly easy for the parametric estimators, as noted above.
2. If errors in ageing are thought likely to be a significant issue, then further statistical work will be required. This would require extensive otolith samples from the full course of a season.
3. Bravington (CCSBT-CPUE/0203/5) suggested that it might be worth developing an estimator that tracks cohort-specific growth effects over time. Such an estimator ought to combine the best features of the parametric-known-growth (efficient but not robust to growth changes) and the parametric-unknown-growth estimators (robust but not efficient). However, such a complex estimator would only be worth developing if the task of re-estimating length-at-age distributions every year proved to be "expensive" in terms of lost precision in  $\hat{p}_a$ . Table 2 suggests that this is not the case; the unknown-growth estimates were almost as efficient as the known-growth estimates. The added complexity is therefore probably not worthwhile pursuing for SBT. An exception would be if it is decided to heavily oversample larger fish, since in that case there is a large difference between precision of the known-growth and unknown-growth estimates. Some further study of variations in size-at-age between cohorts would also be valuable, since if variation is shown to be very limited, the known-growth model could be safely used anyway.
4. It is also probably not worth expending effort on developing less-parametric estimators that do not assume a Normal distribution of length-at-age (Bravington *ibid.*). Polacheck *et al.* (*ibid.*) conclude that a Normal distribution gives a reasonable fit. If any of the fisheries showed very steep gradients in selectivity-at-length (as opposed to selectivity-at-age) then this might be worth revisiting, since the distribution of length-at-age in the catch might then be quite far from Normal. At present, there is no evidence that this is the case.

5. Consideration will ultimately have to be given as to how to incorporate explicit age data into the assessment (e.g. whether to fit the age/length data separately using the models we have described and then carry the likelihood function into the assessment, or to estimate proportions-at-age simultaneously with all other parameters inside some gigantic composite estimator). Although some decision about this will eventually be required, there is no urgent requirement to resolve this before data collection begins, since there are no particular implications for precision or design.

## 4.2 Summary

To get estimates of proportion-at-age with CVs of 25% or less for all ages that constitute at least 5% of the catch, at least the following numbers of randomly-subsampled otoliths would be needed: 100 (GAB), 200 (Japanese long-lines), or 500 (Indonesian long-lines). This assumes that large and representative length frequency samples are also available, and that there are no significant errors in ageing or length measurement. Ageing errors in particular would increase the required sample size. It is also necessary that sampling be spread through the season within the season (recorded by month, say) and to avoid using data that has been aggregated across the whole season; otherwise, the overlap between length-at-age distributions is artificially increased, and estimation of age composition becomes less precise.

Achieving such good precision for such small numbers of otoliths would require use of the new parametric methods introduced in section 2.1.4, because a standard age-length key is substantially less efficient (although in the Indonesian fishery there is not much difference). However, reliable use of the new methods does depend on knowing the "effective sample size" of the length frequency data, which will be smaller than the real sample size because of non-independence between length samples. The data needed for estimating effective sample size could be collected cheaply during age- or length-sampling.

All else being equal, CVs within a fishery are generally higher for older fish, and gains in efficiency of say 25% for some age classes might be achievable through non-uniform subsampling of length groups, i.e. oversampling bigger fish. However, the details of subsampling design have much less impact on precision than the number of otoliths collected. Further details of design or analysis would be best resolved after a scheme has been started, when more data become available.

## 5 REFERENCES

Heyde,C.C. (1997) *Quasi-likelihood and its application. A general approach to optimal parameter estimation*. N.Y.: Springer.

Kimura,D.K. & Chikuni,S. (1987) "Mixtures of empirical distributions: An iterative application of the Age-Length Key" *Biometrics* **43** 23-35.

McLachlan, G.J. & Basford, K.E. (1988) *Mixture models. Inferences and applications to clustering*. N.Y.: Marcel Dekker.

Polacheck, T., Laslett, G.M., & Eveson, J.P. (2003) *An integrated analysis of the growth rates of southern bluefin tuna for use in estimating the catch-at-age matrix in the stock assessment*. Report FRDC 1999/104 to Fisheries Research and Development Corporation, Canberra, Australia.

## 6 APPENDIX: THEORETICAL DETAILS

We are given two data sets from catches of fish. The first of size  $n$  has pairs  $(s_i, a_i)$  of length (or size) measurements ( $s$ ) and ages ( $a$ ). The results are condensed into frequencies  $n_{as}$  giving the number of fish out of a total of  $n_s$  in length class  $s$  having age class  $a$ . Thus  $n = \sum_s n_s$ . The second consists of  $N$  measurements of length. The total condensed to frequencies  $N_s$ , with  $N = \sum_s N_s$ . Typically,  $N$  is much larger than  $n$ . Unless stated otherwise, we shall assume that the age sample is drawn as a subsample from the  $N$  fish that were measured for length, so there would be  $N - n$  fish measured for length only. The aim is to estimate the proportion  $p_a$  of the catch in age class  $a$ . We use the notation  $f(s)$  for the probability that a fish has length  $s$ ,  $f(s|a)$  for the probability that a fish has length  $s$  given age  $a$  and  $f(a|s)$  for the probability that a fish of length  $s$  has age  $a$ .

[**Note:** there are some minor notational differences between this Appendix and the main text. In particular, the main text uses  $p$ 's rather than  $f$ 's, for example writing  $p_{s|a}$  instead of  $f(s|a)$ .]

A distinction must be drawn according to how we regard the sample  $n$  to have been drawn:

- (i) Age-at-length; the  $n_s$  are treated as fixed and age is random. Since recruitment can change substantially with cohort, the age-at-length distribution must relate to the population under consideration.
- (ii) Length-at-age; the totals  $n_a$  are treated as fixed and length is random. This is generally unachievable in most fisheries since age cannot be determined on sight. However, length-at-age distributions can be estimated from a random sample, or may be taken as estimated in some way from previous data.
- (iii)  $n$  is a completely random sample from the population under consideration. It may be treated as random with respect to either age or length, e.g. to construct length-at-age distributions.

We describe briefly the methods commonly used to estimate the age profiles.

### Age-Length Key

Here we must assume that the  $n$  fish are from the same population as the  $N$ . In the spirit of the age-length key (ALK) we estimate

$$p_a = \sum_s f(s) f(a|s), \quad (\text{A1})$$

where  $f(s)$  is estimated by  $N_s/N$ , and  $f(a|s)$  from the  $n$  sample. In the ALK method, the estimate of the latter is simply  $n_{as}/n_s$ . This is a distribution-free method.

### Iterated Age-Length Key

The Iterated Age-Length Key (IALK) uses the same formula (A1) but the estimate of  $f(a|s)$  is via Bayes formula:

$$f(a|s) = p_a f(s|a) / \sum_a \{p_a f(s|a)\} \quad (A2)$$

Under the assumption that  $f(s|a)$  does not change from year to year, the  $n$  sample may be collected in year A, and used together with an  $N$  sample also from year A to estimate  $f(s|a)$ . This may then be applied to an  $N$  sample from a different year, B, to estimate  $p_a$  in year B. The appearance of  $p_a$  means that (A2) is solved by iterative substitution.

### Random sampling

Note that random sampling (e.g. of every 17<sup>th</sup> fish caught) sounds easy but may not always be so in practice, especially for schooling species. If the  $n$  sample really is obtained by random sampling, though, an estimate of the age proportions is provided by the proportion of the ages found in the  $n$  sample. However, the  $N$  length measurements could also be used in the ALK method, as recommended by Kimura & Chikuni (1987). The two estimates should be very similar; any difference would imply that the distribution of length in the  $N$  sample is not the same as in the  $n$  sample, and this would cast doubt on the validity of the random sampling.

## 6.1 Parametric Methods and Mixture Distributions

### 6.1.1 Estimation

The marginal density function is the mixture, with mixing parameters  $\{p_a\}$

$$f(s) = \sum_a p_a f(s|a). \quad (A3)$$

If the  $f(s|a)$  are regarded as known to adequate precision, then the length data alone can in principle provide estimates of  $\{p_a\}$  by maximizing the log-likelihood

$$\Lambda = \sum_s N(s) \log \{f(s)\}. \quad (A4)$$

However, if the densities  $\{f(s|a)\}$  have a large amount of overlap, there would be little information available for the estimation. This point is discussed below. For more information on estimating mixture distributions, see McLachlan & Basford (1988).

If the age-at-length sample is a subsample of the length sample, then the log-likelihood is

$$\Lambda = \sum_s [(N_s - n_s) \log \{f(s)\} + \sum_a n_{as} \log \{p_a f(s|a)\}]. \quad (A5)$$

It is shown in Appendix 1 that the log-likelihood has the same form (up to a constant) if the  $n$  is a subsample from the  $N$  length measurements. We are not primarily concerned with the  $n$  being length-at-age, but if the sample is selected on the basis of age, it contains no information about  $\{p_a\}$ , and this is reflected in the fact that the  $p_a$  is to be dropped in the last term in (A5).

More generally  $f(s|a)$  might have a parametric form with parameters to be estimated. For theoretical simplicity, we will consider only that for the Normal distribution  $N(\mu_a, \sigma_a^2)$ . In that case, the Normal parameters will need to be estimated as well as the proportions. Some reduction in the number of parameters might be reasonable if we can make simplifying assumptions such as the  $\{\sigma_a\}$  all being equal. We may wish to impose the constraints on the  $\{\mu_a\}$  to ensure that they are isotonic (i.e. non-decreasing with age). This is a complication for the theory below, and we assume for the present that the data do invoke these constraints.

Maximization of  $\Lambda$  is subject to the constraint  $\sum_a p_a = 1$ . This may be achieved by using a Lagrange multiplier, or by removing one parameter by substitution. The unbiased likelihood score for  $p_a$  is

$$U_a = \sum_s \{ (N_s - n_s) f(s|a)/f(s) + n_{as}/p_a - N_s \}. \quad (\text{A6})$$

Here, the totals sampled for age in the length class  $s$ ,  $n_s$ , has been taken as fixed. If it is random, it should be replaced by its expectation  $E(n_s)$ . The score may be algebraically rearranged to give the equation for solution by iterative substitution

$$p_a = \sum_s N_{as}^* / N, \quad (\text{A7})$$

where we call the term in the numerator,

$$N_{as}^* = (N_s - n_s) f(a|s) + n_{as}, \quad (\text{A8})$$

the *estimated* number of fish in age class  $a$  at length  $s$  in the combined samples, and  $f(a|s) = p_a f(a|s)/f(a)$  is the ('posterior') probability of  $a$  given  $s$ . This gives (A9)-(A10) an attractive interpretation as the estimated moments. The denominator term in (A7) is  $\sum_s E(N_{as}^*)/p_a$ . Equation (A5) bears some resemblance to the IALK method, and is equivalent to the E-M algorithm.

The likelihood score for  $\mu_a$  and  $\sigma_a^2$  can be rearranged to the form

$$\mu_a = \sum_s N_{as}^* s / \sum_s N_{as}^*, \quad (\text{A9})$$

$$\sigma_a^2 = \sum_a N_{as}^* (s - \mu_a)^2 / \sum_a N_{as}^*. \quad (\text{A10})$$

The right hand side of each of these equations involves the parameters, which have been suppressed in our notation. Using these equations, the parameters  $\{p_a, \mu_a, \sigma_a^2\}$  may be estimated by iterative substitution.

Iterative substitutions can be performed rapidly, so that many thousand are quite feasible. This method of estimation avoids the need to calculate and invert the Hessian matrix, and does not seem to run into the numerical difficulties encountered in using

the Newton-Raphson maximization of the log-likelihood. However, it should be noted that the variance of the estimates are not simply derivable from these equations, since the parameters appear on both sides. One possibility is to use these estimates as starting values for maximizing the log-likelihood using a standard statistical package. We discuss the information matrix and the variance-covariance matrix of the estimates below. These matrices may be computed directly, since formulae are available.

### 6.1.2 No length data

We shall also consider the special case where the length only data are absent or ignored. In the possibly unrealistic case that the whole sample is measured for age and length, then  $N_s = n_s$ . We see that the likelihood (A5) is, up to a constant, that of the multinomial. This is correct on the premise (as we assumed for  $N$ ) that the  $n$  were a random sample from the whole population. *We shall use the information in a multinomial as a standard for assessing the information in contributed by the length sample.*

If, however, the age sample were selected on length, then we should condition on the length sample sizes; this subtracts  $\sum_s n_s \log\{f(s)\}$  from  $\Lambda$  as given in (A5). We then maximize the reduced likelihood

$$\Lambda = \sum_s \sum_a n_{as} \log\{p_a f(s|a)/f(s)\} \quad (\text{A5a})$$

using a Lagrange multiplier, which turns out to be zero. Although they appear sufficient in number, solving the partial derivative equations does not produce estimates – there is still an arbitrary scale for the solution to the estimated  $p$ 's. We obtain the equation

$$p_a = \kappa \sum_s n_{as} / \sum_s \{n_s f(s|a)/f(s)\}. \quad (\text{A7a})$$

The numerator is the same as (A7) with  $N_s = n_s$ , but multiplied by  $\kappa$ , a norming constant to ensure that  $\sum_a p_a = 1$ . The likelihood (A5a) is the same as (A5) but with  $N_s$  replaced by  $\theta$ . This replacement cannot be used in (A7) because it makes the denominator zero. However, the second derivatives (A13)-(A18) below are still valid with this substitution, and the information and asymptotic variance matrices above still hold.

## 6.2 Information for the estimates

The information for the estimates follows from the usual likelihood theory, provided that there is no extra variation. A similar theory applies when there is extra variation and is based on quasi-likelihood, see Heyde (1997). To start with, we shall assume that there is no extra variation. The Fisher information matrix is the expectation of the second derivatives of the log-likelihood, with a modification due to the fact that the  $p_a$  are constrained to sum to 1. This may be handled by eliminating a redundant parameter.

The asymptotic variance matrix for the parameter estimates is the inverse of the information matrix (generalized inverse if the projection method is used).

We consider several cases.

### 6.2.1 Two age classes, length-at-age distributions known

Suppose that there are just two age classes. We represent  $p_1 = p$ ,  $p_2 = 1 - p$ . Assume that the length-given-age distributions are known and so do not require to be estimated. The result of this assumption is that the precision with which we can estimate  $p$  will be optimistic. From the length sample  $l_{aone}$ , the Fisher information for  $p$  per individual is

$$i = \sum_{sw} \{f(s|a=1) - f(s|a=2)\}^2 / \{pf(s|a=1) + (1-p)f(s|a=2)\} \quad (A11)$$

The asymptotic relative efficiency (ARE) compared to measuring age is  $i/\{p(1-p)\}$ . To proceed, it is convenient to replace the summation by an integral; in effect, we assume that the length classes are very small, or that  $s$  is a continuous variate. For Normal distributions with equal variances, this is a function of  $p$  and the separation of the means relative to the standard deviation

$$\Delta = (\mu_1 - \mu_2)/\sigma. \quad (A12)$$

Table A1 gives the values of the ARE for  $\Delta$  and  $p$ . If the number of age classes is greater than 2, the ability to discriminate between these two classes is reduced due to the overlap with the other classes. We see that, even under the optimistic conditions in Table 1, the length data alone is of little value for estimating the age proportions when  $\Delta$  is less than about 0.25.

To interpret Table A1 in practical terms, consider the graph of the Age-Length data. For  $\Delta = 0.75$ ,  $p = 0.5$ , the ARE is about 1/8. That means that means that when the length means for two ages differ by  $0.75\sigma$ , the discrimination obtained from 8 length measurements is roughly equivalent to 1 age measurement. We shall estimate  $\sigma$  later, but suppose that it is about 6 cm. Then this discriminatory power would apply to means separated by about 4.5 cm. For ages over 15, this separation could span several years.

### 6.2.2 $K$ age classes, length-at-age parameters to be estimated

Now we consider the case of  $K \geq 2$  age classes. To simplify the typing, define  $N_{as}^*$  as in (8) and denote

$$c_{ab}(s) = f(a|s)\{\delta_{ab} - f(b|s)\},$$

$$m_{1a}(s) = \partial \log f(s|a) / \partial \mu_a,$$

$$m_{2a}(s) = \partial \log f(s|a) / \partial \sigma_a,$$

$$m_{11a}(s) = \partial^2 \log f(s|a) / \partial \mu_a^2,$$

$$m_{12a}(s) = \partial^2 \log f(s|a) / \partial \mu_a \partial \sigma_a,$$

$$m_{22a}(s) = \partial^2 \log f(s|a) / \partial \sigma_a^2.$$

After multiplication by  $n_s$ ,  $c_{ab}(s)$  is the (a,b) element of the multinomial covariance matrix, but this fact does not seem to provide much insight. Then

$$\partial^2 \Lambda / \partial p_a \partial p_b = - \sum_s \{ (N_s - n_s) f(a|s) f(b|s) + n_{as} \delta_{ab} \} / (p_a p_b), \quad (\text{A13})$$

$$\partial^2 \Lambda / \partial p_a \partial \mu_b = \sum_s \{ (N_s - n_s) c_{ab}(s) / p_a \} m_{1b}(s) \quad (\text{A14})$$

$$\partial^2 \Lambda / \partial p_a \partial \sigma_b = \sum_s \{ (N_s - n_s) c_{ab}(s) / p_a \} m_{2b}(s) \quad (\text{A15})$$

$$\partial^2 \Lambda / \partial \mu_a \partial \mu_b = \sum_s \{ (N_s - n_s) c_{ab}(s) m_{1a}(s) m_{1b}(s) - N_{as}^* \delta_{ab} m_{11a}(s) \}, \quad (\text{A16})$$

$$\partial^2 \Lambda / \partial \mu_a \partial \sigma_b = \sum_s \{ (N_s - n_s) c_{ab}(s) m_{1a}(s) m_{2b}(s) - N_{as}^* \delta_{ab} m_{12a}(s) \} \quad (\text{A17})$$

$$\partial^2 \Lambda / \partial \sigma_a \partial \sigma_b = \sum_s \{ (N_s - n_s) c_{ab}(s) m_{2a}(s) m_{2b}(s) - N_{as}^* \delta_{ab} m_{22a}(s) \}. \quad (\text{A18})$$

If  $f(s|a)$  can be taken as the Normal density centred on the midpoint of the interval, the  $m$ 's have the familiar simple form:

$$m_{1a}(s) = (s - \mu_a) / \sigma_a^2$$

$$m_{2a}(s) = \{ (s - \mu_a)^2 - \sigma_a^2 \} / \sigma_a^3$$

$$m_{11a}(s) = -1 / \sigma_a^2$$

$$m_{12a}(s) = -2m_{1a}(s) / \sigma_a$$

$$m_{22a}(s) = -3m_{1a}^2(s) + \sigma_a^{-2}.$$

This approximation is good enough for some purposes and is valid when the interval is small, as when  $s$  is treated as a continuous variate or the grouping intervals have width 1 cm. The resulting formulae are used in the Tables below for evaluating the standard errors of the estimates and the equivalent sample size when using the likelihood approach. However, when the interval is not small,  $f(s|a)$  is the integral of the Normal density over the interval and the form of the  $m$ 's is more complicated. The computation of the information matrix  $I_{pp}^*$  below seems to be particularly sensitive to this approximation.

Note that, unlike the ALK approach, there is no need to group the lengths into intervals, so that we could treat them as continuous variates. This has a substantial advantage in that the grouping into 5 cm intervals reduces the information about the parameters for the younger ages.

To eliminate the redundancy in the parameters, we now temporarily substitute  $p_K = I - \sum_{a=1}^{K-1} p_a$ . Define the partitioned matrix  $Q = (I_{K-1} - \mathbf{1}_{K-1})$ , where  $I_{K-1}$  is the identity matrix of order  $K - 1$ , and  $\mathbf{1}_{K-1}$  is the vector of  $K - 1$  ones.

The information matrix, partitioned according to the groups of parameters  $\mathbf{p}$  and  $\boldsymbol{\theta} = (\boldsymbol{\mu}^T, \boldsymbol{\sigma}^T)^T$ , is minus the expectation of the matrix of these second order partial derivatives, modified by pre-multiplying by and/or post-multiplying by the matrix  $Q^T$  to get

$$I = \begin{pmatrix} I_{pp} & I_{p\theta} \\ I_{\theta p} & I_{\theta\theta} \end{pmatrix}, \quad (\text{A19})$$

where

$$I_{pp} = -Q E(\partial^2 \Lambda / \partial \mathbf{p} \partial \mathbf{p}^T) Q^T, \quad (\text{A20})$$

$$I_{p\theta} = -Q E(\partial^2 \Lambda / \partial \mathbf{p} \partial \boldsymbol{\theta}^T) \quad (\text{A21})$$

$$I_{\theta p} = I_{p\theta}^T, \text{ and} \quad (\text{A22})$$

$$I_{\theta\theta} = -E(\partial^2 \Lambda / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T). \quad (\text{A23})$$

The asymptotic variance-covariance matrix of the reduced parameters is the inverse of  $I$ . To get back to the original parameters, pre-multiply by  $Q^T$  and/or post-multiply by  $Q$  the corresponding partitions of  $I^{-1}$ .

If  $\boldsymbol{\theta}$  is treated as known without error (or, more generally, if the densities  $f(s|a)$  are known), then the information matrix is confined to  $I_{pp}$ .

The reduced information about  $\mathbf{p}$  due to estimating  $\boldsymbol{\theta}$  is

$$I_{pp}^* = I_{pp} - I_{p\theta} I_{\theta\theta}^{-1} I_{\theta p}, \quad (\text{A24})$$

the second term indicating how much information about  $\mathbf{p}$  has been lost through having to estimate  $\boldsymbol{\theta}$ . If the densities are regarded as known, the second term is to be omitted.

The covariance matrix for the estimates of  $\mathbf{p}$  is  $Q^T (I_{pp})^{-1} Q$  or  $Q^T (I_{pp}^*)^{-1} Q$ , accordingly as  $\{f(s|a)\}$  are (i) known or (ii) estimated. So the leading diagonal elements of this matrix give the asymptotic variances.

The ratio of the multinomial variance  $p_a(I - p_a)$  to the variance of the estimate of  $p_a$  gives the what we call the equivalent sample size (from a multinomial), i.e. if the whole random sample were aged. This equivalent sample size varies according to how much information the length data provide about  $p_a$ , and we compare the two values to describe the loss of information from the length sample due to having to estimate  $\boldsymbol{\theta}$ .

### 6.2.3 ALK

The  $\{n_s\}$  are fixed, and so  $\{n_{as} : a=1\dots K\}$  are multinomial with proportions  $f(a|s)$ . Using standard formulae for the covariances of products, the covariance for the estimated estimated proportions is

$$\text{cov}(p_a, p_b) = \sum_s [\{f(s)/N + f(s)^2\} c_{ab}(s)/n_s + f(a|s)f(b|s)f(y)/N] - p_a p_b / N. \quad (\text{A25})$$

Since all the densities have to be estimated from the empirical frequencies, these covariances are liable to be very poorly estimated if the  $\{N_s\}$  are not large. For comparison with the MLE, one could substitute the estimated Normal densities. This would give some idea how much information is lost by using the ALK (distribution-free) estimate.

For large  $N$ , the terms in  $1/N$  are negligible. Suppose that  $n_s = n f(s)$ . Then the variance of  $p_a$  reduces to  $[p_a(1 - p_a) - \text{var}\{f(a|s)\}] / n$ . Thus ALK is more efficient (having a smaller variance) than the estimator  $\sum_s n_{as}/n$  from the multinomial distribution. That is, under proportional sampling for aging, there is an advantage in using ALK provided  $N$  is large.

### 6.3 Likelihood Under Subsampling

The length sample  $\mathbf{s} = (s_1, \dots, s_N)^T$  assumed random from a large population. Ages, mostly unmeasured, are  $\mathbf{a} = (a_1, \dots, a_N)^T$ . A subsample  $S$  is chosen for aging based on lengths; the subvectors are denoted by  $\mathbf{s}_S$  and  $\mathbf{a}_S$ , the remainder by  $\mathbf{s}_{\setminus S}$  and  $\mathbf{a}_{\setminus S}$ . In this Appendix, we distinguish the age  $a$  as a random variate from the age class  $k$ .

For brevity,  $(A|B)$  is the density of  $A$  given  $B$ . Parameters  $\mathbf{p} = (p_1, \dots, p_K)^T$ ;  $p_k$  = proportion of age class  $k$  in the population from which  $\mathbf{s}$  was drawn. There may be further parameters to be estimated in  $(\mathbf{s}|\mathbf{a})$ . We assume there are no random effects; the distributional properties include:

$$(\text{AA1}) \ (\mathbf{s}) = \prod_i (s_i) \text{ marginally independent, } (s_i) = \sum_k (s_i|k)p_k.$$

$$(\text{AA2}) \ (\mathbf{a}|\mathbf{s}) = \prod_i (a_i|s_i)$$

(AA3)  $S|\mathbf{s}$  may be random or not.  $(S|\mathbf{s})$  depends on  $\mathbf{s}$ . It does not depend upon any of the parameters.

$$(\text{AA4}) \ (\mathbf{a}|\mathbf{s}, S) = (\mathbf{a}|\mathbf{s}) I_S, \text{ where } I_S \text{ is the indicator function.}$$

We wish to calculate the joint density  $(\mathbf{s}, S, \mathbf{a}_S) = (\mathbf{s})(S|\mathbf{s})(\mathbf{a}_S|\mathbf{s}, S)$ .

First,  $(\mathbf{s})$  is given by (AA1). From (AA2) and (AA4),

$$(\mathbf{a}_S | \mathbf{s}_S, S) = \prod_{i \in S} (a_i | s_i) I_S \int \prod_{i \in \setminus S} (a_i | s_i) d\mathbf{a}_{\setminus S}. \quad (\text{A26})$$

The integral (or summation in the discrete case) is 1, and by Bayes' formula,

$$(\mathbf{a}_S | \mathbf{s}, S) = \prod_{i \in S} \{(s_i | a_i) (a_i) / (s_i)\} I_S. \quad (\text{A27})$$

Hence, the combined density is

$$(s, S, a_s) = (S|s) I_S \prod_{i \in S} (s_i) \prod_{i \in S} \{(s_i|a_i) (a_i)\}. \quad (A28)$$

The conclusion is that, apart from a constant term, the log-likelihood is not altered by the subsampling. The distribution  $(s|a)$  is assumed either (i) known or (ii) depend on parameters to be estimated from the likelihood. If  $(s|a)$  is estimated from independent data rather than a subsample, the log-likelihood has essentially the same form.

P	0.1	0.2	0.3	0.4	0.5
$\Delta$					
0.20	0.0036	0.0064	0.0084	0.0095	0.0099
0.25	0.0057	0.0100	0.0130	0.0148	0.0154
0.30	0.0083	0.0144	0.0187	0.0212	0.0220
0.35	0.0113	0.0196	0.0253	0.0286	0.0297
0.40	0.0149	0.0256	0.0329	0.0371	0.0385
0.50	0.0237	0.0400	0.0508	0.0569	0.0589
0.75	0.0559	0.0892	0.1095	0.1207	0.1243
1.00	0.1039	0.1548	0.1838	0.1992	0.2041
1.50	0.2421	0.3174	0.3556	0.3748	0.3807
2.00	0.4113	0.4901	0.5270	0.5450	0.5504
2.50	0.5775	0.6448	0.6748	0.6890	0.6933
3.00	0.7178	0.7680	0.7897	0.7997	0.8028

Table A1. ARE for estimating the proportions in two age classes and known Normal densities. For  $\Delta < 0.2$ , the approximation  $p(I - p) \Delta^2$  may be used.

## 7 TABLES

**Table 1. Comparative estimates of proportions-at-age in 1983 for Greenland turbot data**

Age	Mean length	$\hat{p}_a$ (IALK)	$\hat{p}_a$ (parametric known growth)
4	38.07	0.0353	0.036
5	43.94	0.1903	0.182
6	49.78	0.2281	0.244
7	55.12	0.1291	0.105
8	61.24	0.1125	0.139
9	65.64	0.0380	0.029
10	67.02	0.0525	0.026
11	68.36	0.2142 <sup>1</sup>	0.055
12	71.50		0.000
13	75.00		0.000
14	76.46		0.000
15	78.64		0.000
16	81.02		0.113
17	83.89		0.007
18	83.69		0.059
19	88.15		0.005
20	88.09		0.001

---

<sup>1</sup> Ages 11+ were grouped in Kimura & Chikuni (*op.cit.*)

**Table 2. Precision results for uniform-intensity sampling: n otoliths and N length samples. "p[age]" is proportion at age. "1" and "2" refer to the parametric known-growth and unknown-growth methods respectively. "CV" is the coefficient of variation. "neq" is the size of a randomly-selected sample (random with respect to age and length) that would deliver equal precision to a particular estimator. The ARE columns show the efficiency of the ALK relative to each parametric method.**

**GAB purse seine**

n=100, N=200

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
1	0.0135	61	61	199	199	90	90	0.45	0.45
2	0.1515	17	18	186	179	24	96	0.52	0.54
3	0.6227	6	6	168	154	8	96	0.57	0.62
4	0.1863	17	18	153	140	22	93	0.60	0.66
5	0.0260	51	54	143	130	69	78	0.55	0.60

n=100, N=500

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
1	0.0135	38	38	497	496	46	350	0.71	0.71
2	0.1515	11	12	445	368	14	267	0.60	0.73
3	0.6227	4	5	369	235	6	198	0.54	0.84
4	0.1863	12	15	302	190	16	162	0.54	0.86
5	0.0260	37	48	267	162	53	135	0.51	0.83

n=200, N=500

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
1	0.0135	38	38	498	497	51	283	0.57	0.57
2	0.1515	11	11	459	428	14	274	0.60	0.64
3	0.6227	4	4	404	347	5	254	0.63	0.73
4	0.1863	11	12	357	306	14	236	0.66	0.77
5	0.0260	34	37	329	278	42	211	0.64	0.76

n=100, N=1000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
1	0.0135	27	27	993	992	31	756	0.76	0.76
2	0.1515	8	9	875	628	12	398	0.45	0.63
3	0.6227	3	5	699	294	5	247	0.35	0.84
4	0.1863	9	14	545	217	15	189	0.35	0.87
5	0.0260	28	46	469	180	49	156	0.33	0.87

n=200, N=1000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
1	0.0135	27	27	994	993	31	737	0.74	0.74
2	0.1515	8	9	889	735	10	541	0.61	0.74
3	0.6227	3	4	737	470	4	401	0.54	0.85
4	0.1863	8	11	605	379	12	330	0.55	0.87
5	0.0260	26	34	534	325	36	284	0.53	0.87

Estimation of age profiles of Southern Bluefin Tuna

n=500, N=1000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
1	0.0135	27	27	996	996	39	489	0.49	0.49
2	0.1515	8	8	931	894	11	496	0.53	0.56
3	0.6227	3	3	841	771	3	496	0.59	0.64
4	0.1863	8	8	765	701	9	492	0.64	0.70
5	0.0260	23	24	717	651	28	474	0.66	0.73

n=100, N=2000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
1	0.0135	19	19	1986	1982	22	1473	0.74	0.74
2	0.1515	6	7	1737	1091	11	501	0.29	0.46
3	0.6227	2	4	1359	348	5	275	0.20	0.79
4	0.1863	7	14	1028	236	15	203	0.20	0.86
5	0.0260	21	44	871	191	47	167	0.19	0.88

n=200, N=2000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
1	0.0135	19	19	1987	1983	21	1586	0.80	0.80
2	0.1515	6	7	1751	1255	8	803	0.46	0.64
3	0.6227	2	3	1399	588	3	496	0.35	0.84
4	0.1863	6	10	1090	435	11	381	0.35	0.88
5	0.0260	20	32	938	359	34	320	0.34	0.89

n=500, N=2000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
1	0.0135	19	19	1989	1987	23	1439	0.72	0.72
2	0.1515	6	6	1793	1547	7	1142	0.64	0.74
3	0.6227	2	2	1511	1077	3	904	0.60	0.84
4	0.1863	6	7	1266	893	8	774	0.61	0.87
5	0.0260	18	22	1130	778	23	690	0.61	0.89

**Japanese long-line**

n=100, N=200

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
2	0.0457	33	34	189	183	48	93	0.49	0.50
3	0.1375	19	19	177	166	26	93	0.53	0.56
4	0.1010	25	26	141	130	32	89	0.63	0.69
5	0.1521	21	21	132	124	25	91	0.69	0.73
6	0.1707	20	20	124	118	23	91	0.74	0.77
7	0.1225	25	25	116	113	28	90	0.77	0.80
8	0.0780	32	33	112	110	37	87	0.78	0.79
9	0.0651	36	36	113	111	41	84	0.74	0.76
10	0.0350	50	51	109	107	59	80	0.74	0.75
11	0.0160	77	77	105	104	90	75	0.72	0.73
12	0.0104	96	96	104	103	116	71	0.68	0.69
13	0.0119	89	89	105	104	112	66	0.63	0.64
14	0.0078	111	111	103	103	146	60	0.58	0.58
15	0.0047	144	144	102	102	200	53	0.52	0.52
16	0.0029	184	184	102	101	273	46	0.45	0.46
17	0.0387	42	44	138	130	60	69	0.50	0.53

n=100, N=500

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
2	0.0457	21	23	457	392	27	279	0.61	0.71
3	0.1375	13	15	399	278	17	228	0.57	0.82
4	0.1010	19	23	250	161	25	141	0.56	0.87
5	0.1521	17	19	205	147	20	133	0.65	0.91
6	0.1707	17	19	169	134	20	124	0.73	0.93
7	0.1225	22	24	143	123	25	115	0.80	0.93
8	0.0780	30	32	129	117	33	109	0.85	0.94
9	0.0651	33	35	132	119	36	110	0.83	0.93
10	0.0350	48	50	121	112	52	103	0.85	0.92
11	0.0160	74	76	111	107	80	97	0.87	0.91
12	0.0104	94	95	108	105	101	94	0.87	0.90
13	0.0119	87	88	110	107	94	93	0.85	0.87
14	0.0078	109	110	107	105	120	89	0.83	0.85
15	0.0047	142	143	104	103	159	84	0.80	0.81
16	0.0029	183	183	103	102	209	79	0.76	0.77
17	0.0387	34	40	210	159	44	130	0.62	0.82

Estimation of age profiles of Southern Bluefin Tuna

n=200, N=500

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
2	0.0457	21	22	468	443	28	273	0.58	0.62
3	0.1375	12	13	429	383	16	261	0.61	0.68
4	0.1010	17	18	320	278	20	223	0.70	0.80
5	0.1521	14	15	292	262	16	219	0.75	0.84
6	0.1707	14	14	266	246	15	214	0.80	0.87
7	0.1225	17	18	245	231	19	206	0.84	0.89
8	0.0780	23	23	232	224	24	201	0.86	0.90
9	0.0651	25	25	236	226	27	200	0.85	0.88
10	0.0350	35	36	224	217	38	192	0.86	0.88
11	0.0160	54	54	213	209	58	184	0.87	0.88
12	0.0104	67	68	209	207	73	179	0.85	0.86
13	0.0119	63	63	212	210	69	175	0.82	0.83
14	0.0078	78	78	209	207	88	166	0.80	0.80
15	0.0047	101	102	206	205	117	156	0.76	0.76
16	0.0029	130	130	204	203	154	145	0.71	0.71
17	0.0387	29	30	306	276	35	201	0.66	0.73

n=100, N=1000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
2	0.0457	15	17	902	689	22	445	0.49	0.65
3	0.1375	9	13	762	365	14	311	0.41	0.85
4	0.1010	15	22	418	176	24	158	0.38	0.90
5	0.1521	14	19	304	157	20	146	0.48	0.93
6	0.1707	15	19	221	140	19	133	0.60	0.95
7	0.1225	21	24	169	127	24	121	0.72	0.95
8	0.0780	29	31	144	120	32	115	0.80	0.96
9	0.0651	31	34	146	122	35	116	0.80	0.95
10	0.0350	46	49	130	114	50	109	0.84	0.95
11	0.0160	73	76	116	108	77	103	0.88	0.95
12	0.0104	92	95	112	106	97	100	0.89	0.95
13	0.0119	85	88	115	108	91	101	0.88	0.94
14	0.0078	108	110	109	106	114	98	0.89	0.92
15	0.0047	142	143	105	104	150	94	0.89	0.91
16	0.0029	182	183	104	103	195	91	0.87	0.88
17	0.0387	30	38	284	171	40	155	0.54	0.90

n=200, N=1000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
2	0.0457	15	16	914	784	19	574	0.63	0.73
3	0.1375	9	11	797	557	12	466	0.58	0.84
4	0.1010	13	17	500	323	18	288	0.58	0.89
5	0.1521	12	14	409	294	14	271	0.66	0.92
6	0.1707	12	13	337	268	14	252	0.75	0.94
7	0.1225	16	17	286	246	18	233	0.81	0.95
8	0.0780	21	22	259	234	23	223	0.86	0.96
9	0.0651	23	25	264	237	25	225	0.85	0.95
10	0.0350	34	35	242	224	36	213	0.88	0.95
11	0.0160	53	54	222	213	55	202	0.91	0.95
12	0.0104	66	67	216	210	69	198	0.91	0.94
13	0.0119	61	62	221	213	65	198	0.90	0.93
14	0.0078	77	78	214	210	81	192	0.90	0.91
15	0.0047	101	101	209	207	107	185	0.89	0.89
16	0.0029	129	130	206	205	139	178	0.86	0.87
17	0.0387	24	28	420	317	30	279	0.67	0.88

n=500, N=1000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
2	0.0457	15	15	947	916	21	492	0.52	0.54
3	0.1375	8	9	884	829	11	493	0.56	0.59
4	0.1010	11	12	705	650	14	488	0.69	0.75
5	0.1521	9	9	661	621	11	490	0.74	0.79
6	0.1707	9	9	619	591	10	490	0.79	0.83
7	0.1225	11	11	582	563	12	489	0.84	0.87
8	0.0780	15	15	561	548	16	485	0.87	0.89
9	0.0651	16	16	567	553	17	482	0.85	0.87
10	0.0350	23	23	544	535	24	476	0.87	0.89
11	0.0160	34	34	524	519	36	469	0.90	0.90
12	0.0104	43	43	518	515	45	462	0.89	0.90
13	0.0119	40	40	524	520	43	454	0.87	0.87
14	0.0078	50	50	517	515	54	441	0.85	0.86
15	0.0047	64	64	512	510	71	424	0.83	0.83
16	0.0029	82	82	508	507	92	406	0.80	0.80
17	0.0387	19	20	689	649	23	459	0.67	0.71

Estimation of age profiles of Southern Bluefin Tuna

n=100, N=2000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
2	0.0457	11	13	1793	1221	19	595	0.33	0.49
3	0.1375	7	12	1482	439	13	368	0.25	0.84
4	0.1010	11	22	741	184	23	166	0.22	0.90
5	0.1521	11	19	480	161	19	152	0.32	0.94
6	0.1707	13	18	301	143	19	137	0.45	0.96
7	0.1225	19	24	202	129	24	124	0.61	0.96
8	0.0780	27	31	159	121	32	117	0.74	0.97
9	0.0651	30	34	160	123	35	119	0.75	0.97
10	0.0350	45	49	139	115	50	112	0.80	0.97
11	0.0160	71	75	121	108	76	105	0.87	0.97
12	0.0104	91	95	116	106	96	103	0.89	0.97
13	0.0119	83	88	120	108	89	104	0.87	0.96
14	0.0078	107	110	112	106	112	102	0.91	0.96
15	0.0047	141	143	106	104	146	99	0.93	0.95
16	0.0029	182	183	104	103	189	97	0.93	0.94
17	0.0387	26	37	380	178	38	168	0.44	0.95

n=200, N=2000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
2	0.0457	11	12	1805	1377	15	909	0.50	0.66
3	0.1375	6	9	1523	730	10	631	0.41	0.86
4	0.1010	10	16	837	353	17	319	0.38	0.91
5	0.1521	10	13	609	313	14	294	0.48	0.94
6	0.1707	10	13	441	281	13	268	0.61	0.95
7	0.1225	15	17	338	254	17	244	0.72	0.96
8	0.0780	20	22	287	239	23	232	0.81	0.97
9	0.0651	22	24	292	244	25	236	0.81	0.97
10	0.0350	33	35	260	228	35	221	0.85	0.97
11	0.0160	51	53	233	215	54	209	0.90	0.97
12	0.0104	65	67	224	212	68	205	0.92	0.97
13	0.0119	60	62	230	216	63	208	0.90	0.96
14	0.0078	76	78	218	211	79	203	0.93	0.96
15	0.0047	100	101	211	208	104	197	0.93	0.95
16	0.0029	129	129	208	205	134	192	0.93	0.94
17	0.0387	21	27	568	342	28	321	0.57	0.94

n=500, N=2000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
2	0.0457	11	11	1838	1635	13	1199	0.65	0.73
3	0.1375	6	7	1627	1248	8	1021	0.63	0.82
4	0.1010	9	11	1074	775	11	692	0.64	0.89
5	0.1521	8	9	913	712	9	655	0.72	0.92
6	0.1707	8	9	779	655	9	615	0.79	0.94
7	0.1225	10	11	680	604	11	576	0.85	0.95
8	0.0780	14	14	625	578	15	556	0.89	0.96
9	0.0651	15	16	637	586	16	561	0.88	0.96
10	0.0350	22	22	590	556	23	535	0.91	0.96
11	0.0160	33	34	548	530	35	513	0.94	0.97
12	0.0104	42	43	535	523	43	504	0.94	0.96
13	0.0119	39	40	545	531	40	507	0.93	0.96
14	0.0078	49	49	531	523	51	496	0.94	0.95
15	0.0047	64	64	520	516	66	484	0.93	0.94
16	0.0029	82	82	514	511	85	472	0.92	0.92
17	0.0387	16	18	948	765	19	687	0.72	0.90

**Indonesian long-line**

n=100, N=200

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
8	0.0017	229	233	112	108	735	11	0.10	0.10
9	0.0055	128	130	110	107	264	26	0.24	0.24
10	0.0096	98	99	107	105	147	48	0.45	0.46
11	0.0114	92	92	104	102	115	65	0.63	0.64
12	0.0167	76	76	103	102	88	77	0.75	0.75
13	0.0354	51	52	103	102	57	83	0.81	0.82
14	0.0344	53	53	102	101	57	87	0.85	0.86
15	0.0314	55	55	101	100	59	89	0.88	0.88
16	0.0362	51	52	100	100	54	90	0.89	0.90
17	0.0503	43	43	100	100	46	90	0.90	0.90
18	0.0754	35	35	100	100	37	91	0.90	0.90
19	0.0754	35	35	100	100	37	91	0.90	0.91
20	0.0832	33	33	100	100	35	90	0.90	0.90
21	0.0702	36	36	100	100	38	90	0.90	0.90
22	0.0817	33	34	100	100	35	90	0.90	0.90
23	0.0759	35	35	100	100	37	90	0.89	0.90
24	0.0517	43	43	100	100	45	89	0.89	0.89
25	0.2539	17	17	103	102	18	90	0.87	0.88

Estimation of age profiles of Southern Bluefin Tuna

n=100, N=500

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
8	0.0017	206	227	138	114	416	34	0.25	0.30
9	0.0055	119	126	127	113	171	62	0.48	0.55
10	0.0096	94	98	117	108	112	83	0.70	0.76
11	0.0114	89	91	109	104	98	91	0.83	0.87
12	0.0167	74	76	107	103	79	95	0.89	0.92
13	0.0354	50	51	109	103	53	98	0.90	0.95
14	0.0344	52	53	105	102	54	98	0.93	0.96
15	0.0314	55	55	102	101	56	98	0.95	0.97
16	0.0362	51	51	102	100	52	98	0.96	0.97
17	0.0503	43	43	101	100	44	98	0.97	0.97
18	0.0754	35	35	101	100	35	98	0.97	0.98
19	0.0754	35	35	100	100	35	98	0.97	0.98
20	0.0832	33	33	100	100	34	98	0.97	0.98
21	0.0702	36	36	101	100	37	98	0.97	0.98
22	0.0817	33	33	101	100	34	98	0.97	0.98
23	0.0759	35	35	101	100	35	98	0.97	0.98
24	0.0517	43	43	101	100	43	98	0.97	0.98
25	0.2539	16	17	110	103	17	100	0.91	0.97

n=200, N=500

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
8	0.0017	158	164	234	219	329	54	0.23	0.25
9	0.0055	89	91	228	218	131	105	0.46	0.48
10	0.0096	69	70	218	212	83	150	0.69	0.71
11	0.0114	64	65	210	206	71	172	0.82	0.83
12	0.0167	53	54	207	204	57	184	0.89	0.90
13	0.0354	36	36	209	205	38	190	0.91	0.93
14	0.0344	37	37	205	202	38	192	0.94	0.95
15	0.0314	39	39	202	201	40	193	0.95	0.96
16	0.0362	36	36	201	201	37	193	0.96	0.96
17	0.0503	31	31	201	200	31	194	0.96	0.97
18	0.0754	25	25	201	200	25	194	0.96	0.97
19	0.0754	25	25	200	200	25	194	0.97	0.97
20	0.0832	23	23	200	200	24	193	0.97	0.97
21	0.0702	26	26	200	200	26	193	0.96	0.97
22	0.0817	24	24	201	200	24	193	0.96	0.96
23	0.0759	25	25	201	200	25	193	0.96	0.96
24	0.0517	30	30	201	200	31	193	0.96	0.96
25	0.2539	12	12	209	205	12	196	0.93	0.96

Estimation of age profiles of Southern Bluefin Tuna

n=100, N=1000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
8	0.0017	186	225	169	116	325	56	0.33	0.48
9	0.0055	113	125	141	115	148	83	0.59	0.72
10	0.0096	91	97	126	110	104	95	0.76	0.87
11	0.0114	87	91	114	105	94	98	0.86	0.93
12	0.0167	73	75	111	103	77	100	0.90	0.96
13	0.0354	49	51	114	104	52	101	0.89	0.97
14	0.0344	51	52	108	102	53	100	0.93	0.98
15	0.0314	54	55	104	101	56	99	0.95	0.99
16	0.0362	51	51	103	101	52	99	0.97	0.99
17	0.0503	43	43	102	100	44	99	0.97	0.99
18	0.0754	35	35	102	100	35	99	0.98	0.99
19	0.0754	35	35	101	100	35	99	0.98	0.99
20	0.0832	33	33	101	100	33	99	0.98	0.99
21	0.0702	36	36	101	100	37	99	0.98	0.99
22	0.0817	33	34	101	100	34	99	0.98	0.99
23	0.0759	35	35	102	100	35	99	0.97	1.00
24	0.0517	42	43	102	100	43	99	0.98	1.00
25	0.2539	16	17	116	103	17	102	0.88	1.00

n=200, N=1000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
8	0.0017	146	161	276	227	238	104	0.38	0.46
9	0.0055	84	89	255	226	107	158	0.62	0.70
10	0.0096	66	69	234	217	74	186	0.79	0.86
11	0.0114	63	65	218	208	67	194	0.89	0.93
12	0.0167	52	53	214	206	55	197	0.92	0.96
13	0.0354	35	36	218	206	37	201	0.92	0.97
14	0.0344	37	37	210	203	38	199	0.95	0.98
15	0.0314	39	39	205	202	39	198	0.97	0.98
16	0.0362	36	36	203	201	37	198	0.98	0.99
17	0.0503	31	31	202	201	31	198	0.98	0.99
18	0.0754	25	25	202	200	25	198	0.98	0.99
19	0.0754	25	25	201	200	25	198	0.98	0.99
20	0.0832	23	23	201	200	24	198	0.99	0.99
21	0.0702	26	26	201	200	26	198	0.98	0.99
22	0.0817	24	24	202	200	24	198	0.98	0.99
23	0.0759	25	25	203	200	25	198	0.98	0.99
24	0.0517	30	30	202	200	30	198	0.98	0.99
25	0.2539	12	12	220	206	12	203	0.93	0.99

Estimation of age profiles of Southern Bluefin Tuna

n=500, N=1000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
8	0.0017	102	104	560	539	176	190	0.34	0.35
9	0.0055	57	58	552	536	75	319	0.58	0.59
10	0.0096	44	44	533	523	50	411	0.77	0.78
11	0.0114	41	41	518	512	44	452	0.87	0.88
12	0.0167	34	34	513	508	35	472	0.92	0.93
13	0.0354	23	23	516	510	24	481	0.93	0.94
14	0.0344	24	24	508	505	24	486	0.96	0.96
15	0.0314	25	25	504	502	25	488	0.97	0.97
16	0.0362	23	23	502	501	23	489	0.97	0.98
17	0.0503	19	19	502	501	20	490	0.98	0.98
18	0.0754	16	16	501	501	16	490	0.98	0.98
19	0.0754	16	16	501	500	16	490	0.98	0.98
20	0.0832	15	15	501	500	15	490	0.98	0.98
21	0.0702	16	16	501	500	16	490	0.98	0.98
22	0.0817	15	15	502	501	15	489	0.98	0.98
23	0.0759	16	16	502	501	16	489	0.97	0.98
24	0.0517	19	19	501	500	19	489	0.98	0.98
25	0.2539	8	8	516	509	8	490	0.95	0.96

n=100, N=2000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
8	0.0017	165	224	216	117	277	77	0.35	0.65
9	0.0055	108	125	155	117	137	97	0.63	0.83
10	0.0096	88	96	135	111	101	102	0.76	0.92
11	0.0114	86	91	118	105	93	101	0.86	0.96
12	0.0167	72	75	114	104	76	101	0.89	0.98
13	0.0354	48	51	119	104	52	103	0.86	0.99
14	0.0344	50	52	111	102	53	101	0.91	0.99
15	0.0314	54	55	106	101	55	100	0.95	0.99
16	0.0362	51	51	104	101	52	100	0.96	0.99
17	0.0503	43	43	104	100	43	100	0.96	1.00
18	0.0754	35	35	103	100	35	100	0.97	1.00
19	0.0754	35	35	102	100	35	100	0.98	1.00
20	0.0832	33	33	101	100	33	100	0.99	1.00
21	0.0702	36	37	101	99	36	100	0.99	1.01
22	0.0817	33	34	102	99	34	100	0.98	1.01
23	0.0759	34	35	103	99	35	100	0.98	1.02
24	0.0517	42	43	102	98	43	100	0.98	1.02
25	0.2539	15	17	124	102	17	103	0.83	1.02

Estimation of age profiles of Southern Bluefin Tuna

n=200, N=2000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
8	0.0017	132	159	339	232	198	149	0.44	0.64
9	0.0055	80	89	282	231	97	191	0.68	0.83
10	0.0096	64	69	251	220	71	203	0.81	0.92
11	0.0114	62	64	228	210	66	202	0.89	0.96
12	0.0167	52	53	221	207	54	202	0.91	0.98
13	0.0354	35	36	228	207	37	204	0.90	0.99
14	0.0344	36	37	216	204	37	202	0.94	0.99
15	0.0314	38	39	208	202	39	200	0.96	0.99
16	0.0362	36	36	206	201	37	200	0.97	0.99
17	0.0503	30	31	204	201	31	200	0.98	0.99
18	0.0754	25	25	203	200	25	200	0.98	1.00
19	0.0754	25	25	202	200	25	199	0.99	1.00
20	0.0832	23	23	201	200	24	199	0.99	1.00
21	0.0702	26	26	202	200	26	200	0.99	1.00
22	0.0817	24	24	203	200	24	200	0.99	1.00
23	0.0759	24	25	204	200	25	200	0.98	1.00
24	0.0517	30	30	203	199	30	200	0.98	1.00
25	0.2539	11	12	233	206	12	206	0.89	1.00

n=500, N=2000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
8	0.0017	95	102	652	563	131	343	0.53	0.61
9	0.0055	54	57	615	560	63	452	0.73	0.81
10	0.0096	42	44	572	538	46	491	0.86	0.91
11	0.0114	40	41	539	519	42	497	0.92	0.96
12	0.0167	33	34	530	513	34	501	0.95	0.98
13	0.0354	23	23	537	515	23	506	0.94	0.98
14	0.0344	23	24	520	508	24	502	0.97	0.99
15	0.0314	25	25	510	504	25	499	0.98	0.99
16	0.0362	23	23	506	502	23	499	0.98	0.99
17	0.0503	19	19	504	501	19	498	0.99	0.99
18	0.0754	16	16	503	501	16	498	0.99	0.99
19	0.0754	16	16	502	500	16	498	0.99	0.99
20	0.0832	15	15	502	501	15	498	0.99	0.99
21	0.0702	16	16	502	501	16	498	0.99	0.99
22	0.0817	15	15	504	501	15	499	0.99	0.99
23	0.0759	16	16	505	501	16	499	0.99	1.00
24	0.0517	19	19	504	501	19	498	0.99	1.00
25	0.2539	7	8	540	514	8	510	0.94	0.99

**Table 3. Summary of CVs in Table 2 for the parametric unknown-growth estimator and the Japanese fishery**

Age 3

n\N	200	500	1000	2000
100	19	15	13	12
200	.	13	11	9
500	.	.	9	7

Age 6

n\N	200	500	1000	2000
100	20	19	19	18
200	.	14	13	13
500	.	.	9	9

Age 9

n\N	200	500	1000	2000
100	36	35	34	34
200	.	25	25	24
500	.	.	16	16

**Table 4. Precision results for unbalanced subsampling (relatively more big fish): Japanese long-lines. Column headings are as for Table 2.**

n=100, N=200

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
2	0.0457	44	49	108	87	66	48	0.44	0.55
3	0.1375	25	33	97	56	40	39	0.40	0.68
4	0.1010	38	49	62	37	62	23	0.37	0.62
5	0.1521	24	27	94	78	39	38	0.40	0.48
6	0.1707	18	19	144	134	26	73	0.51	0.55
7	0.1225	21	21	169	164	25	116	0.69	0.71
8	0.0780	26	26	174	172	30	129	0.74	0.75
9	0.0651	28	29	178	175	34	127	0.71	0.72
10	0.0350	40	40	174	173	47	125	0.71	0.72
11	0.0160	60	60	171	170	71	121	0.71	0.71
12	0.0104	75	75	169	169	91	114	0.67	0.68
13	0.0119	70	70	171	170	88	106	0.62	0.62
14	0.0078	87	87	169	169	115	97	0.57	0.57
15	0.0047	112	112	168	168	157	86	0.51	0.51
16	0.0029	143	143	168	167	213	76	0.45	0.45
17	0.0387	35	35	204	198	51	95	0.47	0.48

n=100, N=500

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
2	0.0457	24	29	376	247	47	95	0.25	0.38
3	0.1375	14	28	315	78	32	60	0.19	0.76
4	0.1010	23	45	167	43	56	28	0.17	0.66
5	0.1521	18	25	166	86	34	47	0.28	0.55
6	0.1707	16	18	193	148	23	96	0.50	0.65
7	0.1225	19	20	204	181	21	157	0.77	0.87
8	0.0780	24	25	199	186	26	172	0.86	0.93
9	0.0651	26	27	205	191	29	176	0.86	0.92
10	0.0350	38	39	192	183	41	167	0.87	0.92
11	0.0160	59	59	180	175	62	159	0.89	0.91
12	0.0104	74	74	176	173	78	155	0.88	0.89
13	0.0119	68	69	179	175	74	153	0.85	0.87
14	0.0078	85	86	175	173	93	146	0.84	0.85
15	0.0047	111	111	172	171	124	138	0.80	0.81
16	0.0029	142	143	170	169	163	130	0.76	0.77
17	0.0387	29	32	292	245	36	196	0.67	0.80

Estimation of age profiles of Southern Bluefin Tuna

n=200, N=500

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
2	0.0457	26	30	306	232	39	137	0.45	0.59
3	0.1375	15	22	267	126	25	98	0.37	0.78
4	0.1010	24	34	161	78	41	53	0.33	0.68
5	0.1521	16	19	214	161	25	87	0.41	0.54
6	0.1707	13	13	308	276	17	170	0.55	0.61
7	0.1225	14	15	354	338	16	273	0.77	0.81
8	0.0780	18	18	361	352	20	305	0.84	0.86
9	0.0651	20	20	368	360	22	306	0.83	0.85
10	0.0350	28	28	357	351	30	302	0.85	0.86
11	0.0160	42	42	345	343	46	297	0.86	0.87
12	0.0104	53	53	342	340	57	290	0.85	0.85
13	0.0119	49	49	345	343	54	281	0.81	0.82
14	0.0078	61	61	342	340	69	270	0.79	0.79
15	0.0047	79	79	339	338	91	255	0.75	0.75
16	0.0029	101	101	337	336	120	238	0.71	0.71
17	0.0387	24	24	443	421	30	277	0.63	0.66

n=100, N=1000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
2	0.0457	16	21	822	489	42	116	0.14	0.24
3	0.1375	10	27	675	89	31	66	0.10	0.75
4	0.1010	16	44	333	46	55	30	0.09	0.65
5	0.1521	14	25	268	88	34	50	0.18	0.56
6	0.1707	14	18	250	153	22	102	0.41	0.66
7	0.1225	17	20	236	188	21	167	0.71	0.89
8	0.0780	23	25	220	193	25	184	0.83	0.95
9	0.0651	25	27	226	198	28	189	0.84	0.95
10	0.0350	37	38	206	187	39	179	0.87	0.95
11	0.0160	57	59	187	177	60	169	0.90	0.95
12	0.0104	72	74	182	175	76	165	0.91	0.95
13	0.0119	67	68	185	178	71	166	0.90	0.94
14	0.0078	84	85	178	174	89	161	0.90	0.92
15	0.0047	110	111	174	172	117	155	0.89	0.91
16	0.0029	142	142	171	170	152	150	0.87	0.88
17	0.0387	25	30	388	272	32	244	0.63	0.90

n=200, N=1000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
2	0.0457	17	21	752	495	32	202	0.27	0.41
3	0.1375	10	20	629	157	23	124	0.20	0.79
4	0.1010	16	32	334	87	39	58	0.17	0.67
5	0.1521	13	18	332	172	24	97	0.29	0.56
6	0.1707	11	13	385	295	16	195	0.51	0.66
7	0.1225	13	14	408	362	15	318	0.78	0.88
8	0.0780	17	18	399	372	18	352	0.88	0.94
9	0.0651	19	19	409	382	20	360	0.88	0.94
10	0.0350	27	27	384	365	28	346	0.90	0.95
11	0.0160	41	42	359	350	43	331	0.92	0.95
12	0.0104	52	52	352	346	54	325	0.92	0.94
13	0.0119	48	49	358	350	51	325	0.91	0.93
14	0.0078	60	61	350	346	64	315	0.90	0.91
15	0.0047	79	79	344	341	83	305	0.89	0.89
16	0.0029	101	101	340	338	108	294	0.86	0.87
17	0.0387	21	23	583	490	24	419	0.72	0.86

n=500, N=1000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
2	0.0457	20	22	541	436	27	296	0.55	0.68
3	0.1375	11	15	484	282	17	224	0.46	0.79
4	0.1010	17	22	312	186	26	131	0.42	0.71
5	0.1521	11	12	468	389	16	211	0.45	0.54
6	0.1707	8	9	718	669	11	398	0.55	0.60
7	0.1225	9	9	844	821	11	631	0.75	0.77
8	0.0780	12	12	872	860	13	715	0.82	0.83
9	0.0651	13	13	888	876	14	715	0.81	0.82
10	0.0350	18	18	871	864	19	732	0.84	0.85
11	0.0160	27	27	853	849	29	746	0.87	0.88
12	0.0104	34	34	847	845	36	741	0.87	0.88
13	0.0119	31	31	854	850	34	720	0.84	0.85
14	0.0078	39	39	847	845	42	708	0.83	0.84
15	0.0047	50	50	842	840	55	688	0.82	0.82
16	0.0029	64	64	838	837	72	662	0.79	0.79
17	0.0387	16	16	1020	991	20	598	0.59	0.60

Estimation of age profiles of Southern Bluefin Tuna

n=100, N=2000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
2	0.0457	11	15	1712	936	40	128	0.07	0.14
3	0.1375	7	26	1392	95	30	70	0.05	0.73
4	0.1010	12	44	654	46	54	30	0.05	0.66
5	0.1521	11	25	452	87	33	51	0.11	0.58
6	0.1707	12	18	341	154	22	104	0.31	0.67
7	0.1225	16	19	278	192	20	172	0.62	0.90
8	0.0780	22	25	243	196	25	189	0.78	0.96
9	0.0651	24	27	248	202	27	196	0.79	0.97
10	0.0350	35	38	220	190	39	184	0.84	0.97
11	0.0160	56	59	195	179	59	174	0.89	0.97
12	0.0104	71	74	188	176	75	170	0.91	0.97
13	0.0119	66	68	193	179	69	173	0.89	0.96
14	0.0078	84	85	182	175	87	168	0.92	0.96
15	0.0047	110	111	175	172	114	164	0.94	0.95
16	0.0029	141	142	172	170	147	160	0.93	0.94
17	0.0387	22	29	517	288	30	272	0.53	0.94

n=200, N=2000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
2	0.0457	11	15	1643	977	30	240	0.15	0.25
3	0.1375	7	19	1349	177	22	135	0.10	0.76
4	0.1010	12	31	666	91	38	60	0.09	0.66
5	0.1521	10	18	537	176	24	100	0.19	0.57
6	0.1707	10	13	500	305	15	205	0.41	0.67
7	0.1225	12	14	473	376	15	337	0.71	0.90
8	0.0780	16	18	440	385	18	371	0.84	0.96
9	0.0651	18	19	453	396	19	383	0.85	0.97
10	0.0350	26	27	412	374	28	363	0.88	0.97
11	0.0160	41	42	375	354	42	345	0.92	0.97
12	0.0104	51	52	363	349	53	339	0.93	0.97
13	0.0119	47	48	371	355	49	342	0.92	0.96
14	0.0078	60	60	357	349	62	334	0.94	0.96
15	0.0047	78	79	347	343	81	326	0.94	0.95
16	0.0029	100	101	343	340	104	318	0.93	0.94
17	0.0387	18	21	777	543	22	506	0.65	0.93

n=500, N=2000

age	p[age]	CV1	CV2	neq1	neq2	CV (ALK)	neq (ALK)	ARE1	ARE2
2	0.0457	12	15	1434	982	21	483	0.34	0.49
3	0.1375	7	13	1211	371	14	302	0.25	0.81
4	0.1010	12	21	665	211	25	146	0.22	0.69
5	0.1521	9	11	717	424	15	240	0.33	0.57
6	0.1707	7	8	894	725	10	479	0.54	0.66
7	0.1225	9	9	974	888	10	780	0.80	0.88
8	0.0780	11	11	966	917	12	866	0.90	0.94
9	0.0651	12	12	989	939	13	887	0.90	0.94
10	0.0350	17	17	937	904	18	862	0.92	0.95
11	0.0160	26	27	887	870	27	836	0.94	0.96
12	0.0104	33	33	872	860	34	826	0.95	0.96
13	0.0119	31	31	884	871	32	827	0.94	0.95
14	0.0078	38	38	868	860	40	813	0.94	0.94
15	0.0047	50	50	855	850	52	795	0.93	0.94
16	0.0029	64	64	848	844	66	778	0.92	0.92
17	0.0387	14	15	1332	1174	16	1002	0.75	0.85