Modeling Age-at-Length and Length-at-Age Data Incorporating Multiple Years

3 authors:

Quang C. Huynh
Blue Matter Science Ltd
14 PUBLICATIONS **65** CITATIONS

SEE PROFILE

John Hoenig
Virginia Institute of Marine Science
169 PUBLICATIONS **7,841** CITATIONS

SEE PROFILE

Phil Sadler
Virginia Institute of Marine Science
6 PUBLICATIONS **48** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Mathematics of fish tagging experiments View project

Growth assessment of tropical tunas with AOTTP data View project

A general theory of age-length keys: combining the forward and inverse keys to estimate age composition from incomplete data

Lisa E. Ailloud [ORCID] [1,*] and John M. Hoenig[2]

[1]International Commission for the Conservation of Atlantic Tunas, AOTTP, Corazón de María, 8, 28002 Madrid, Spain
[2]Virginia Institute of Marine Science, William & Mary, P.O. Box 1346, Gloucester Point, VA 23062, USA

*Corresponding author: tel: +34 914165600; e-mail: lisa.ailloud@iccat.int

There are two approaches to estimating age composition from a large number of length observations and a limited number of age determinations: the forward and the inverse age-length keys. The forward key looks at the distribution of age within each length bin while the inverse key looks at the distribution of length at each age. The former is more precise but has stringent requirements for the way data are collected. The latter approach is more widely applicable. We review the theory of the two keys with particular attention to necessary assumptions and the restrictions on when the methods are applicable. We show it is possible to combine the two approaches into a combined forward-inverse age-length key. This approach can be used to estimate age composition in several years simultaneously. It takes advantage of the efficiency of the forward key in years when that is appropriate, applies the inverse key to years with no age data, and uses a blending of the two approaches for years with moderate amounts of age data.

Keywords: age composition, age-length keys, sparse data, stock assessment.

Introduction

Age-structured stock assessment models rely on estimates of the age composition of the catch as their primary input. Age composition, simply put, describes the proportion of a population belonging to each age class. Estimates of age composition can be obtained from fisheries-dependent or fisheries independent data. They are key to understanding demographic variations in recruitment, growth, mortality, and the reproductive potential of a stock. Observed changes in age composition through time can also provide insight on how a stock is responding to exploitation and what capacity it has to withstand and recover from external perturbations (Jennings et al., 1998; Greenstreet et al., 1999; Rouyer et al., 2011; Durant et al., 2013). Having a good understanding of spawner age composition can also help managers gauge how a stock may be responding to management actions and rebuilding programmes (Marteinsdottir and Thorarinsson, 1998; Hixon et al., 2014).

Age composition can be estimated by ageing a simple random sample of the population and taking the resulting proportions at age of the sample as an estimate of the age composition for the sampled population. Ages are commonly estimated by counting growth rings deposited on an annual basis in the otoliths, scales, fin rays or spines of bony fishes (Quist et al., 2012) or the vertebrae, or spines of cartilaginous fishes (Cailliet and Goldman, 2004). Obtaining a reliable sample of ages from which to estimate age composition requires all age groups to be well represented in the sample. This involves sampling a very large number of fish because older animals tend to be much less abundant in the catch than younger ones. Yet, determining ages is costly and time consuming, so obtaining a simple random sample of the population is not a realistic goal for most stocks.

A more cost-efficient way of sampling the fishery to obtain estimates of age composition is through double sampling (Fridriksson, 1934; Tanaka, 1953). With double sampling, an

estimate of the true classifier is obtained by utilizing its relationship with a covariate that is less reliable but easily obtained (Tenenbein, 1972). For fish, the reading of ages is labour intensive, but lengths are easy to measure and correlated with ages, so the double sampling technique proves useful. In the first stage, length information is collected on a large random sample obtained from the population of interest. In a second stage, ages are recorded on a much smaller subsample of fish randomly selected through length-stratified sampling. This is the concept on which the theory of age-length keys (ALK) was first developed (Fridriksson, 1934).

In the literature, the term "age-length key" has come to refer to one specific type of ALK: the forward or "classic" ALK. However, there are three main types of ALK that can be used to estimate age composition:

(1) *Forward keys*—which describe the probability of age given size (Fridriksson, 1934)

(2) *Inverse keys*—which describe the probability of size given age (Clark, 1981; Bartoo and Parker, 1983; Hoenig and Heisey, 1987; Kimura and Chikuni, 1987)

(3) *Combined forward-inverse age-length (FIAL) keys*—which couple both the concepts of forward and inverse keys into one using a maximum likelihood framework (Hoenig et al., 2002)

While applications of the forward key are common in the peer-reviewed literature, the inverse key has only occasionally been applied. Haeseker et al. (2003) use Hoenig and Heisey's (1987) inverse key approach to estimate age composition in sea lampreys, and Murta and Vendrell (2009) use the Kimura and Chikuni (1987) inverse key approach to age fish eggs. The combined FIAL key has been used tangentially to estimate disease prevalence (Pestal et al., 2003) from error-prone tests but its application to estimating age composition has, to our knowledge, never been documented in the peer-reviewed literature. An application of the FIAL key to western Atlantic bluefin tuna is presented in Ailloud et al. (2019). In a grey literature document, Murta et al. (2016) test the relative performance of the forward key, inverse keys, and the FIAL key to demonstrate the use of the ALKr package in R. While this package contains useful functions for implementing the forward key (Fridriksson, 1934) and different versions of the inverse key (Clark, 1981; Bartoo and Parker, 1983; Hoenig and Heisey, 1987; Kimura and Chikuni, 1987; Gascuel, 1994), the implementation of the FIAL key is restrictive as it only returns an ALK for years in which no age data have been collected. It does not estimate an ALK for years in which age-length samples are incomplete such that some length groups do not have any age observations. Computer code in AD Model Builder (ADMB, Fournier et al., 2012), able to accommodate sparse data (i.e. when length samples are representative but age samples are small or missing in many years, and some length groups do not have any age observations) can be found in the Supplementary data section of this article.

Technical reports and grey literature indicate that the use of the forward and inverse keys is widespread in stock assessment. Unfortunately, those reports also show that the assumptions behind each method are poorly described. Practitioners are commonly found violating the basic assumption of forward keys when they borrow keys from adjacent years to estimate age composition in years for which age data are unavailable or pool data over multiple years

to increase sample sizes. Though some authors admit their lack of rigor in doing so, they often improperly justify the procedure by stating that growth is not likely to have changed significantly between the years (or areas) of collection. Yet, since forward keys describe the distribution of age at size, changes in growth are not the primary concern in this case; what is of concern are differences in age structure between years (or areas) due to changes in survival and recruitment (Kimura, 1977). Similarly, reports often develop or express the desire to develop separate keys by gear due to differences in size selectivity (e.g. ASMFC, 2010; Wyanski et al., 2000); yet, size selectivity does not, in fact, preclude a forward key developed from one gear being applied to a different gear as long as, within a length category, the fish available to each gear are from the same population (Westrheim and Ricker, 1978).

When the forward and inverse key methods are tried, and neither method is found to be satisfactory, the FIAL key has seldom been sought as a potential solution, probably because of a lack of familiarity with the approach and (until now) lack of software for implementing the method. In Carpi et al. (2015), the authors elect to borrow a forward key built in one year to estimate age composition in adjacent years after discovering that the Kimura and Chikuni (1987) inverse key approach did not perform well on their dataset. This is a clear indication that a more fisheries-oriented description of the FIAL key is needed.

Any errors present in the estimates of age composition are bound to propagate through an assessment, ultimately affecting the evaluation of stock status and management advice. It is therefore important that practitioners understand the advantages and limitations of each method, and any restrictions to their use. Our objective is therefore to describe how each method is derived and what the underlying assumptions are. We will clarify the strengths and weaknesses of each method in estimating age composition, with particular emphasis on the utility of the FIAL key, which can prove useful when age data are incomplete and the inverse key is producing dubious results.

Data requirements for constructing age-length keys

To construct ALK, at least two samples must be obtained from the population of interest. The population may be all fish that are landed, or all fish in the water depending on what are the research questions. In the simplest case, a large sample of N fish is obtained for which the lengths have been measured (we will term this the *length frequency* sample) and a smaller sample of n fish is also obtained for which lengths j ($j = 1, 2, 3, \ldots, J$) and ages i ($i = 1, 2, 3, \ldots, I$) have been determined (we will term this the *age-length* sample). The age-length sample is generally collected through a length-stratified random sampling design (using pre-specified length bins) from the length-frequency sample, which can be done in one of two ways:

(1) Using "fixed" subsampling, where a fixed number of fish is selected to be aged for each length bin

(2) Using "proportional" subsampling, where the number of fish selected for ageing for each length bin is proportional to the sample size of fish belonging to that length bin.

Fixed subsampling is often done by quota sampling where fish skeletal parts are collected until, say, ten fish have been sampled from each length bin. (This has obvious problems if age

composition varies over time or space and sampling ends when the last quota is met.) Proportional subsampling is often accomplished by using systematic sampling, i.e. every *m*th fish is sampled for skeletal parts. Kimura (1977) compared proportional subsampling to fixed subsampling, using total variance (sum over age of estimated variances, Vartot; Kimura, 1977) of the estimated proportions at age as a measure of overall precision, and concluded that the proportional subsampling was statistically more efficient. However, fixed subsampling was strictly defined as obtaining the same number of samples from each length bin. Yet, that need not be the case. The number of samples obtained from each length bin can be optimized according to a priori knowledge of the number of age groups found in each length bin. Length bins known to have a single age group dominating the catch do not require large sample sizes to obtain precise estimates of age composition, while length bins that comprise a mixture of many ages will benefit from larger sample sizes to lower the variance in the estimated proportions at age. [Technically, the variance of a proportion p, is $p(1 − p)/n$ where n is the sample size; this variance has a maximum at $p = 0.5$, and declines to 0 as p approaches 0 or 1. Hence, a length group dominated by a single age will have p close to 1 and, thus, small variance, even at small sample sizes.] Used as such, fixed subsampling can be made more efficient than proportional subsampling [see Lai (1993) "length-based age subsampling"]. If one has past age-length samples and one assumes that average age composition is similar between years, one can use a priori knowledge of the number of age groups found in each length bin and the variance formula above to calculate an optimal allocation of sample sizes per length bin needed to maximize precision given time and cost limitations (Lai, 1993).

The notation used in this article is summarized in Table 1. For any one year, the age-length sample can be summarized in a two-way contingency table where the age categories form the rows, and the length categories form the columns (Table 2). The cell counts, n_{ij}, correspond to the number of fish in the sample that fall within age class i and length bin j. The expressions $n_{i.}$ and $n_{.j}$ correspond to the total sample sizes of fish by age class (collapsed over all length bins) and length bin (collapsed over all ages), respectively. Here, the $n_{i.}$ are random while the $n_{.j}$ are treated as fixed. The total sample size of the age-length sample is denoted by n.

For any one year, the length frequency sample can be summarized in a vector Y of length J (Table 3). The vector entries, y_j, correspond to the sample sizes of fish belonging to each length bin j for that year, for $j = 1, 2, \ldots, J$. The total sample size of the length frequency sample is denoted by N.

A third type of sample can be used with the FIAL key. This sample, which is primarily of theoretical interest, is a random subsample of the population for which only age information is available. We will term this sample the *age only* sample and represent it using a vector X of length I (Table 4). The vector entries, x_i, correspond to the sample sizes of fish belonging to each age class i. The total sample size of the age only sample is denoted by M.

The above defined notation is used throughout.

The forward age-length key
Methodology
The forward ALK was first developed by Fridriksson (1934). The method works on the premise that given a random sample of N fish for which only lengths have been measured and a subsample

Table 1. Summary of notations used.

	Symbol	Value
Age class	i	$1, 2, \ldots, I$
Length bin	j	$1, 2, \ldots, J$
Year	k	$1, 2, \ldots, K$
Number of fish of age i and length bin j in the age-length sample	n_{ij}	
Number of fish of length j in the length frequency sample	y_j	
Number of fish of age i in the age only sample	x_i	
Total size of the age-length sample	n	
Total size of the length frequency sample	N	
Total size of the age only sample	M	
Total number of fish belonging to the ith age class of the age-length sample	$n_{i.}$	
Total number of fish belonging to the jth length bin of the age-length sample	$n_{.j}$	

of n fish whose lengths and ages have been measured, the probability $P(i|j)$ that a fish is age i given that it belongs to length bin j is the same for both samples. This probability can be estimated from the age-length sample as:

$$\hat{P}(i|j) = \hat{q}_{ij} = {n_{ij}}/{n_{.j}} \tag{1}$$

where $\hat{q}_{ij}$ are the estimated probabilities of age given length that populate the cells of the forward ALK. All other notations are defined in Table 1.

The probabilities of age given length from the forward ALK are then simply multiplied by the marginal probabilities $\hat{P}(j) = {y_j}/{N}$ to obtain an estimate of age composition from the forward key, $\hat{A}$. This can be expressed using matrix algebra as follows:

$$\hat{A} = \mathbf{Q}Y/N \tag{2}$$

where $\mathbf{Q}$ is the I by J matrix with elements $\hat{q}_{ij}$. In the equations above, the sample n may be obtained using simple random sampling or length-stratified random sampling. Equation (2) can be shown to give maximum likelihood estimates; it is presented in the form above to emphasize the logic of the approach.

Assumptions, applications, and limitations
Forward keys require that representative age-length and length frequency samples be collected on a yearly basis (or seasonally if resources allow). A key constructed from one year of data cannot be applied to a different year's catch because the population composition changes from year to year. The probability of age given size, $P(i|j)$, is affected by the relative proportion of each age class in the population as a whole, which fluctuates from year to year with changes in recruitment, age-specific mortality rates, and growth. As Kimura (1977) and Westrheim and Ricker (1978) noted, the forward ALK tends to preserve the age composition of the population from which it was derived. Ignoring these guidelines and applying a single forward ALK to multiple years of length frequency data, or pooling several years of age-length data to construct a single forward ALK, can seriously underestimate the variance in estimated proportions at age and result in severe bias (Aanes and Vølstad, 2015).

Table 2. An illustration of the age-length sample for any one year.

		Length (j)				
		1	2	...	J	Totals by age
Age (i)	1	n_{11}	n_{12}	...	n_{1J}	$n_{1.}$
	2	n_{21}	n_{22}	...	n_{2J}	$n_{2.}$
	...	...	...	...	...	...
	I	n_{I1}	n_{I2}	...	n_{IJ}	$n_{I.}$
Totals by length		$n_{.1}$	$n_{.2}$	...	$n_{.J}$	n

Each cell contains the sample size n_{ij} of fish of age i belonging to length bin j. Row sums, the total number of samples in each age class, are denoted by $n_{i.}$. Column sums, the total number of samples in each length bin, are denoted by $n_{.j}$. The total size of the age-length sample is denoted by n.

Table 3. An illustration of the length frequency sample for any one year.

		Length (j)		
1	2	...	J	Total
y_1	y_2	...	y_J	N

The entries, y_j, correspond to the sample sizes of fish belonging to each length bin j for that year. The total sample size of the length only sample is denoted by N.

Table 4. An illustration of the age only sample, X, for any one year.

		Age (i)		
1	2	...	I	Total
x_1	x_2	...	x_I	M

The vector entries, x_i, correspond to the sample sizes of fish belonging to each age class i for that year. The total sample size of the age only sample is denoted by M.

that was developed from one gear and apply it to a different gear, even if the two gears have different selectivity patterns, so long as the two selectivity curves within a length bin are parallel. With narrow length bins, selectivity is almost constant, hence, the requirement of parallel selectivity curves is met.

We have seen that if age-length data are missing in certain years, the forward key method will not allow age composition to be estimated for those years. But an additional issue arises when samples are not being collected following a thorough sampling protocol because gaps in data within a year can still preclude the forward key from being used. Thus, if a length bin has not been sampled for age, one will not be able to assign ages to the portion of the catch corresponding to that length bin. For example, with Atlantic Bluefin tuna, length composition data are collected routinely but spines and otoliths are collected opportunistically so there are gaps in the size coverage. More generally, in multifleet fisheries, some fleets may be hard to sample resulting in gaps in coverage. This is where the inverse key and the FIAL key are advantageous.

The question of "what is the optimal number of age and length samples needed to construct a reliable forward key?" has been explored by several authors. Tanaka (1953) looked at the sample sizes needed to reach a given level of precision in the estimates of proportions at age. Lai (1993) derived a length-based optimal sampling design for forward keys for both the fixed and proportional sampling schemes, given a total allowable cost or desired level of precision. Oeberst (2000) developed a universal cost function to determine the size and structure of the sample required to reach a certain level of precision using the cost ratio of age determinations to length measurements. Coggins et al. (2013) simulated stocks with varying life history traits and exploitation histories to evaluate the sample sizes needed to estimate von Bertalanffy growth parameters and the instantaneous rate of total mortality from age composition estimated using forward ALKs. However, these studies have all assumed random sampling when, in reality, one almost always obtains samples through cluster sampling (Chih, 2010). With cluster sampling, sampling efficiency is low because the non-independence of fish sampled from the same cluster lowers the effective sample size. As such, sample size calculations based on random sampling are too optimistic for most real-world applications, with much bigger samples needed than indicated. Alternatively, one could incorporate effective sample sizes into sampling designs as proposed by Chih (2010).

While small changes in growth or survival are not likely to significantly affect the construction of the ALK, variable year class strength is of major concern (Westrheim and Ricker, 1978). Consider a simple example. Say the first year of a study coincides with a very good recruitment year such that 75% of the fish found in the first length bin are of age 0, $P(i = 0 | j = 1) = 0.75$, and 25% are age 1, $P(i = 1 | j = 1) = 0.25$. Now imagine that in the next year, the population experiences a complete failure in recruitment and no fish of age zero are observed in length bin 1. In that case, $P(i = 0 | j = 1)$ will now equal 0. So, $P(i|j)$ can drastically vary from year to year with recruitment. Note that in the second year, 75% of the fish in the first length class will be assigned to age 0 if the ALK from the first year is applied to the length sample from the second year even though 0% of the fish are age 0.

While one should not apply a key from one year to a different year, one can apply a key that was developed from one gear to a different gear so long as the gears are fishing the same population. The two gears could have different size-based selectivities but if, within a length bin, the fish available to each gear are from the same population, then the age composition within that length bin should be the same in the catch from both gears (Kimura 1977; Westrheim and Ricker, 1978). Reed and Wilson (1979), as a response to the Westrheim and Ricker (1978) paper, noted that if the probability of capture within a length bin is age dependent, then the age composition of the catch within that length bin could differ between the two gears. This point was also illustrated by Aanes and Vølstad (2015) who compared ALKs developed from a longline and gillnet survey both targeting the same population of eastern North Atlantic cod. While this observation is valid in theory, it is rarely of practical interest since the bias can be largely avoided by using narrow length bins [see Reed and Wilson (1979) for a demonstration of how the bias becomes negligible as the bin width is gradually reduced]. Therefore, for all intents and purposes, it is deemed acceptable to borrow a key

Key points
(1) The age-length and the length frequency samples must originate from the *same statistical* population, i.e. within a length class, the underlying age composition must be the same for the two samples. In other words, the two samples must be drawn from the same available population. This implies that:

(a) A forward key developed from one year cannot be applied to another year.

(b) A forward key developed from one area cannot be applied to another area if the two areas are characterized by differences in age composition (e.g. differences in availability due to age-dependent migration patterns, or differences in area-dependent survivorship).

(2) A forward key developed from one gear can be used to age catch from a different gear even if the two gears have different size selectivities, so long as the two selectivity curves within a length bin are parallel. With narrow length bins, selectivity is almost constant, hence, the requirement of parallel selectivity curves is met.

The inverse age-length key
Methodology

Length information is usually collected on an annual basis but, not uncommonly, there are some years with missing age data or age data based on inadequate sample sizes. This is where the inverse key becomes useful. The inverse key describes the probability $P(j|i)$ that a fish is of length j given that it belongs to age class i. Contrary to the probability of age given size, the probability of size given age is not affected by variability in recruitment and survivorship. What does, however, affect the probability of size given age is spatiotemporal variations of size at age. These could be caused by changes in growth rates, or changes in mean size at age due to changes in fishing practices, for example. So the inverse key can be applied to samples from populations with differing age compositions than the population from which it was derived, so long as size at age does not vary considerably among sampling events.

The inverse ALK approach was first conceived by Clark (1981). The probability of size given age can be estimated from an age-length sample taken in year k (or pooled over multiple years) using the method of moments as:

$$\hat{P}(j|i) = \hat{r}_{ij} = {}^{n_{ij}}/_{n_{i.}} \tag{3}$$

where the $\hat{r}_{ij}$ are the probabilities of length given age that populate the cells of the inverse ALK matrix, $\boldsymbol{R}$. All other notations are defined in Table 1. Let E^* denote the vector containing estimates of the marginal probabilities $\hat{P}(j)$ obtained from the length frequency sample taken in year $k^{'}$ $(k^{'} \neq k)$:

$$E^* = \left[\frac{y_1}{N}, \frac{y_2}{N}, \dots, \frac{y_J}{N}\right]^T \tag{4}$$

Then the elements of the estimated length composition (e_j^*) can be expressed as:

$$e_j^* = \sum_{i=1}^{I} \hat{P}(j|i)\hat{P}(i) \tag{5}$$

which in matrix notation, yields:

$$E^* = \boldsymbol{R}^T \tilde{A} \tag{6}$$

where $\tilde{A}$ is the estimated age composition from the inverse key. This system can be solved by taking the generalized inverse of $\boldsymbol{R}^T$:

$$\tilde{A} = (\boldsymbol{R}\boldsymbol{R}^T)^{-1}\boldsymbol{R}E^* \tag{7}$$

which is the least squares solution of Equation (6) provided that the number of length bins in the age-length sample is greater than or equal to the number of age classes and that $\boldsymbol{R}^T$ is of full column rank (i.e. each of the columns of the matrix is linearly independent of the others).

The above demonstrates the logic of the inverse key, and the comparison of Equations (2) and (7) highlights the difference in the approaches. However, there are more efficient ways to estimate the parameters in the inverse key approach than Equation (7).

Noticing that the ordinary least squares estimator could yield infeasible (i.e. negative) estimates, Clark (1981) solved the system by restricted least squares, restricting proportions to be non-negative. Though the inverse key approach was groundbreaking at the time, there were two issues hindering its wider use. The first was that calculating the generalized inverse is prone to numerical instability. The second was that an assumption implicit to the least-squares approaches is that the independent variable (in this case, the age-length sample) is known without error and that all the error is in the dependent variable (in this case, the length frequency sample). In reality, the length frequency sample is the larger sample and thus thought to be known fairly precisely, whereas the age-length sample is typically small and therefore more likely to be subject to sampling error.

Kimura and Chikuni (1987) and Hoenig and Heisey (1987) concurrently sought to address these issues by finding maximum likelihood estimates using the expectation-maximization algorithm (Dempster *et al.*, 1977). Kimura and Chikuni (1987) kept the inverse key fixed during the iterative process, still only allocating uncertainty in the likelihood to the length frequency sample; whereas Hoenig and Heisey (1987) allowed for the inverse key, together with the probabilities at age $\hat{P}(i)$, to be updated at each iteration of the algorithm, modelling uncertainty in the likelihood in both the length frequency sample and the age-length sample.

Assumptions, applications, and limitations

Hoenig and Heisey (1987) originally thought that their estimator is invalid if length stratification is used. However, Hoenig *et al.* (2002) showed that the estimator is in fact valid in this case. Thus, of the various approaches to inverse ALK, only the ones proposed by Hoenig and Heisey (1987) and Hoenig *et al.* (2002) allow for sampling error in both the length frequency and the age-length samples.

Key points

(1) The number of length bins (J) must be greater than or equal to the number of age classes (I) in order to obtain a unique solution (in some cases, a plus group will need to be implemented).

(2) The age-length and the length frequency samples do not need to have been collected in the same year. They can be collected from two populations with different age compositions as long as size at age does not differ between the two populations.

(3) The Hoenig and Heisey (1987) method is the superior method for applying inverse keys when there is a single length frequency and a single age-length sample as it allows

for uncertainty in both the length frequency sample and the age-length sample.

The combined forward-inverse age-length (FIAL) key
Methodology

The FIAL key links the concepts of forward and inverse keys using Bayes rule in a maximum likelihood framework (see Hoenig *et al.*, 2002). In years without age data, it uses the distribution of length at age whereas in years with age data, it essentially uses the information on age given length but penalizes the estimates if they deviate from the distribution of length at age. This is possible because $P(j|i)$ and $P(i|j)$ are related by Bayes rule $(P(i|j)_k = \frac{P(j|i)P(i)_k}{\sum_{h=1}^{I} P(j|h)P(h)_k}$ such that the likelihoods for both the forward and inverse keys can be written in terms of $P(j|i)$. The forward key approach and the inverse key approach can both be expressed as the product of independent multinomials, so the FIAL key is also a product multinomial.

Let us illustrate this using an example described in Hoenig *et al.* (2002). Imagine a population, sampled over 2 years, for which three datasets are available. In the first year (denoted by the subscript "1" in the equations to follow), a random sample of n_1 fish is measured and aged. In the second year (denoted by the subscript "2"), length frequency is recorded on a large random sample of size M_2, and age-length information is obtained from a much smaller random sample (n_2), possibly stratified by length.

The forward key method cannot be used to estimate age composition in the first year, since no length frequency data are available for that year (but, of course, the age composition can be estimated from the aged random sample). The age composition in year 2 can be estimated from the aged sample from year 1 (using the inverse key) or from year 2 (using the forward key) or from both aged samples using the FIAL key. The likelihood for year 2 using the forward key approach (Λ_1) is proportional to:

$$\Lambda_1 \propto \prod_{i=1}^{I} \prod_{j=1}^{J} [P(i|j)_2 P(j)_2]^{n_{ij2}} \prod_{j=1}^{J} P(j)_2^{y_{j2}} \quad (8)$$

where the first part of the likelihood matches the model estimate of the joint probability of ages and lengths with observations from the age-length sample (n_{ij2}), and the second part of the likelihood matches the model estimate of the marginal probability of lengths with observations from the length frequency sample (y_{j2}). Estimates of age composition for that year (i.e. probabilities $\hat{P}(i)_2$) can then be obtained using the invariance principle of maximum likelihood estimation:

$$\hat{P}(i)_2 = \sum_{j=1}^{J} \hat{P}(i|j)_2 \hat{P}(j)_2 \quad (9)$$

The estimates obtained from Equations (8) and (9) are the same as those obtained from Equation (2).

With the inverse key, we do not need the age-length sample and length frequency samples to have been collected in the same year. By assuming that the conditional probability of size given age stays constant from year to year, we can use the age-length sample from year 1 to analyse the length frequency from year 2.

Let $P(j|i)_{12}$ be the probability of size given age that is common to both years. The likelihood for the data using the inverse key approach (Λ_2) can be written as:

$$\Lambda_2 \propto \prod_{i=1}^{I} \prod_{j=1}^{J} [P(j|i)_{12} P(i)_1]^{n_{ij1}} \prod_{j=1}^{J} \left[\sum_{i=1}^{I} P(j|i)_{12} P(i)_2\right]^{y_{j2}} \quad (10)$$

where the first part of the likelihood matches the model estimate of the joint probability of ages and lengths with observations from the age-length sample from year 1 (n_{ij1}), and the second part of the likelihood matches the model estimate of the marginal probability of lengths with observations from the length frequency sample from year 2 (y_{j2}). The number of length bins (J) must be greater than or equal to the number of age classes (I).

The FIAL key allows for all three datasets to be analysed simultaneously, thus the likelihood for all the data (Λ_3) would simply be:

$$\Lambda_3 \propto \prod_{i=1}^{I} \prod_{j=1}^{J} [P(j|i)_{12} P(i)_1]^{n_{ij1}} \prod_{i=1}^{I} \prod_{j=1}^{J} [P(j|i)_{12} P(i)_2]^{n_{ij2}}$$
$$\times \prod_{j=1}^{J} \left[\sum_{i=1}^{I} P(j|i)_{12} P(i)_2\right]^{y_{j2}} \quad (11)$$

where the first part of the likelihood matches the model estimate of the joint probability of ages and lengths with observations from the age-length sample from year 1 (n_{ij1}), the second part of the likelihood matches the model estimate of the joint probability of ages and lengths with observations from the age-length sample from year 2 (n_{ij2}), and the third part of the likelihood matches the model estimate of the marginal probability of lengths with observations from the length frequency sample from year 2 (y_{j2}). Note that Equation (11) is written in terms of $P(j|i)$ but the parts of the likelihood pertaining to the n_{ij2} and the Y_{j2} are equivalent to the forward key described by Equations (8) and (9) (see Appendix 1). There are $IJ + 2I$ parameters in the model described in Equation (11). But, since each row of the $P(j|i)_{12}$ matrix and each of the $P(i)$ vectors must, by definition, add up to 1, only $IJ + I - 2$ parameters need to be estimated.

This likelihood can be generalized to the case where k years are surveyed and where, in addition to the age-length and length frequency samples, age-only samples (x_{ik}) are collected in certain years:

$$\Lambda_g \propto \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{k=1}^{K} [P(j|i)P(i)_k]^{n_{ijk}} \prod_{j=1}^{J} \prod_{k=1}^{K} \left[\sum_{i=1}^{I} P(j|i)P(i)_k\right]^{y_{jk}}$$
$$\times \prod_{i=1}^{I} \prod_{k=1}^{K} P(i)_k^{x_{ik}} \quad (12)$$

where $P(i)_k$ pertains to the age composition in the kth year, n_{ijk} corresponds to the number of fish cross-classified as ij in the kth year, and $P(j|i)$ is assumed constant throughout the years. Hoenig *et al.* (2002) show how the general model applies even when fixed sub-sampling by length (i.e. length stratification) is employed.

Assumptions, applications, and limitations
The FIAL key can be very useful for situations where, in certain years, only part of the total length frequency was sampled for

ages. It can also be very useful for estimating historical age composition for fisheries in which age data has only recently been collected. However, like the inverse key, the FIAL key makes the strong assumption that the distributions of size at age do not vary through time and space. Violating this assumption can lead to biases in the estimates of age composition, ultimately affecting the appraisal of stock status.

Therefore, if working with a fishery for which certain years contain little to no age-length data, while other years contain reliable and representative age-length samples, the best approach would be to use both the forward key and the FIAL key approaches. This would be achieved by applying the FIAL key over all available samples, such that the superior age-length samples inform estimates of the overall probability of size given age, but then replace the estimates of age composition for years in which there are good age data with those obtained from a forward ALK analysis. This would relax the assumption of constant size at age at least for the most informed years. One could also use this strategy to test whether or not size at age has changed with time by comparing estimates of age composition from the FIAL key with age composition estimates from the forward key in years for which good age-length data are available.

In Ailloud *et al.* (2019), the FIAL key is evaluated by simulation and applied to the complicated case of western Atlantic bluefin tuna. While the method performs well even with error-prone simulated data, it has much greater difficulty converging to the global minimum when using the real dataset for western Atlantic bluefin tuna, where most age-length samples were obtained from opportunistic data collection programmes. This exercise emphasizes the importance of striving to follow a statistically robust sampling protocol when collecting age-length samples. For any method, if the data are flawed, the results may be misleading. In general, it is difficult to predict how ALK methods will perform when assumptions are violated. We therefore recommend using simulation to determine the seriousness of various types of assumption violations if such violations are suspected in the data.

Key points
(1) The number of length bins (J) must be greater than or equal to the number of age classes (I) in order to obtain a unique solution.
(2) Size-at-age is assumed constant among samples.
(3) The estimator is valid even if length stratification is used.

Discussion and conclusions
The forward ALK is the preferred method for estimating age composition if populations are adequately sampled and age-length data are available for each year. It is the method that makes the fewest assumptions and is therefore expected to result in the most robust estimates of age composition. However, more often than not, age data are missing for at least part of the time series for which age composition estimates are needed. In that situation, the FIAL key can allow scientists to bridge the gap between years with little to no age data and years with good age data to derive estimates of age composition. Should complete and representative samples be available only for certain years, our recommendation would be to use results from the FIAL key (using data from all years) to obtain estimates for years with inadequate age-length data, and use the results from the forward key in years for which complete and representative age-length samples are available. In

practice, age-length samples are rarely fully representative of the target population and model assumptions are often violated. As such, we strongly recommend using simulation to predict what impact these violations will have on model performance (see Ailloud *et al.*, 2019).

Stock Synthesis (SS, Methot and Wetzel, 2013), a widely used stock assessment modelling platform, uses a framework that closely resembles the FIAL key. The model makes use of both the forward and inverse keys in a complicated procedure that attempts to capture age and size selectivity (Methot, 2000). However, one major difference with the FIAL key is SS's use of a parametric inverse key. Instead of using the age-length data empirically as raw proportions of size at age, SS estimates growth parameters using the raw data, accounting for age and size selectivity, and calculates probabilities of size at age based on the mean and residual variance of the fitted curve (Methot, 2000). Another notable difference is that, unlike SS, the FIAL key does not make use of an underlying model of the stock to estimate the age composition of the survey/fishery. ALKs estimate the age composition (of a recreational, commercial, or survey catch) without trying to explain the causes of age composition (in terms of mortality, selectivity, and other factors); the keys do not need information on selectivity and other factors to estimate age composition because they are simply expanding observed proportions to the entire catch. It is advantageous to use all available information to describe and explain stock dynamics, as done by stock synthesis; it is also of interest to estimate parameters while making minimal assumptions to see if estimation is robust to the assumptions, a strength of the FIAL key.

Supplementary data
Supplementary material is available at the *ICESJMS* online version of the manuscript.

References
Aanes, S., and Vølstad, J. H. 2015. Efficient statistical estimators and sampling strategies for estimating the age composition of fish. Canadian Journal of Fisheries and Aquatic Sciences, 72: 938–953.

Ailloud, L. E., Lauretta, M. V., Walter, III, J. F., and Hoenig, J. M. 2019. Estimating age composition for multiple years when there are gaps in the ageing data: the case of western Atlantic bluefin tuna. ICES Journal of Marine Science, 10.1093/icesjms/fsz069.

ASMFC (Atlantic States Marine Fisheries Commission). 2010. Atlantic Croaker 2010 Benchmark stock assessment. http://www.asmfc.org/uploads/file/5282798aatlanticCroaker2010BenchmarkStockAssessment.pdf (last accessed 25 June 2018).

Bartoo, N. W., and Parker, K. R. 1983. Stochastic age-frequency estimation using the von Bertalanffy growth equation. Fishery Bulletin, 81: 91–96.

Cailliet, G. M., and Goldman, K. J. 2004. Age determination and validation in chondrichthyan fishes. *In* Biology of Sharks and Their Relatives, pp. 399–447. Ed. by J. C. Carrier, J. A. Musick, and M. R. Heithaus. CRC Press, Boca Raton, FL.

Carpi, P., Santojanni, A., Donato, F., Colella, S., Keč, V. Č., Zorica, B., Leonori, I. *et al.* 2015. A joint stock assessment for the anchovy stock of the northern and central Adriatic Sea: comparison of two catch-at-age models. Scientia Marina, 79: 57–70.

Chih, C. P. 2010. Incorporating effective sample sizes into sampling designs for reef fish. Fisheries Research, 105: 102–110.

Clark, W. G. 1981. Restricted least-squares estimates of age composition from length composition. Canadian Journal of Fisheries and Aquatic Sciences, 38: 297–307.

Coggins, L. G., Gwinn, D. C., and Allen, M. S. 2013. Evaluation of age–length key sample sizes required to estimate fish total mortality and growth. Transactions of the American Fisheries Society, 142: 832–840.

Dempster, A. P., Laird, N. M., and Rubin, D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B, Statistical Methodology, 39: 1–38.

Durant, J. M., Hidalgo, M., Rouyer, T. A., Hjermann, D. Ø., Ciannelli, L., Eikeset, A. M., Yaragina, N. A. *et al.* 2013. Population growth across heterogeneous environments: effects of harvesting and age structure. Marine Ecology Progress Series, 480: 277–287.

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielsen, A. *et al.* 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Journal Optimization Methods and Software, 27: 233–249.

Fridriksson, A. 1934. On the calculation of age distribution within a stock of cod by means of relatively few age-determinations as a key to measurements on a large scale. Rapports et Proces-verbaux des Réunions du Conseil International pour l'Éxploration de la Mer, 86: 1–14.

Gascuel, D. 1994. Une méthode simple d'ajustement des clés taille/-âge: application aux captures d'albacores (*Thunnus albacares*) de l'Atlantique Est. Canadian Journal of Fisheries and Aquatic Sciences, 51: 723–733.

Greenstreet, S. P., Spence, F. E., and McMillan, J. A. 1999. Fishing effects in northeast Atlantic shelf seas: patterns in fishing effort, diversity and community structure. V. Changes in structure of the North Sea groundfish species assemblage between 1925 and 1996. Fisheries Research, 40: 153–183.

Haeseker, S. L., Jones, M. L., and Bence, J. R. 2003. Estimating uncertainty in the stock-recruitment relationship for St. Marys River sea lampreys. Journal of Great Lakes Research, 29: 728–741.

Hixon, M. A., Johnson, D. W., and Sogard, S. M. 2014. BOFFFFs: on the importance of conserving old-growth age structure in fishery populations. ICES Journal of Marine Science, 71: 2171–2185.

Hoenig, J. M., and Heisey, D. M. 1987. Use of a log-linear model with the EM algorithm to correct estimates of stock composition and to convert length to age. Transactions of the American Fisheries Society, 116: 232–243.

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

Jennings, S., Reynolds, J. D., and Mills, S. C. 1998. Life history correlates of responses to fisheries exploitation. Proceedings of the Royal Society B: Biological Sciences, 265: 333–339.

Kimura, D. K. 1977. Statistical assessment of the age–length key. Journal of the Fisheries Research Board of Canada, 34: 317–324.

Kimura, D., and Chikuni, S. 1987. Mixtures of empirical distributions: an iterative application of the age-length key. Biometrics, 43: 23–35.

Lai, H. L. 1993. Optimal sampling design for using the age-length key to estimate age composition of a fish population. Fishery Bulletin, 91: 382–388.

Marteinsdottir, G., and Thorarinsson, K. 1998. Improving the stock-recruitment relationship in Icelandic cod (*Gadus morhua*) by including age diversity of spawners. Canadian Journal of Fisheries and Aquatic Sciences, 55: 1372–1377.

Methot, R. D. 2000. Technical description of the Stock Synthesis assessment program. NOAA Tech Memo. NMFS-NWFSC-43. https://www.nwfsc.noaa.gov/assets/25/6189_06162004_143158_tm43.pdf (last accessed 25 February 2019).

Methot, R. D., and Wetzel, C. R. 2013. Stock Synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fisheries Research, 142: 86–99.

Murta, A. G., and Vendrell, C. 2009. Using the EM algorithm to age fish eggs. ICES Journal of Marine Science, 66: 607–616.

Murta, A. G., Loff, J. F., Neves, M., and Wise, L. 2016. ALKr: A R package of methods based on age-length keys to estimate the age structure of fish populations. https://www.iccat.int/GBYP/Docs/Modelling_Phase_3_R_Package_ALKr.pdf (last accessed 16 August 2018).

Oeberst, R. 2000. A universal cost function for the optimization of the number of age determinations and length measurements for age-length-key tables. Archive of Fishery and Marine Research, 48: 43–60.

Pestal, G. P., Taylor, D. M., Hoenig, J. M., Shields, J. D., and Pickavance, R. 2003. Monitoring the prevalence of the parasitic dinoflagellate *Hematodinium* sp. in snow crabs *Chionoecetes opilio* from Conception Bay, Newfoundland. Diseases of Aquatic Organisms, 53: 67–75.

Quist, M. C., Pegg, M. A., and DeVries, D. R. 2012. Age and growth. *In* Fisheries Techniques, 3rd edn. Ed. by A. V. Zale, D. L. Parrish, and T. M. Sutton. American Fisheries Society, Bethesda, MD. 2012. pp. 677–731.

Reed, W. J., and Wilson, K. H., 1979. Comment on "Bias in age-length keys." Journal of the Fisheries Research Board of Canada, 36: 1159–1160.

Rouyer, T., Ottersen, G., Durant, J. M., Hidalgo, M., Hjermann, D. Ø., Persson, J., Stige, L. C. *et al.* 2011. Shifting dynamic forces in fish stock fluctuations triggered by age truncation? Global Change Biology, 17: 3046–3057.

Tanaka, S. 1953. Precision of age-composition of fish estimated by double sampling method using the length for stratification. Nippon Suisan Gakkaishi, 19: 657–670.

Tenenbein, A. 1972. A double sampling scheme for estimating from misclassified multinomial data with applications to sampling inspection. Technometrics, 14: 187–202.

Westrheim, S. J., and Ricker, W. E. 1978. Bias in using an age–length key to estimate age-frequency distributions. Journal of the Fisheries Research Board of Canada, 35: 184–189.

Wyanski, D. M., White, D. B., and Barans, C. A. 2000. Growth, population age structure, and aspects of the reproductive biology of snowy grouper, *Epinephelus niveatus*, off North Carolina and South Carolina. South Carolina State Documents Depository. https://dc.statelibrary.sc.gov/handle/10827/10557 (last accessed June 2018).

Handling editor: Francis Juanes

Appendix 1. The forward key as a special case of the FIAL likelihood

The likelihood for the forward key was given in [Equation (8)](#) as

$$\Lambda_1 \propto \prod_{i=1}^{I} \prod_{j=1}^{J} [P(i|j)_2 P(j)_2]^{n_{ij2}} \prod_{j=1}^{J} P(j)_2^{y_{j2}} \tag{A1}$$

We note that the likelihood is parameterized in terms of $P(i|j)$ and $P(j)$, i.e. in terms of the distribution of age given length and the distribution of length. There are $J(I-1)$ conditional probabilities $P(i|j)$ and $J-1$ probabilities of length $P(j)$ for a total of $IJ - J + J - 1 = IJ - 1$ parameters to estimate.

The portion of the FIAL key pertaining to the forward key is contained in the second and third parts of the likelihood in [Equation (11)](#):

$$\Lambda_4 \propto \prod_{i=1}^{I} \prod_{j=1}^{J} [P(j|i)_2 P(i)_2]^{n_{ij2}} \prod_{j=1}^{J} \left[\sum_{i=1}^{I} P(j|i)_2 P(i)_2 \right]^{y_{j2}} \tag{A2}$$

This is written in terms of $P(j|i)$ and $P(i)$. There are $I(J-1)$ conditional probabilities to estimate and $I-1$ parameters $P(i)$ to estimate, thus $IJ - 1$ parameters.

Now in Equation (A2),

$$P(j|i)_2 P(i)_2 = P(i, \ j) \ = P(i|j)P(j)$$

and

$$\sum_{i=1}^{I} P(j|i)_2 P(i)_2 = P(j)_2.$$

Substituting these into [Equation (A2)](#) gives

$$\Lambda_4 \propto \prod_{i=1}^{I} \prod_{j=1}^{J} [P(i|j)_2 P(j)_2]^{n_{ij2}} \prod_{j=1}^{J} \left[P(j)_2 \right]^{y_{j2}}$$

which is the same as [Equation (A1)](#).