Original Article

Estimating age composition for multiple years when there are gaps in the ageing data: the case of western Atlantic bluefin tuna

Lisa E. Ailloud [1]*, Matthew V. Lauretta[2], John F. Walter III[2], and John M. Hoenig[3]

[1]International Commission for the Conservation of Atlantic Tunas, AOTTP, Corazón de María, 8, 28002 Madrid, Spain
[2]Southeast Fisheries Science Center, National Marine Fisheries Service, Sustainable Fisheries Division, 75 Virginia Beach Drive, Miami, FL 33149, USA
[3]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.

Age–length key (ALK) methods generally perform well when length samples and age samples are representative of the underlying population. It is unclear how well these methods perform when lengths are representative but age samples are sparse (i.e. age samples are small or missing in many years, and some length groups do not have any age observations). With western Atlantic bluefin tuna, the available age data are sparse and have been, for the most part, collected opportunistically. We evaluated two methods capable of accommodating sparse age data: a novel hybrid ALK (combining forward ALKs and cohort slicing) and the combined forward-inverse ALK. Our goal was to determine if the methods performed better than cohort slicing, which has traditionally been used to obtain catch-at-age for Atlantic bluefin tuna, given the data limitations outlined above. Simulation results indicated that the combined forward-inverse ALK performed much better than the other methods. When applied to western Atlantic bluefin tuna data, the combined forward-inverse ALK approach was able to track cohorts and identified an inconsistency in the ageing of some samples.

Keywords: age–length keys, age composition, Atlantic bluefin tuna, simulation, sparse data, stock assessment.

Introduction

Atlantic bluefin tuna (ABFT) are managed by member nations of the International Commission for the Conservation of Atlantic Tunas (ICCAT). Two stocks are currently recognized: the eastern stock with spawning grounds in the Mediterranean Sea and the western stock with spawning grounds in and around the Gulf of Mexico (Carlsson et al., 2006; Rooker et al., 2007). Additional spawning grounds have recently been discovered in the NW Atlantic, but the origin and extent of these recruits have not yet been characterized fully (Richardson et al., 2016; Walter et al., 2016). Though the two stocks mix throughout most of their foraging range (Block et al., 2005; Dickhut et al., 2009; Rooker et al., 2008; Wilson et al., 2015), they are managed as two separate units delineated by the 45°W meridian. The eastern stock is estimated to be ~10 times larger than the western stock (Fromentin and

Powers, 2005) and mixing rates have been found to vary across ages, space, and time (Siskey et al., 2016). ABFT are relatively long lived (up to 34 years of age; Ailloud et al., 2017) and carry out extensive migrations across the Atlantic Ocean where they are targeted by a wide range of fisheries that differentially harvest multiple age groups.

Age data derived from the reading of hard parts (spines and otoliths) are needed to accurately characterize the age composition of catches of each stock. However, the complex, highly migratory nature of these fish, and the multinational nature of the fisheries, present challenges for data collectors (Anon, 2014; Rodriguez-Marin et al., 2015). For ABFT, a small proportion of fish are sampled for length but the quality of these samples is poorly known (Justel-Rubio and Ortiz, 2013), but, in the assessment the catch and length-frequency distributions are assumed to

be well known, which is the assumption we will make in this paper. Data on ages of individual fish, however, are undoubtedly sparse. We define sparse age data as being characterized by years with no age data, or very small sample sizes, and years where some length bins have not been sampled for age. The stock assessment contains over 40 years of length-frequency data, yet only 20 of these years contain age information, and, for many years, only some sizes were aged. The earliest records of age data date to 1974 in the West Atlantic (Table 1) and 1984 in the East Atlantic. It was not until 2010 that age data started being collected on a large scale and annual basis for western ABFT. ICCAT has now made it a priority to collect age data to improve estimates of the population age structure; but although efforts are in place to try to obtain larger and more representative samples of hard parts (Anon, 2014; Busawon *et al.*, 2014), historical samples have, for the most part, been obtained from opportunistic sampling programmes instead of a formal sampling design (see Supplementary Figure S1 for a depiction of available western ABFT samples by year, gear and area).

ABFT stocks have traditionally been assessed using a virtual population analysis (VPA). Unlike integrated assessments, which are able to convert length frequencies to age frequencies internally, the VPA requires annual catch-at-age as an input and projects numbers backwards in time from the oldest to the youngest ages to reconstruct the population size by age. Cohort slicing has conventionally been used to produce these catch-at-age estimates (see Mohn and Savard, 1989). A growth model is used to specify size bins corresponding to each age class, and the catch-at-size data are assigned ages accordingly. The technique, which proves useful when age data are sparse or unavailable, makes the strong assumption that there is no overlap in size between adjacent age classes. Violations of this assumption tend to (i) underestimate recruitment variability (Mohn, 1994; Restrepo, 1995) and (ii) underestimate the contribution of younger fish while overestimating the contribution of older fish (Goodyear, 1987; Kell and Kell, 2011; Ailloud *et al.*, 2015). As these errors propagate through the assessment, they can translate into bias in parameter estimates derived from cohort-sliced catch-at-age data (Ailloud *et al.*, 2015), potentially affecting the evaluation of stock status and future projections.

If data on age and length of individual fish are available, age–length keys (ALKs) offer a better alternative for estimating catch-at-age (Ailloud and Hoenig, 2019). ALKs describe the distribution of age given size (forward ALK—Fridriksson, 1934; Kimura, 1977; Westrheim and Ricker, 1978), size given age (inverse ALK—Hoenig and Heisey, 1987; Kimura and Chikuni, 1987) or

Table 1. Actual age–length samples available for the West (fish captured in the western Atlantic/Gulf of Mexico).

Age/Year	1974	1975	1976	1977	1978	1996	1997	1998	1999	2000	2002	2009	2010	2011	2012	2013	2014	2015	Grand Total
0																			
1		26					1	8											35
2		53		1	1	12	6	10	1					15	8	1	16	4	128
3		9	11	3	4	5	6	3	3					50	63	13	38	21	229
4		4	5	6			9	6	2				3	65	90	37	30	90	347
5		3	4	3	1	4	8	1				1	10	67	58	34	35	24	253
6	2	1	5	1	3	3	3					5	4	51	30	16	14	10	148
7			1	1		12	2	2	1		2	7	22	52	49	11	22	6	190
8		1				15	3	1			3	9	54	100	57	47	24	1	315
9		3	1			15	2				3	10	83	184	55	51	29	11	447
10		1	2	1		16	1				8	5	78	111	65	51	54	17	410
11		4		2	1	2					8	8	39	63	44	62	59	37	329
12		2	1		1	2	1				7	9	23	32	32	45	41	51	247
13		1	1	1	4	1					11	8	16	27	17	33	32	39	191
14		1		1	7						5	11	12	20	12	26	19	26	140
15		2			4						2	9	19	23	11	6	16	20	112
16		3	3		2	1					2	16	15	27	24	13	12	7	125
17		8	3	1	2						3	1	11	38	15	27	5	7	121
18		8	3		6	1						4	4	16	20	32	13	7	114
19		9	6		2						1	6	10	10	12	18	11	8	93
20		3	3	1	4							3	9	16	4	14	12	9	78
21		3	1		6	1						1	3	20	9	15	7	4	70
22		2	2		9							1	3	15	11	7	3	6	59
23		4	4		10								1	4	3	11	3	2	42
24		2	7		5										1	2	9	7	33
25		1	2	2	6											3	3	1	18
26			1		8											2	1	3	15
27		2	1	2	4														9
28					7														7
29			1		5									1	1		1		9
30					2							1							3
31																			
32					2														2
33																			
34					1									1	1				3
Grand Total	2	156	68	26	98	75	34	43	21	6	55	115	420	1 009	692	585	506	411	4 322

both (combined forward-inverse ALK; Hoenig *et al.*, 2002). The age composition of a large sample of measured fish is estimated by summarizing the relationship between age and length of a much smaller subsample of fish for which ages have been determined, and then applying this relationship to the larger sample of fish for which only lengths are available. These keys are ideally constructed using length-stratified random sampling to achieve greater precision. A forward key from one year cannot be applied to a different year for which age data are missing because forward keys tend to preserve the age composition of the samples from which they were derived (Kimura, 1977; Westrheim and Ricker, 1978). As such, forward keys require age data to be collected every year and to cover the range of lengths observed and, thus, cannot alone be used to estimate age composition for western ABFT. We therefore explored two alternative estimation methods that can accommodate sparse age data: a novel hybrid ALK and the combined forward-inverse age–length (FIAL) key. The hybrid key (described below) forms a weighted average of cohort slicing and forward ALKs, whereas the FIAL key combines the forward and inverse approaches into one likelihood function.

While ALKs should, in theory, offer improvements over cohort slicing, it is unclear whether that holds true when age data are not collected following a statistically robust sampling design. Some authors have attempted to develop estimation methods that can accommodate non-random sampling designs (e.g. Hirst *et al.*, 2012), but our objective was to, instead, test how well different methods perform when fed poor quality data. For western ABFT, most age samples were obtained opportunistically. Moreover, of the 20 years of age data available, only 4 of those years have high sample sizes of aged fish (>500) and good coverage across size classes. The rest are characterized by low sample sizes that do not span the range of sizes observed in the catch (Table 1).

Our objective was therefore to determine whether the hybrid key or the combined forward-inverse key can offer improvements over cohort slicing for estimating age composition in western ABFT given the existing data limitations. In the first stage, we simulated catch-at-age and catch-at-length data and annual reference age–length samples patterned after the biology and sampling scheme of western ABFT and compared the performance of each method for estimating catch-at-age. We then tested the selected method against the dataset available for western ABFT, and compared the resulting estimates of age composition with those obtained from cohort slicing. Implications for the 2017 stock assessment results are discussed.

Material and methods
The following notation will be used:

> i refers to age
> j refers to length
> k refers to year
> m refers to month

When multiple subscripts are used, the appropriate ones are in the order i, j, k, m.

Catch-at-age estimation
Cohort slicing (CS). CS was performed on a monthly basis using the algorithm *AgeIT* developed by ICCAT (Ortiz and Palma,

2011). The algorithm defines length bins for each age group and month using a growth curve and an assumed birth month. It then compares the catch-at-size data against the lower and upper size limits associated with each age class to assign ages to the catch. For this exercise, the observed monthly catch-at-size data were given as an input and the growth curve from Ailloud *et al.* (2017) with a May birth month as per the 2017 assessment.

The hybrid ALK (HY). This novel, yet simple, approach makes use of the improved estimates produced by forward ALKs while using the convenience of cohort slicing to fill gaps where needed. With HY, if the sample size of otoliths in a given length bin falls below the accepted threshold (here, $T = 20$), the probability of age given size for that length bin in year k, $\hat{P}(i|j)_k^{HY}$, is estimated as the weighted sum of the probability of age given size obtained by analysing the data using forward ALKs, $\hat{P}(i|j)_k^{ALK}$, and the probability of age given size obtained from the cohort-sliced catch-at-age estimates, $\hat{P}(i|j)_k^{CS}$. If CS were conducted on an annual basis, $\hat{P}(i|j)_k^{CS}$ would simply be a matrix of zeros and ones, but with CS being conducted on a monthly basis the $\hat{P}(i|j)_k^{CS}$ cells can, in fact, fall between 0 and 1. The procedure can be expressed as follows:

$$\hat{P}(i|j)_k^{HY} = \begin{cases} \dfrac{n_{j,k}}{T}\hat{P}(i|j)_k^{ALK} + \dfrac{T - n_{j,k}}{T}\hat{P}(i|j)_k^{CS}, & \text{for } n_{j,k} < 20 \\ \hat{P}(i|j)_k^{ALK}, & \text{otherwise} \end{cases} \quad (1)$$

where $n_{j,k}$ is the sample size of otoliths in the jth length bin in the kth year and T is the threshold of 20 otolith samples per length bin for using just the estimate from the ALK.

The combined FIAL key. The method of Hoenig *et al.* (2002) combines the concepts of forward and inverse keys (see Ailloud and Hoenig, 2019). While the forward key looks at the distribution of ages in a size bin to obtain estimates of $P(i|j)$, the inverse key looks at the distribution of sizes given age to obtain estimates of $P(j|i)$. It is assumed that the $P(j|i)$ do not change over time (i.e. no variation in size-at-age over time) such that an inverse key developed from data from one (or more) year(s) can be applied to any year. One thinks of the logic of the inverse method as finding the weighting factors for the separate length-at-age distributions that cause the sum of the distributions to match the overall length-frequency distribution as closely as possible, with the weighting factors being the age composition. Hoenig *et al.* (2002) showed that the $P(j|i)$ can be expressed in terms of $P(i|j)$ and vice versa using Bayes Rule. Consequently, the forward and inverse approaches can be combined into one likelihood function and the catch-at-age can be estimated for both years with age data and years without age data as well as for years where only sparse age data are available.

Let the number of fish sampled in year k whose lengths j and ages i were both recorded be represented by the array $n_{i,j,k}$, the number of fish sampled in year k for which only lengths were recorded be represented by the matrix $y_{j,k}$, and the number of fish sampled in year k for which only ages were recorded be represented by the matrix $x_{i,k}$ (the $x_{i,k}$ are mainly of theoretical interest—we explain below why this can be useful). And let $P(i)_k$ represent the probability of age i in year k. The kernel of the log-likelihood (Λ) is then defined as the product of three components, a, b, and c:

$$\Lambda = abc \quad (2)$$

where

$$a = \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{k=1}^{K} \left[P(j|i)P(i)_k \right]^{n_{i,j,k}} \qquad (2a)$$

$$b = \prod_{j=1}^{J}\prod_{k=1}^{K} \left[\sum_{i=1}^{I} P(j|i)P(i)_k \right]^{y_{j,k}} \qquad (2b)$$

$$c = \prod_{i=1}^{I}\prod_{k=1}^{K} P(i)_k^{x_{i,k}} \qquad (2c)$$

In the above listed equations, a matches the model estimate of the joint probability of ages and lengths with observations from the age–length sample available for each year ($n_{i,j,k}$), b matches the model estimate of the marginal probability of lengths with observations from the length-frequency sample available for each year ($y_{j,k}$), and c matches the model estimate of the marginal probability of age ($P(i)$) for each year with counts of fish for which only ages are available each year ($x_{i,k}$). Ages (i) range from 0 to 16+ (where "16+" combines all fish ages 16 and above), j refers to 15 cm length bins ($j \in$ {(20, 35), [35, 50), …, [335, 349)}), and k refers to years ($k = 1974, 1975,…, 2015$).

The optimization was carried out in AD Model Builder (ADMB). To check for proper convergence, the optimization was run with different starting values until three consecutive iterations converged on the same log-likelihood value. All $x_{i,k}$ were set to 1 fish to keep $P(i)_k$ estimates off zero. This facilitated finding the global maximum of the likelihood. To save memory space and avoid boundary problems, the $P(j|i)$ matrix was set up as a ragged array in ADMB. Only the elements of $P(j|i)$ corresponding to non-zero elements in the matrix of age data collapsed overall years ($\sum_k n_{i,j,k}$) were estimated. In other words, it was assumed that if a fish of age i and length j had never been observed in the overall age sample then the probability of being age i for a fish of length j was zero.

The proportions-at-age estimates resulting from CS, HY, and FIAL henceforth will be referred to as $\hat{p}^{CS}$, $\hat{p}^{HY}$, and $\hat{p}^{FIAL}$, respectively.

Simulation

We used a simulation analysis to reproduce population dynamics patterned after western ABFT and test the relative performance of the three different catch-at-age estimation methods. Recruitment (age 1), growth and mortality data from 1974 to 2015 were obtained from the 2017 western ABFT VPA base case scenario (ICCAT, 2017). These data were used to simulate true catch-at-age, and then generate observed catch-at-size (subject to measurement error), and age–length samples (subject to random ageing error and error in the subsampling of the catch; i.e. clustering and unequal probability of selection among size classes). Different scenarios regarding recruitment variability, changes in mean size-at-age over time, magnitude of measurement error, and balance in the age samples were explored (see Base case and alternative scenarios section). For each scenario, catch-at-size data and an age–length sample were generated 100 times (with error) and performance measures [root mean square error (RMSE) in the estimated proportions-at-age] were summarized over the 100 runs to evaluate the performance of each estimation method for each of the 8 scenarios.

Data generation

Annual recruitment values for age 0 fish in year k ($N_{0,k}$) were back-calculated using estimated numbers of age 1 fish ($N_{1,k}$) assuming a natural mortality rate (M_0) of 0.41 for age 0 fish and a fishing mortality rate ($F_{0,k}$) equal to 25% of the fishing mortality on age 1 fish for that year ($F_{1,k}$):

$$N_{0,k} = \frac{N_{1,k+1}}{e^{-(M_0 + F_{0,k})}} \qquad (3a)$$

where

$$F_{0,k} = 0.25 F_{1,k} \qquad (3b)$$

and where i stands for age and k stands for year. Because the most recent 3 years of recruitment (2013–2015) are not well estimated in the VPA, they were replaced by the geometric mean recruitment (age 1) for the period 2006–2012 (96 637 fish). Numbers-at-age were projected forward to age 30 using a monthly (m) time step for total mortality (Z), assuming a birth month of May:

$$N_{i,k,m+1} = N_{i,k,m}e^{-Z_{i,k,m}} \qquad (4)$$

where

$$Z_{i,k,m} = M_{i,k,m} + F_{i,k,m} \qquad (5)$$

Annual mortality rates were modified to accommodate a monthly time step (as used in the actual Bluefin tuna assessment): natural mortality was assumed uniform over the year, while fishing mortality was assumed to follow a symmetric triangular distribution over the year with a mode at month 6 (i.e. highest F in the summer and lowest F in the winter).

Catch-at-age ($C_{i,k,m}$) for each age, year and month was calculated as

$$C_{i,k,m} = \frac{F_{i,k,m}}{Z_{i,k,m}} \left(1 - e^{-Z_{i,k,m}}\right) N_{i,k,m} \qquad (6)$$

Mean size-at-age and standard deviation in size-at-age were obtained from the Richards growth equation in Ailloud *et al.* (2017) to calculate probabilities of size given age for each year and month. Size-at-age was assumed to be normally distributed and no seasonality in growth was incorporated into the growth equation. The resulting probabilities were used to convert catch-at-age into catch-at-size, creating what we will refer to as the "true" catch-at-age-and-size. A normally distributed error term, $\varepsilon_{L,x} \sim N\left(0, \sigma_L^2 = 25 \text{ cm}\right)$, was then added to the lengths of individual fish (x) to simulate measurement error and produce the "observed" catch-at-size data ($C'_{j,k,m}$) to be used in our age composition estimation models. A variance of 25 cm was chosen because it was nearly double that reported in Ailloud *et al.* (2017). The larger variance was adopted to reflect a situation where many of the measurements are taken shipboard by fishers or untrained staff.

The objective when generating age–length samples for the simulation was to mimic the data availability of western ABFT. Western ABFT data are characterized by fewer samples in the earlier years and more numerous samples in recent years. Moreover, because fish caught together (same gear/area) tend to be more similar in size than fish in the overall population (see
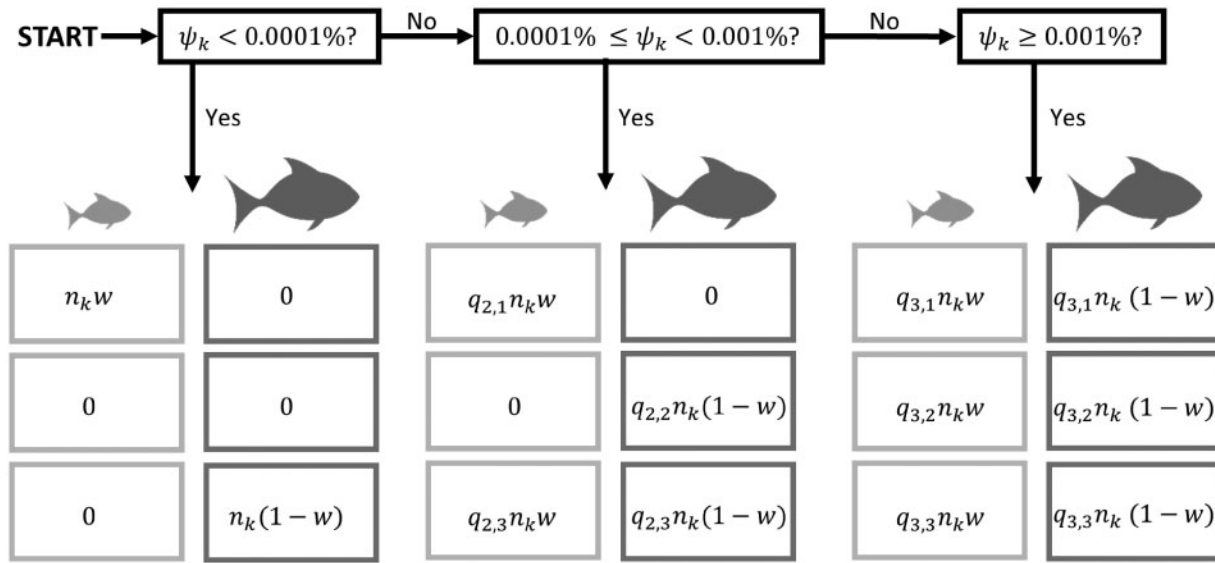
Figure 1. Illustration of one realization of the simulation sampling scheme. Light grey boxes represent clusters belonging to the small fish (SM) group, dark grey boxes represent clusters belonging to the large fish (LG) group. Ψ_k is the number of fish aged as a percent of total catch in year k. The number inside each box represents the sample size of fish to be extracted from each cluster. n_k is the total sample size of aged fish available in year k. w is the proportion of small fish in the sample (for the base case scenario, the sample is balanced between large and small fish, thus $w = 0.5$; this number is later changed in alternative scenarios 2 and 3, when samples are purposely skewed towards small and large fish, respectively). $q_{(c,z)}$ is a randomly selected (without replacement) fraction used to create uneven samples in each cluster ($q_{(1,z)} \in \{0,0,1\}$ in the case where one cluster from the LG and SM groups are selected, $q_{(2,z)} \in \{0, \frac{1}{3}, \frac{2}{3}\}$ in the case where two clusters from the LG and SM groups are selected, and $q_{(3,z)} \in \{\frac{1}{6}, \frac{2}{6}, \frac{3}{6}\}$ in the case where three clusters from the LG and SM groups are selected). The choice of which clusters get zero samples taken out of them would, of course, change from run to run.

Supplementary Figure S1) we generated samples with high intra-cluster correlation.

The following steps (represented graphically in Figure 1) were used to generate sparse and non-independent age–length samples:

(1) Annual sample sizes (n_k) were set equal to the actual sample sizes of aged fish available for western ABFT (last row of Table 1).

(2) For each year in which $n_k > 0$, individual fish from the observed age–length data were split into six non-overlapping clusters of unequal sizes using a K-means clustering algorithm (Hartigan and Wong, 1979) based on fish length. This algorithm partitions fish into K groups based on length such that the sum of squares from points to the assigned cluster means is minimized. The three clusters with the lowest cluster means were termed the "small fish" (SM) group, and the three clusters with the highest cluster means were termed the "large fish" (LG) group.

(3) For each year, we calculated the number of fish aged as a percent of annual catch (termed ψ_k) and, from that metric, created the following rule for selecting clusters to be sampled:

   (a) If $\psi_k < 0.0001\%$: one cluster was randomly sampled from each of the SM and LG groups.

   (b) If $0.0001\% \leq \psi_k < 0.001\%$: two clusters were randomly sampled from each of the SM and LG groups.

   (c) If $\psi_k \geq 0.001\%$: all six clusters were sampled.

While leaving entire portions of the length spectrum unsampled may seem extreme, it is fairly realistic for the case

of western ABFT. For example, in 1976, 68 fish were aged ($0.0001\% \leq \psi_{1976} < 0.001\%$) with no samples falling below 95 cm FL or in the 165–212 cm FL range.

(4) To create high intra-sample correlation, fish present within each cluster sampled were ordered by size and, after selecting the first fish randomly, all subsequent fish were sampled (without replacement) with probabilities proportional to the inverse difference in lengths between observations and the first fish sampled. Unequal sizes for each age–length sample were devised as follows, where n_k^{SM} and n_k^{LG} represent the sample sizes of fish aged from the small fish group and the large fish group, respectively, in year k:

$$n_k^{\text{SM}} = \sum_{z=1}^{3} q_{c,z} n_k w \qquad (7\text{a})$$

$$n_k^{\text{LG}} = \sum_{z=1}^{3} q_{c,z} n_k (1 - w) \qquad (7\text{b})$$

with

$$n_k = n_k^{\text{SM}} + n_k^{\text{LG}} \qquad (8)$$

where c is the number of clusters sampled in year k, w is the proportion of small fish in the sample (for the base case scenario, the sample is balanced between large and small fish, thus $w = 0.5$; this number is later changed in alternative scenarios 2 and 3, detailed in the next section, when samples are purposely skewed towards small and large fish to mimic

the sample availability of eastern and western ABFT, respectively) and $q_{c,z}$ is a randomly selected (without replacement) fraction used to create uneven samples in each cluster ($q_{1,z} \in \left\{0, 0, 1\right\}$ in the case where one cluster from the LG and SM groups are selected, $q_{2,z} \in \left\{0, \frac{1}{3}, \frac{2}{3}\right\}$ in the case where two clusters from the LG and SM groups are selected, and $q_{3,z} \in \left\{\frac{1}{6}, \frac{2}{6}, \frac{3}{6}\right\}$ in the case where three clusters from the LG and SM groups are selected). The $q_{c,z}$ values for the small-fish clusters are selected independently of the $q_{c,z}$ values for the large-fish clusters. The z subscript is simply there to indicate the order in which the fraction was picked: the first fraction to be selected has for subscript $z = 1$, the second $z = 2$, and the third $z = 3$.

(5) A normally distributed error term for each fish x, $\varepsilon_{A,x} \sim N(0, \sigma_A^2)$, was added to the true age of individual fish (A_x) to simulate ageing error where

$$\sigma_A = A \cdot CV \tag{9}$$

where the coefficient of variation of ageing error (CV) was assumed constant across ages and set to 10% to mimic the threshold error rate used for accepting age readings in bluefin tuna (Busawon *et al.*, 2015).

Base case and alternative scenarios

Eight scenarios were explored as simulations. For each scenario, a single true population was simulated, from which 100 different observed populations (and associated age–length samples) were generated.

Scenario 1—Base case. All dynamics match those described in the above section.

Scenario 2—Mainly small fish. Age data sampling is skewed towards smaller fish. w, the proportion of small fish found in the sample, defined above, is set to 0.7 instead of the value of 0.5 used in the base case.

Scenario 3—Mainly large fish. Age data sampling is skewed towards larger fish. w is set to 0.3 instead of the value of 0.5 used in the base case.

Scenario 4—Large recruitment variability. Recruitment variability is magnified by calculating the average recruitment over the 42-year time series of observations ($\overline{N}_0$) and inflating, by 50%, the size of the recruitment deviate in each year k:

$$N'_{0,k} = N_{0,k} + \frac{1}{2}\ \left(N_{0,k} - \overline{N}_0\right) \tag{10}$$

where $N'_{0,k}$ is the new recruitment value for year k.

Scenario 5—Small decrease in mean size-at-age over time. Mean size-at-age ($L_{i,k}$) is assumed to have been 10% higher at the beginning of the time series ($k = 1974$) compared with modern days ($k = 2015$), thus the new mean size-at-age i in year k ($L'_{i,k}$) is calculated as:

$$L'_{i,k} = L_{i,2015} + \frac{1}{10} \cdot \frac{2015 - k}{2015 - 1974} L_{i,2015} \tag{11}$$

Scenario 6—Large decrease in mean size-at-age over time. Mean size-at-age ($L_{i,k}$) is assumed to have been 20% higher

at the beginning of the time series compared with modern days. The new mean size-at-age i in year k ($L'_{i,k}$) is calculated following Equation (11) but with a factor of 1/5, replacing 1/10.

Scenario 7—Large measurement error in recorded lengths. A higher rate of measurement error in the observed catch-at-size data. σ_L^2 was increased to 100 cm from the 25 cm used in the base case scenario.

Scenario 8—Additional age data available. Ten additional years of age data were simulated to explore how each method's performance is expected to change as additional, more representative data become available in the future. Recruitment values and associated fishing mortality rate vectors were randomly sampled (with replacement) from the most recent 20 year period (1996–2015) to populate the 10 year projection. One thousand age–length records were generated for each year beyond 2015.

Performance metrics

Performance was measured using the RMSE. For each age and year combination, the RMSE associated with the proportion-at-age estimates for any given method and scenario was given by

$$\text{RMSE}_{i,k} = \sqrt{\frac{1}{100} \sum_{l=1}^{100} (\hat{p}_{i,k,l} - p_{i,k})^2} \tag{12}$$

where $p_{i,k}$ is the true proportion at age i in year k and $\hat{p}_{i,k,l}$ is an estimate of it from the lth run ($l = 1, 2, \ldots, 100$) of a given scenario. The smaller the RMSE the more accurate our estimate of $p_{i,k}$.

RMSE values were then collapsed over years ($\text{RMSEage} = \sqrt{\sum_k \text{RMSE}_{i,k}^2}$) and ages ($\text{RMSEyear} = \sqrt{\sum_i \text{RMSE}_{i,k}^2}$), as well as both years and ages ($\text{RMSEtot} = \sqrt{\sum_k \sum_i \text{RMSE}_{i,k}^2}$) to produce summary performance metrics for each estimation method and scenario.

To quantify the overall performance of FIAL and HY relative to CS, we calculated the percent gain in efficiency for each method, in each scenario. The calculation is analogous to that defined by Cochran (1977) for variances:

$$\%E = 100 \frac{\text{MSEtot}^{CS} - \text{MSEtot}^{\Omega}}{\text{MSEtot}^{\Omega}} \tag{13}$$

where MSEtot^{CS} is the mean squared error associated with CS and MSEtot^{Ω} is the mean squared error associated with either one of the alternative estimation methods ($\Omega \in \{\text{FIAL, HY}\}$).

To formally test whether one method outperformed the other, an additional metric was defined:

$$\text{RMSEtot}_l^{\Omega} = \sqrt{\sum_k \sum_i (\hat{p}_{i,k,l}^{\Omega} - p_{i,k}^{\Omega})^2} \tag{14}$$

where $\text{RMSEtot}_l^{\Omega}$ is the component of RMSEtot associated with the lth run and method Ω ($\Omega \in \{\text{CS, HY, FIAL}\}$). For each scenario, a pairwise comparison of RMSEtot_l^{CS} and RMSEtot_l^{HY}, and a pairwise comparison of RMSEtot_l^{CS} and RMSEtot_l^{FIAL}, were made to count the number of runs for which RMSEtot_l^{HY} and

RMSEtot$_l^{FIAL}$ were smaller than RMSEtot$_l^{CS}$, respectively. We determined significance (at the $\alpha = 0.05$ level) using a two-sided sign test.

Application to real data

A total of 4283 age–length samples (99.9% otoliths, 0.1% spines) collected in the western Atlantic was used for this analysis, with the earliest samples dating back to 1974 (Table 1). These samples comprised a mixture of eastern- and western-origin fish, which was not an issue for this study because the objective was to characterize the age composition of the catches from the western Atlantic rather than the age composition of the western stock. All samples were aged following the standardized reading protocol and ages were adjusted for proper year class assignment (Luque et al., 2014; Busawon et al., 2015; Rodriguez-Marin et al., 2016). In the samples, only 5% of the fish had sizes that were directly measured as straight fork length; for the remaining fish, measurements were obtained from converted length (i.e. curved fork length and snout length) or weight measurements. As in the 2017 ICCAT assessment, apparent outliers (39 records of fish with sizes falling beyond 3 standard deviations of the mean of the sample for each age) were removed for the analysis as they were thought to be unrealistic and could have a negative impact on the estimation process.

The official catch-at-size data used in the 2017 assessment was used as an input for the combined forward-inverse ALK analysis. ICCAT has devoted a great deal of effort to try to correct for biases in the catch-at-size data, including some imputation of missing length data for certain years (Shemla and McAllister, 2006). These records were therefore assumed to be the best information available to date for bluefin tuna. Because of the lack of samples of fish of age 0, $P(j|i = 0)$ was fixed to probabilities computed using the mean and standard deviation of size-at-age 0 obtained from Ailloud et al. (2017) growth curve equation and assuming normality in the distribution of size-at-age. The FIAL algorithm was run multiple times with different starting values to check for convergence.

Results
Simulation

One hundred simulation runs was judged sufficient to guarantee stability of the performance metrics (see Supplementary Figure S2a–c). Overall, FIAL and HY outperformed CS across all eight scenarios (Figure 2). FIAL performed best, with the lowest RMSEtot, followed by HY and CS (Figure 2). Results from the sign test confirm this: values for RMSEtot$_l^{FIAL}$ and RMSEtot$_l^{HY}$ were found to be significantly smaller than RMSEtot$_l^{CS}$ in all eight scenarios (all P values <0.001). Depending on the scenario, FIAL was 52–451% more efficient than CS, while HY was 11–21% more efficient than CS (Table 2). The difference in performance between FIAL and CS was most pronounced in scenario 8, where additional years of age data brought considerable improvements to the performance of FIAL, as well as in scenarios 3 and 4, the scenarios in which the age sample is skewed towards larger individuals and recruitment variability is inflated by 50%, respectively (Table 2). These three scenarios were where FIAL performed best, both relative to the other methods and relative to other scenarios. The difference in performance between FIAL and CS was least pronounced in scenarios 6 and 7 (Table 2). These were the scenarios where the population experienced large
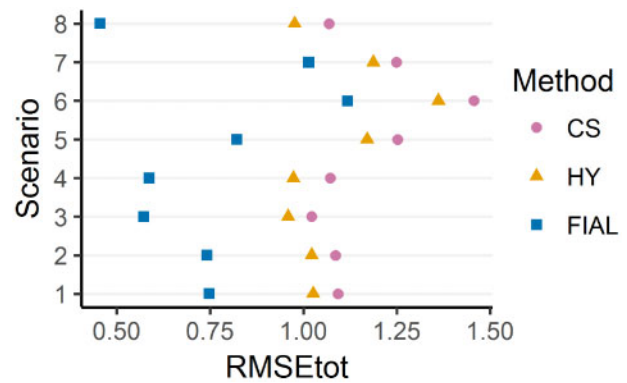

Figure 2. RMSEtot results across methods (CS, HY, and FIAL) and scenarios (1 through 8).

changes in mean size-at-age through time (scenarios 6) and where large observation errors were added to the catch-at-length data (scenario 7).

FIAL either outperformed or performed similarly to CS and HY across most age groups (Figure 3 and Supplementary Figure S3a–h). The greatest differences in performance between the methods were observed in the younger age groups (ages 2–4, which contribute a large portion of the total catch) and in the plus group (Figure 3). CS tended to put large amounts of error in the plus group while the error in FIAL was split between the plus group and the age before the plus group (age 15; Figure 3). For FIAL, errors were lower and more evenly distributed among age groups in scenario 8, where additional years of age data were simulated.

RMSE values by year for each method and scenario are shown in Figure 4a and Supplementary Figure S4a–g. RMSE$_{year}$ values are, in most years, higher for CS than for HY and FIAL and show a more erratic pattern with CS. As expected, HY performs better than CS in years where age data are available. In scenarios 6 and 7, where large changes in mean size-at-age and large errors in the catch-at-size are simulated, all three methods perform poorly. The difference in performance between CS and FIAL is most pronounced in the earlier years, where the stock is experiencing very high levels of fishing mortality on very young ages (Figure 4a and Supplementary Figure S4a-h). FIAL performs slightly better when age samples are skewed towards older fish compared with when age samples are skewed towards smaller fish (Figure 4a and Supplementary Figure S4a). FIAL performs considerably better with additional years of age data (scenario 8; Supplementary Figure S4g).

Application to real data

CS and FIAL were applied to the western ABFT catch-at-size data from 1974 to 2015. With the FIAL analysis, different starting values for the parameters were used for each run, and runs with reasonably low final maximum gradient component (<0.1) were retained. The algorithm showed difficulty converging to a consistent global minimum across trials (Supplementary Figure S5). Estimates of $\hat{P}(i)_k$ from the top 5 runs showed nearly identical results, suggesting the best result is likely close to or at the global minimum (Supplementary Figure S6). Estimates of $\hat{P}(j|i)$ (i.e. the inverse key) for the best run are shown in Supplementary Figure S7. Mean sizes at age calculated from the inverse key revealed a

slight discrepancy among the mean-size-at-age of older ages: mean size-at-age 15 was found to be slightly larger than the mean size-at-age 16+ (Supplementary Figure S7).

There was evidence of both strong and weak cohorts moving through the catch in the FIAL results (Table 3). Estimates of catch-at-age derived from the two methods were plotted against one another and presented in Supplementary Figure S7. CS and FIAL were often found to be a year off from each other

in characterizing the origin of strong year classes. For example, in 1975 and 1976, a strong 1973 cohort was clearly apparent in the CS results while that peak was attributed to a 1972 cohort in the FIAL results. Similarly, in 2007–2009 CS identified a strong 2003 cohort while FIAL interpreted it to be a strong 2002 cohort (Supplementary Figure S8; Table 3).

Discussion

With simulated data designed to emulate several real world complexities, the FIAL key performed significantly better than the other two methods. The method also provided useful results when applied to the ABFT dataset, albeit with some difficulties in achieving convergence. The FIAL key was able to track strong cohorts and this led to the discovery of a systematic ageing error.

The fact that convergence was more difficult with the real dataset than the simulated datasets provided some indication that the simulated data may not capture the full degree of idiosyncrasies contained in the actual data, such as time-varying or seasonal growth. Looking at length-at-age distributions in the real dataset revealed evidence of bimodality in certain years, which could have biological relevance, or could simply be a result of

Table 2. Percent relative efficiency (%E) of HY and FIAL compared with CS.

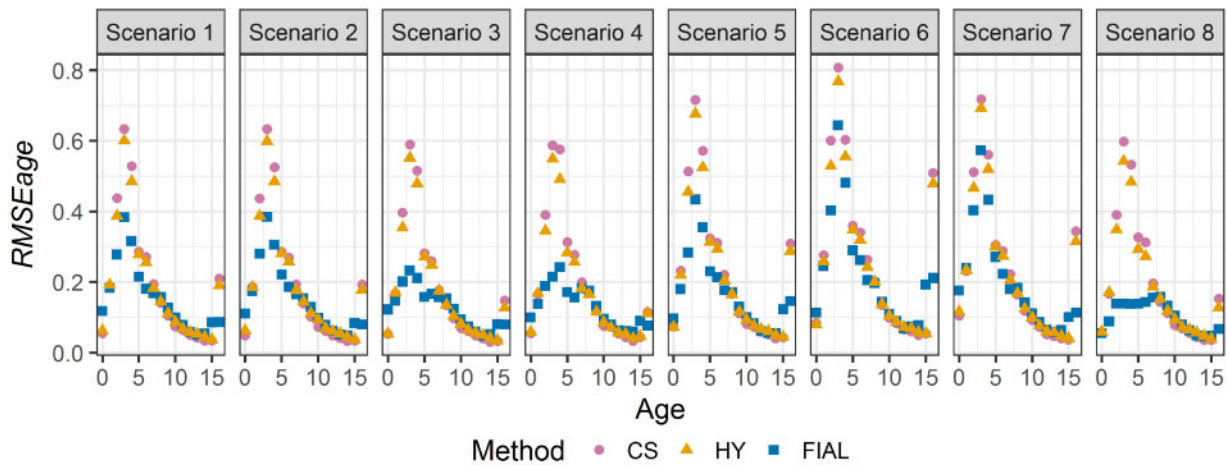| Scenario	%E_{HY}	%E_{FIAL}
1	13	114
2	13	114
3	14	219
4	21	234
5	15	133
6	14	70
7	11	52
8	20	451

Figure 3. RMSEage results across methods (CS, HY, and FIAL) and scenarios (1 through 8).
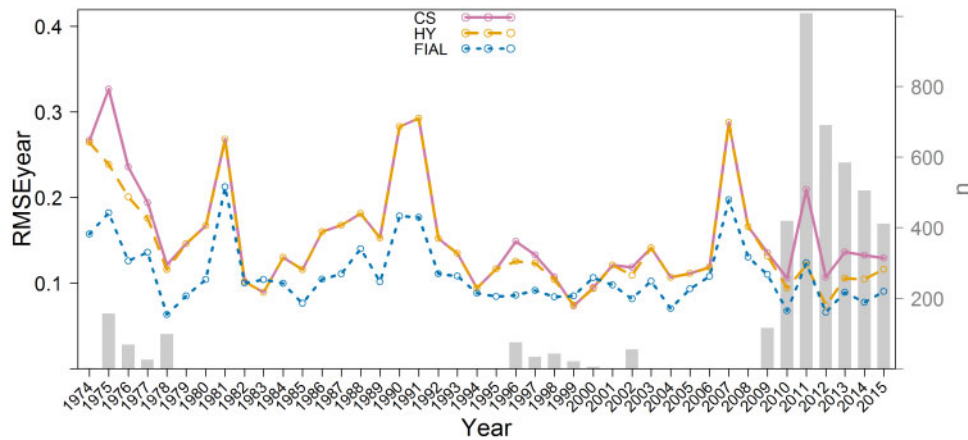

Figure 4. RMSE values by year and estimation method (lines) resulting from scenario 1 (base case) plotted against otolith sample sizes (*n*) available for that year (grey histogram).

Table 3. Catch-at-age estimates (in numbers) resulting from the FIAL analysis applied to western Atlantic bluefin tuna data.

Year/Age	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16+
1974	38	12232	57133	72101	87	1296	290	5	189	399	1491	3635	104	3488	6946	833	695
1975	2	2938	44858	170357	337	1250	227	115	37	106	1634	1383	1135	3562	3471	457	712
1976	0	855	10362	37339	75381	246	449	246	342	22	647	82	3976	3570	5676	773	898
1977	29	266	1912	29912	8350	18986	25525	1512	77	29	231	225	343	2755	9026	1823	1113
1978	55	108	7507	11009	17694	12309	7931	9425	137	26	64	208	575	324	7327	3081	1547
1979	5	85	3945	12437	14909	4840	21531	111	3061	2170	10	38	87	4582	4525	3214	1654
1980	320	0	4164	14468	13397	9674	5452	1679	3077	8858	1482	31	38	421	5743	2952	945
1981	0	1024	6924	12545	14208	9765	5457	4795	3191	1680	4196	1646	43	76	148	1862	7256
1982	0	1952	2745	3838	1510	745	4	628	969	236	796	1584	145	10	10	1695	1182
1983	0	2918	1160	2668	2762	794	1091	1821	1308	1639	1828	751	1216	117	127	1135	2697
1984	28	197	1260	6933	2642	1664	3347	2004	2341	80	622	1202	1525	1118	27	649	1496
1985	5	872	57	6918	13761	431	4009	5722	2166	28	24	247	1705	2021	1039	13	1223
1986	565	0	2073	5686	8725	1659	8221	1550	1866	1735	709	57	672	1684	930	113	769
1987	65	1584	1442	14720	7196	7722	7806	3188	1839	1284	1261	235	402	1352	729	101	675
1988	69	3131	3659	10329	12676	3805	3351	5136	4851	1811	990	626	572	1191	884	261	732
1989	93	592	533	14494	1631	117	5365	1353	2705	3627	1788	1362	619	1227	751	674	475
1990	129	935	1909	3835	22397	1626	3776	1367	3142	1441	3254	699	534	899	1050	249	531
1991	111	310	4597	4259	21763	2337	1937	1675	3270	2397	3348	1800	523	1091	499	455	201
1992	98	0	640	3887	5129	218	1484	1726	2410	2280	1747	1731	1236	707	625	444	341
1993	133	7	443	1061	1569	7048	1804	2392	2755	1804	4643	212	332	307	641	739	227
1994	77	607	2198	213	1056	3017	2717	2155	1889	3312	2411	1515	225	414	292	587	201
1995	281	349	924	484	2672	4539	4357	4141	1816	1176	3200	1174	411	171	531	866	658
1996	205	238	548	11655	1209	3196	5280	2408	836	3383	576	2325	563	1045	365	599	592
1997	75	0	322	689	7400	999	1458	1811	2164	2590	2819	143	1259	932	597	683	328
1998	54	18	475	789	3601	4133	1301	785	2605	2645	5445	703	707	881	195	641	525
1999	64	0	136	486	1484	1751	2275	3640	1696	2786	2575	3905	807	114	19	26	1668
2000	49	0	135	276	962	395	2308	3654	3076	1467	2609	1527	2105	34	9	18	2410
2001	13	0	1814	3	1323	5214	2712	695	2753	4171	562	1748	2104	1080	188	35	1470
2002	0	0	1119	6024	4636	2651	8096	2833	908	4400	4207	514	1340	1689	294	142	1083
2003	0	0	214	2309	3465	4952	2674	1675	517	1623	4330	438	206	1138	545	117	989
2004	0	0	856	2636	5255	5082	3975	1782	3004	1084	1552	1344	494	110	739	369	588
2005	0	652	242	5718	3100	2687	574	854	1175	627	1087	2083	991	115	542	507	560
2006	0	0	253	401	1347	1298	3006	2299	650	1665	1358	1076	626	686	116	698	712
2007	0	0	87	118	3613	15759	168	2296	1556	558	497	544	471	175	86	200	1040
2008	0	0	96	825	1539	803	7776	3507	309	2875	1251	567	430	151	43	392	1313
2009	0	0	79	94	1376	2162	767	5155	4682	169	1389	695	388	550	483	275	811
2010	0	0	72	1269	573	1515	1738	608	1070	2837	1653	1553	399	563	634	241	574
2011	0	0	7	490	762	3189	1230	2490	790	4870	1628	700	553	404	422	141	778
2012	0	35	81	175	1793	1604	979	265	1494	951	3461	1121	520	361	502	68	734
2013	0	13	36	128	339	1484	383	581	370	1537	932	2084	819	350	360	41	759
2014	0	10	90	1245	110	963	582	487	224	388	1801	1533	1980	552	285	119	617
2015	0	0	1	31	1390	252	225	427	602	974	536	2115	1727	1235	508	241	657

Lighter shades indicate lower catches and darker shades indicate higher catches. A strong 2002 cohort is clearly apparent (outlined in black).

observation errors in the recorded ages or lengths. The observed bimodalities are likely to exacerbate convergence problems as they blur the distinction between the size distributions of adjacent age classes. Similarly, the inconsistency in the mean sizes at age, where the mean size-at-age 15 was estimated to be slightly higher than the mean size-at-age 16+, is likely to cause convergence problems. This issue could be resolved by sampling additional large fish, which are greatly needed to adequately characterize probability of size-at-age over the oldest age groups and which can lead to greater accuracy overall (as was apparent in the simulation results for scenario 3). Testing alternative bin lengths or perhaps even exploring the use of unequal bin sizes across lengths may also allow for increased accuracy and precision in the estimated probabilities of size given age in older fish.

For the application to the real data, parameters associated with the probabilities at size for age 0 fish had to be fixed because there were no age 0 fish in the sample. It is important that data be collected on age 0 fish so that these parameters can be estimated. While age 0 fish are present in the historical catches, this was before age data was being actively collected. Today, the purse seine fishery that used to target age 0 fish is no longer operating in the Atlantic owing to minimum size restrictions, so although annual age collection is in place it has proven very difficult to obtain samples of very small fish.

If size-at-age is suspected to have changed through time, our simulation showed that all three estimation methods would be negatively affected. Attempts to uncover and characterize significant temporal changes in size-at-age in western ABFT have met with difficulty (Siskey et al., 2016). With the fishery having shifted from historically targeting very small fish to targeting medium to large fish in more recent years, it is difficult to conduct a statistically robust comparison of size-at-age between different time

periods. That being said, in the near future, as data collection improves, the assumption of no variation in size-at-age could be relaxed by beginning to use forward keys in concert with the FIAL key. Forward keys do not make any assumptions about variation in size-at-age; thus, as representative annual samples become available, forward keys could be used to estimate age composition in the most recent years, whereas the FIAL key could continue to be used to estimate age composition in historical years.

The larger concern that came out of this exercise was the confusion over the birth year of the strong cohort seen moving through the fishery in recent years: the FIAL key analysis pointed to a strong 2002 cohort, yet ageing experts working on eastern ABFT were of the opinion that the 2003 year class was the strong one. The main difference between age samples from the eastern and western Atlantic was the type of structure being aged. Western Atlantic samples were mainly composed of otoliths (99%) while samples from the eastern Atlantic were mainly composed of spines (90%). Thus, this brought us to questioning whether the differing signals observed in the East and the West were indicative of a true difference in recruitment history or whether it was simply the product of a difference in methodology. Paired otolith-spine samples (i.e. samples taken from the same fish; available from Rodriguez-Marin *et al.*, 2016) revealed that age readings from otoliths were, on average, slightly higher than the corresponding age readings from spines. Age estimates from spine readings are thought to be more reliable than age readings from otolith samples in young ABFT (Dr Rodriguez-Marin, personal communication). That is because the otoliths of young ABFT often contain visible false annuli (i.e. bands that were not deposited on an annual basis) that can easily be misinterpreted as being annual and thus result in overestimated ages. Beyond age 7, spines are considered less precise than otoliths as the innermost rings begin to resorb (Rooker *et al.*, 2007). It therefore appears that the differences observed between East and West stem from a difference in methodology rather than a difference in recruitment. A more thorough evaluation of this problem is needed to settle this issue.

While the solution to sparse age data might be to move exclusively to integrated statistical catch at length models—such as Stock Synthesis (Methot and Wetzel, 2013), CASAL (Bull *et al.*, 2012), or SCAL (Butterworth and Rademeyer, 2015) and others (ASAP—Legault and Restrepo, 1998; BAM—Williams and Shertzer, 2015)—each still has to make some basic assumptions about size-at-age often similar to the FIAL key, and the idiosyncrasies of working with real data, such as lower mean sizes at older ages and potential time-varying process error observed in ABFT, can be just as problematic for the more complicated integrated catch at length approaches as they are for the simple ALK approaches. Moreover, one downside of using integrated analyses is that it then becomes difficult to tease out inconsistencies among sources of data and to evaluate failures of assumption and their effects on the model outputs. In the case of ABFT, by looking at just the age–length data, one could see an inconsistency in ageing among calcified hard parts; this likely would have been overlooked in the results from an integrated model. Using the FIAL key alongside other more sophisticated integrated models can therefore provide valuable insight into the behaviour of more complex, integrated models.

Conclusion

The FIAL key outperformed the other ageing methods with the simulated data and it improved age composition estimates of western ABFT. Two main concerns should be addressed for the model to be used operationally: (i) issues with age assignment between hard part types must be resolved, and (ii) young-of-the-year fish must be sampled so that the probability of size-at-age 0 can be estimated. In addition, to reduce age composition bias, scientists and fishers could make a concerted effort to improve the representativeness of the sampling, and therefore the data available to assess the stock. Annual data collection efforts must prioritize maximizing sampling coverage across sizes (particularly very small and very large fish), space and time, and collection should strive to follow a robust length stratified sampling design whenever possible. Any bias and imprecision in the catch-at-size estimates will also reduce the precision of the catch-at-age estimates, regardless of the estimation method used to obtain age composition. It is therefore equally important to ensure that the length sampling is as high quality as possible (i.e. representative samples with large effective sample size). Lastly, efforts to characterize the stock origin of each age–length sample is also underway and should be continued as it will allow scientists to disentangle the origin of strong year class signals, which is crucial to determining accurately the productivity potential of the stocks.

Supplementary data

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

Acknowledgements

References

Ailloud, L. E., and Hoenig, J. M. 2019. A general theory of age-length keys: combining the forward and inverse keys to estimate age composition from incomplete data. ICES Journal of Marine Science, 76: 1515–1523.

Ailloud, L. E., Smith, M. W., Then, A. Y., Omori, K. L., Ralph, G. M., and Hoenig, J. M. 2015. Properties of age compositions and mortality estimates derived from cohort slicing of length data. ICES Journal of Marine Science, 72: 44–53.

Ailloud, L. E., Lauretta, M. V., Hanke, A. R., Golet, W. J., Allman, R. J., Siskey, M. R., Secor, D. H. *et al*. 2017. Improving growth estimates for western Atlantic bluefin tuna using an integrated modeling approach. Fisheries Research, 191: 17–24.

Anon. 2014. Bluefin tuna biological sampling program: commercial and recreational fisheries. Collective Volume of Scientific Papers, 70: 394–395. https://www.iccat.int/Documents/CVSP/CV070_2014/n_2/CV070020394.pdf (last accessed 17 February 2019).

Block, B. A., Teo, S. L., Walli, A., Boustany, A., Stokesbury, M. J., Farwell, C. J., Weng, K. C. et al. 2005. Electronic tagging and population structure of Atlantic bluefin tuna. Nature, 434: 1121.

Bull, B., Francis, R. I. C. C., Dunn, A., McKenzie, A., Gilbert, D. J., Smith, M. H., Bian, R., and Fu, D. 2012. CASAL (C++ algorithmic stock assessment laboratory): CASAL user manual v2.30-2012/03/21. NIWA Technical Report 135. http://docs.niwa.co.nz/library/public/NIWAtr135.pdf (last accessed 18 June 2018).

Busawon, D. S., Neilson, J. D., Andrushchenko, I., Hank, A., Secor, D. H., and Melvin, G. 2014. Evaluation of Canadian sampling program for bluefin tuna, results of natal origin studies 2011–2012 and assessment of length-weight conversions. ICCAT Collective Volume of Scientific Papers, 70: 202–219. https://www.iccat.int/Documents/CVSP/CV070_2014/n_1/CV070010202.pdf (last accessed 17 February 2019).

Busawon, D. S., Rodriguez-Marin, E., Luque, P. L., Allman, R., Gahagan, B., Golet, W., Koob, E., Siskey, M., Ruiz, M., Quelle, P., and Neilson, J. 2015. Evaluation of an Atlantic bluefin tuna otolith reference collection. ICCAT Collective Volume of Scientific Papers Collective Volume of Scientific Papers, 71: 960–982. https://www.iccat.int/Documents/CVSP/CV071_2015/n_2/CV071020960.pdf (last accessed 18 June 2018).

Butterworth, D. S., and Rademeyer, R. A. 2015. An updated statistical catch-at-length assessment for Western Atlantic Bluefin Tuna. ICCAT Collective Volume of Scientific Papers, 71: 1813–1831.

Carlsson, J., McDowell, J. R., Carlsson, J. E., and Graves, J. E. 2006. Genetic identity of YOY bluefin tuna from the eastern and western Atlantic spawning areas. Journal of Heredity, 98: 23–28.

Cochran, W. G. 1977. Sampling Techniques, 3rd edition. John Wiley, New York.

Dickhut, R. M., Deshpande, A. D., Cincinelli, A., Cochran, M. A., Corsolini, S., Brill, R. W., Secor, D. H. et al. 2009. Atlantic bluefin tuna (*Thunnus thynnus*) population dynamics delineated by organochlorine tracers. Environmental Science & Technology, 43: 8522–8527.

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. Conseil International pour l'Éxploration de la Mer, 86: 1–14.

Fromentin, J. M., and Powers, J. E. 2005. Atlantic bluefin tuna: population dynamics, ecology, fisheries and management. Fish and Fisheries, 6: 281–306.

Goodyear, C. P. 1987. Status of the Red Drum Stocks of the Gulf of Mexico. USDOC, NMFS, SEFC, Miami Laboratory Contribution CRD 86/87-34. 113 pp. https://grunt.sefsc.noaa.gov/P_QryLDS/download/CRD33_CRD-86_87-34.pdf? id=LDS (last accessed 18 June 2018).

Hartigan, J. A., and Wong, M. A. 1979. Algorithm AS 136: a k-means clustering algorithm. Journal of the Royal Statistical Society: Series C (Applied Statistics), 28: 100–108.

Hirst, D., Storvik, G., Rognebakke, H., Aldrin, M., Aanes, S., and Vølstad, J. H. 2012. A Bayesian modelling framework for the estimation of catch-at-age of commercially harvested fish species. Canadian Journal of Fisheries and Marine Science, 69: 2064–2076.

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.

ICCAT. 2017. Report of the 2017 Atlantic bluefin tuna stock assessment meeting. International Commission for the Conservation of Atlantic Tunas. ICCAT Collective Volume of Scientific Papers, 74: 2372–2535. https://iccat.int/Documents/Meetings/Docs/2017_BFT_ASS_REP_ENG.pdf (last accessed June 2018).

Justel-Rubio, A., and Ortiz, M. 2013. Review and preliminary analyses of size frequency samples of bluefin tuna (*Thunnus* thynnus) 1952–2010. ICCAT Collective Volume of Scientific Papers, 69: 297–330. https://www.iccat.int/Documents/CVSP/CV069_2013/n_1/CV069010297.pdf (last accessed February 2019).

Kell, L. T., and Kell, A. 2011. A comparison of age slicing and statistical age estimation for Mediterranean Swordfish (*Xiphias gladius*). ICCAT Collective Volume of Scientific Papers, 66: 1522–1534.

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.

Legault, C. M., and Restrepo, V. R. 1998. A flexible forward age-structured assessment program. ICCAT Collective Volume of Scientific Papers, 49: 246–253.

Luque, P. L., Rodriguez-Marin, E., Landa, J., Ruiz, M., Quelle, P., Macias, D., and Ortiz de Urbina, J. M. 2014. Direct ageing of *Thunnus thynnus* from the eastern Atlantic Ocean and western Mediterranean Sea using dorsal fin spines. Journal of Fish Biology, 84: 1876–1903.

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.

Mohn, R. K., and Savard, L. 1989. Length Based Population Analysis of Sept-Iies Shrimp (Gulf of St Lawrence). NAFO SCR-Res Doc, 89/92. 14 pp. https://archive.nafo.int/open/sc/1989/SCR-89-092.PDF (last accessed 19 February 2019).

Mohn, R. 1994. A comparison of three methods to convert catch at length data into catch-at-age. ICCAT Collective Volume of Scientific Papers, 42(1): 110–119.

Ortiz, M., and Palma, C. 2011. Summary of comparison and verification of the AgeIT program for age-slicing of bluefin tuna catch-at-size (CAS) information. ICCAT Collective Volume of Scientific Papers, 66: 918–934.

Restrepo, V. R. 1995. Application of cohort slicing and tuned VPA to simulated data that includes variability in length at age. ICCAT Collective Volume of Scientific Papers, 44: 67–71.

Richardson, D. E., Marancik, K. E., Guyon, J. R., Lutcavage, M. E., Galuardi, B., Lam, C. H., Walsh, H. J. et al. 2016. Discovery of a spawning ground reveals diverse migration strategies in Atlantic bluefin tuna (*Thunnus thynnus*). Proceedings of the National Academy of Sciences of the United States of America, 113: 3299–3304.

Rodriguez-Marin, E., Ortiz, M., de Urbina, J. M. O., Quelle, P., Walter, J., Abid, N., Addis, P. et al. 2015. Atlantic bluefin tuna (*Thunnus thynnus*) biometrics and condition. PLoS One, 10: e0141478.

Rodriguez-Marin, E., Quelle, P., Ruiz, M., Busawon, D., Golet, W., Dalton, A., and Hanke, A. 2016. Updated Comparison of Age Estimates From Paired Calcified Structures From Atlantic Bluefin Tuna. ICCAT SCRS/2016/134. http://www.repositorio.ieo.es/e-ieo/handle/10508/10663 (last accessed 18 June 2018).

Rooker, J. R., Alvarado Bremer, J. R., Block, B. A., Dewar, H., De Metrio, G., Corriero, A., Kraus, R. T. et al. 2007. Life history and stock structure of Atlantic bluefin tuna (*Thunnus thynnus*). Reviews in Fish Biology and Fisheries, 15: 265–310.

Rooker, J. R., Secor, D. H., De Metrio, G., Schloesser, R., Block, B. A., and Neilson, J. D. 2008. Natal homing and connectivity in Atlantic bluefin tuna populations. Science, 322: 742–744.

Shemla, A., and McAllister, M. K. 2006. Bayesian generalized linear models to standardize and impute missing data in the Atlantic bluefin tuna (*Thunnus thynnus thynnus*) TaskII catch and effort database. ICCAT Collective Volume of Scientific Papers, 59: 750–768.

Siskey, M. R., Wilberg, M. J., Allman, R. J., Barnett, B. K., and Secor, D. H. 2016. Forty years of fishing: changes in age structure and

stock mixing in northwestern Atlantic bluefin tuna (*Thunnus thynnus*) associated with size-selective and long-term exploitation. ICES Journal of Marine Science, 73: 2518–2528.

Walter, J. F., Porch, C. E., Lauretta, M. V., Cass-Calay, S. L., and Brown, C. A. 2016. Implications of alternative spawning for bluefin tuna remain unclear. Proceedings of the National Academy of Sciences of the United States of America, 113: E4259–E4260.

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.

Williams, E. H., and Shertzer, K. W. 2015. Technical documentation of the Beaufort Assessment Model (BAM). NOAA Technical Memorandum MNFS-SEFSC-671, doi: 10.7289/V57M05W6. ftp://ftp.library.noaa.gov/noaa_documents.lib/NMFS/SEFSC/TM_NMFS_SEFSC/NMFS_SEFSC_TM_671.pdf (last accessed April 2019).

Wilson, S. G., Jonsen, I. D., Schallert, R. J., Ganong, J. E., Castleton, M. R., Spares, A. D., Boustany, A. M. *et al.* 2015. Tracking the fidelity of Atlantic bluefin tuna released in Canadian waters to the Gulf of Mexico spawning grounds. Canadian Journal of Fisheries and Aquatic Sciences, 72: 1700–1717.

Handling editor: Francis Juanes