ICES Journal of Marine Science, 61: 798–811 (2004)
doi:10.1016/j.icesjms.2004.03.002

Estimation of birthdates and catch-at-age using length frequency analysis (LFA), with application for skipjack tuna (*Katsuwonus pelamis*) caught in the Southwest Atlantic

Humber A. Andrade and Paul G. Kinas

Andrade, H. A., and Kinas, P. G. 2004. Estimation of birthdates and catch-at-age using length frequency analysis (LFA), with application for skipjack tuna (*Katsuwonus pelamis*) caught in the Southwest Atlantic. – ICES Journal of Marine Science, 61: 798–811.

Converting length frequencies into age frequencies is an important component of a fisheries assessment. In this paper we use a length frequency analysis (LFA) to estimate birthdates after converting length data into catch-at-age, and use simulation studies to compare model-selection criteria and to examine the reliability of the resulting estimates. Deviance and an adaptation of the Akaike Information Criterion performed best. LFA results in useful estimates of birthdates and of catch-at-age if reliable length frequency data and estimates of growth parameters are available. The analysis is applied to skipjack tuna (*Katsuwonus pelamis*) caught in the Southwest Atlantic Ocean. Although spawning is reported to be seasonal in subtropical waters, the birthdates of the fish caught there were spread uniformly across the year. Young skipjack become vulnerable to fishing mainly in the first quarter of each year. Recruitment of strong year classes did not affect fishery yields equally in the Southwest Atlantic and the Caribbean Sea, so the assumption of a unit western stock for management purposes and the stock structure of skipjack in the Atlantic need further evaluation.

Keywords: length frequency analysis, recruitment, skipjack tuna.

Received 5 August 2003; accepted 10 March 2004.

H. A. Andrade: UNIVALI-CTTMar, Rua Uruguai 458, C.P. 360, 88302202, Itajaí-SC, Brazil. P. G. Kinas: FURG-Depto. Matemática, Av. Itália s/n, C.P. 474, 96201900 Rio Grande-RS, Brazil. Correspondence to H. A. Andrade: tel: +55 47 3417714; fax: +55 47 3417715; e-mail: humber@cttmar.univali.br.

Introduction

Decomposing length frequency histograms into age classes using length frequency analysis (LFA) used to be *ad hoc* (Ricker, 1975), but it has evolved into a formalized procedure by extending it to the theory of mixtures of probability distributions (Hasselblad, 1966), usually of Gaussian form. In one of the first applications using a coherent statistical approach (e.g. likelihood), McNew and Summerfelt (1978) found it hard to produce consistent estimates where growth rates varied over the period analysed. They also had some difficulty dealing with the unknown number of age groups and with the estimation of too many parameters, including a standard deviation for each age class. More realistic estimates were obtained by MacDonald and Pitcher (1979), who modified the procedure by forcing constraints on parameter estimations.

Further improvements in LFA were made in the papers of Schnute and Fournier (1980) and Fournier and Breen (1983), by imposing some model structure and assumptions on growth, mortality, and standard deviations of length for each age group. Growth parameters can be poorly estimated from frequency data, resulting in incorrect age composition. Therefore, it is generally advisable not only to impose a growth structure (e.g. von Bertalanffy), but also to assume model parameters gathered from independent sources (Quinn and Deriso, 1999). Mortality assumptions proved to be useful sometimes (e.g. Fournier and Breen, 1983), but not in all situations (e.g. Johnson and Quinn, 1987).

In most cases, the addition of model structure enhances the LFA by increasing the speed of convergence in the iterative procedure, and by helping to obtain consistent solutions (e.g. to obtain the same estimates no matter what the starting values used in the iteration were; Fournier and Breen, 1983; Johnson and Quinn, 1987). In practice, estimates from LFA are reliable and useful only if auxiliary information (e.g. a growth model and parameters) is available (Quinn and Deriso, 1999).

Several other length-based procedures are presented in a volume edited by Pauly and Morgan (1987), and a further review of LFA applications was provided by Rosenberg and Beddington (1988). During the past two decades, LFA has proved to be useful in stock assessment with the help of software packages developed primarily to estimate growth parameters and age composition from length frequency data (i.e. ELEFAN – Pauly and David, 1981; MULTIFAN – Fournier *et al.*, 1990; MULTIFAN–CL – Fournier *et al.*, 1998). In most of these procedures, LFA is used to estimate growth parameters or catch-at-age tables to support the stock assessment. However, in this paper we extended LFA to back-calculate birthdates, by handling the average age of distinct age classes more freely than in traditional formulations, and by including simulation studies to check the reliability of the estimates.

We illustrate the proposed procedure by assessing a real question. Skipjack tuna (*Katsuwonus pelamis*) inhabiting the Atlantic Ocean have been divided by the International Commission for the Conservation of Atlantic Tunas (ICCAT) into two distinct stocks (ICCAT, 2000, 2003), a more productive eastern stock, and a western stock, which is the subject of the present work. Although there is a possibility of mixing, the hypothesis of separate East and West Atlantic stocks has been maintained as most plausible (ICCAT, 2003), motivated by:

(i) the large distances between fishing grounds;
(ii) environmental restrictions;
(iii) the existence of a spawning area in the East as well as in the West Atlantic;
(iv) the lack of additional evidence (e.g. no significant transatlantic migrations in the tagging data).

ICCAT also suggested that still smaller management units could be considered on the basis of biological characteristics of the species and the distinct areas in which fishing takes place.

The reproductive behaviour of Atlantic skipjack is usually assumed to be opportunistic, because mature oocytes are found throughout the year (Cayré and Farrugio, 1986; Goldberg and Au, 1986). The opportunistic behaviour associated with a highly migratory lifestyle would imply that skipjack could be in suitable oceanographic conditions to spawn partially several times a year. Therefore, pronounced seasonal variation in the birth rate and recruitment would not be expected, at least in equatorial waters, where environmental conditions favour continuous spawning activity (Cayré *et al.*, 1986).

Some authors have modelled recruitment as erratic over time, with no main recruitment season (e.g. Fromentin and Restrepo, 2001). In contrast, results of surveys of the West Atlantic stock, although supporting the hypothesis of continuous spawning throughout the year, suggest peak spawning from November to March (Kikawa and Nishikawa, 1980; Matsuura, 1982, 1986; Goldberg and Au, 1986). The theory of a main spawning season also supports the suggestion of Matsumoto *et al.* (1984) that skipjack are year-round spawners in the tropics and spring-to-autumn spawners in the subtropics. Most skipjack in the Southwest Atlantic are caught in the subtropics (south of 20°S; ICCAT, 2003), giving a basis for the seasonal oscillation in the gonadal activity observed by Goldberg and Au (1986) and egg and larva densities (Matsuura, 1982, 1986). Different spawning patterns in tropical and subtropical areas are also seen in other oceans (Schaefer, 2001; Stéquert *et al.*, 2001).

Currently available information suggests distinct reproductive and perhaps recruitment patterns for different fractions of the skipjack population in the Atlantic Ocean. Whereas recruitment should be assumed to be erratic for the tropical East Atlantic stock, a seasonal peak seems plausible for skipjack exploited in the subtropical Southwest Atlantic. It is important for the construction of assessment models and for fishery management decisions to determine whether reproductive (spawning and recruitment) pulses exist for the West Atlantic stock. In order to address this question, the ideal source of information would be reliable, high-precision catch-at-age data, such as those provided in some cases by daily ring studies of otoliths (e.g. Uchiyama and Struhsaker, 1981), which allow the back-calculation of birthdates. If birthdates were available, it would be easier to determine whether the spawning season supports the fishery. However, such precise catch-at-age data are not available for West Atlantic skipjack, so we investigate the issue by applying a proposed extension of LFA to length frequency distributions of reported landings. Ianelli (1993) also tried to do this, and it is his results that encouraged us to proceed with exploring this potential use of LFA.

Data and analyses

Fishing data

Data on the tuna fishery in the Atlantic are available from ICCAT. In the West Atlantic, skipjack are caught mainly in the north (Caribbean Sea), and off the southeastern coast of South America (Figure 1). Length frequency data were collected from May 1995 to December 2001 in Itajaí (26°55′S 48°40′W), the main fishing harbour on the Brazilian coast. We assume them to be representative of all skipjack caught in the Southwest Atlantic and landed in Brazil. However, despite length data being easy to obtain, there are at least three difficulties in obtaining length frequency distributions representative of the total catch in the fishery: the boats can fish in more than one area during a trip; skipjack form schools, and different schools can display different modal lengths; and while fish are offloaded to shore, knowledge of source area and school is scanty.

In order to investigate the bias that could be introduced into the data by the landing process, we plotted the standard error (s.e.) of the average length as a function of the sample

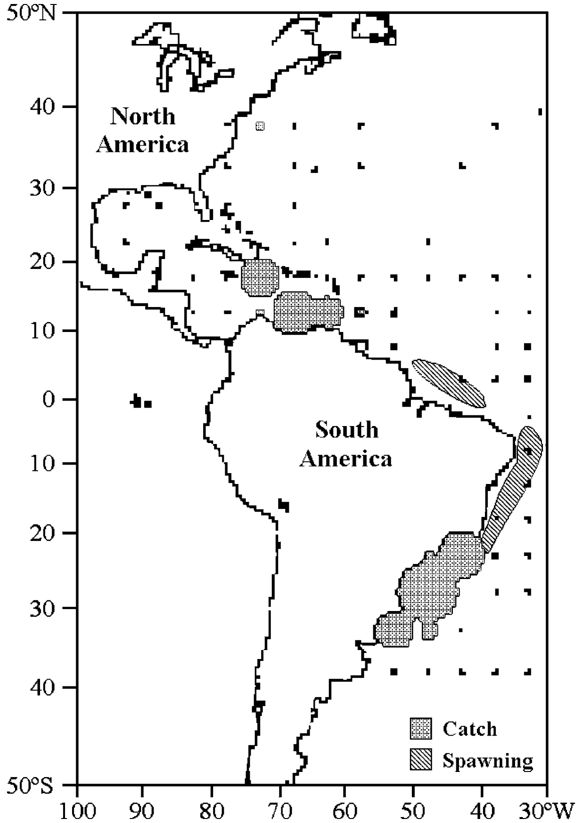

Figure 1. Skipjack tuna catches and spawning area in the West Atlantic. Spots indicate small catches. This schematic map was constructed by overlaying the spawning map of Matsuura (1986) on the map of catches available from ICCAT (2003).

size. Measuring 110 skipjack ensured that the s.e. was minimized. We also tested whether the length frequency determined early during a landing operation differed from that obtained when offloading was almost finished. In almost half the landings analysed, there were peaks representative of different length classes, so the sampling procedure adopted wherever possible was to measure at least 110 skipjack during the first half of the offloading process and another 110 during the second half. The effective number of skipjack measured per month and year are shown in Table 1.

Length–weight relationships for the Southwest Atlantic (Andrade and Campos, 2002) were used to estimate the total weight of the fish in the samples. Ratios between the total weight caught and the weight of the samples were used to extrapolate the total monthly length frequency from length frequency samples (Sparre *et al.*, 1989).

Model building and fitting

To derive the maximum likelihood function for LFA, we used the approach of Schnute and Fournier (1980), and Fournier and Breen (1983), reviewed in MacDonald (1987). Hence, we assume the probability density function of length l for a given age class i to be normal:

$$f_i(l) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(l-\mu_i)^2}{2\sigma_i^2}\right] \tag{1}$$

where $i = 1, \ldots, A$ refers to the index for the age class with mean age a_i, μ_i the mean length predicted for age class i, and σ_i the corresponding standard deviation (s.d.). We assume that μ_i is related to a_i by a von Bertalanffy growth equation and that σ_i is proportional to the square root of a_i:

$$\sigma_i = \sigma\sqrt{a_i} \tag{2}$$

The above function is an extension of that of Fournier and Breen (1983), though where we use the more general mean age a_i, Fournier and Breen used the age-class index i.

In all, k length intervals $(l^{(j-1)}, l^{(j)}]$, indexed by j $(j = 1, \ldots, k)$, are defined. The probability that a length measurement for a fish in age class i is in length interval j is

$$F_{ji} = \int_{l^{(j-1)}}^{l^{(j)}} f_i(l)\, dl \tag{3}$$

If N is the number of fish measured and g_i the true proportion of fish in age class i, then the expected number of fish of size class j is

$$L_j = N \sum_{i=1}^{A} F_{ji} g_i \tag{4}$$

The proportion of fish in size class j is then

$$p_j = \frac{L_j}{N} = \sum_{i=1}^{A} F_{ji} g_i \tag{5}$$

Finally, if the k-dimensional vector of observed length frequency $L^{obs} = \{L_j^{obs}\}$ is assumed to follow the multinomial distribution

$$f(L^{obs}) = \binom{N}{L_1^{obs} \ldots L_k^{obs}} \prod_{j=1}^{k} p_j^{L_j^{obs}} \tag{6}$$

then the negative log-likelihood function to be minimized is

$$\ell(L^{obs} | \{\mu_i\}, \{g_i\}, \sigma) = -\sum_{j=1}^{k} L_j^{obs} \log\left(\sum_{i=1}^{A} F_{ji} g_i\right) \tag{7}$$

We used the von Bertalanffy parameters estimated by Vilela and Castello (1991) to describe the growth process. Hence, the model used to predict mean length μ_a for a given mean age a was

Table 1. Number of skipjack tuna measured per year and month.

Month	1995	1996	1997	1998	1999	2000	2001	Total
January	—	954	879	1 351	2 449	3 516	250	9 399
February	—	1 254	795	1 174	1 115	1 975	2 093	8 406
March	—	1 398	1 011	1 301	1 108	749	1 072	6 639
April	—	569	900	1 070	1 390	924	831	5 684
May	177	551	151	—	545	1 264	495	3 183
June	239	668	634	618	462	497	399	3 517
July	566	—	600	317	1 660	684	697	4 524
August	676	104	691	861	617	—	910	3 859
September	292	100	597	456	478	232	298	2 453
October	—	—	882	213	—	—	—	1 095
November	629	485	—	418	1309	130	553	3 524
December	578	86	376	175	—	—	454	1 669
Total	3 157	6 169	7 516	7 954	11 133	9 971	8 052	53 952

$$\mu_a = 87.078[1 - \exp^{-0.22(a+2.071)}] \tag{8}$$

where μ_a is in cm and a is in years.

To fit an LFA to data, A, the number of age classes, needs to be specified in advance. In order to deal with this unknown number, we fitted a set of models from $A = 1$ to a maximum as large as necessary to represent all discernible age classes, and used an index (e.g. the Akaike Information Criterion, AIC; Akaike, 1974) to choose the best fit.

For A age classes, there are $2 \times A$ parameters to be estimated. In preliminary applications of the LFA, the solutions appeared to be highly dependent upon the starting values of the parameters, because the shape of the negative log-likelihood function is not informative or has several local minima. Therefore, each optimization procedure was tried with $10 \times A$ random starting parameter sets. For instance, for $A = 3$, 30 random starting parameter sets were simulated and the optimization was applied 30 times. Seidel's updating method was used as a "tuning" mechanism to find the minimum negative log-likelihood for each set of starting parameters. For a given set of starting parameters $(\{\mu_i'\}, \{g_i'\}, \sigma)$, the optimization function was applied to estimate $\{\mu_i\}$, $\{g_i\}$, and σ. Then, the estimated $\{\mu_i\}$ replaced the initial $\{\mu_i'\}$, the optimization was applied again, and the estimated $\{g_i\}$ replaced the initial $\{g_i'\}$, and so on. The procedure was repeated until further variation in the likelihood function was negligible. Finally, the smallest negative log-likelihood obtained from all $10 \times A$ optimizations was selected.

Simulation study

To choose among models fitted for different values of A, some criterion had to be used. However, the reliability of the model-selection criterion had to be examined too, so an operating model was used to simulate data and to test the performance of different criteria. The simulation study allowed assessment of the bias and the uncertainty in the resulting parameter estimates. Our operating model was defined after fitting the proposed model to real data using the AIC as the optimization criterion, assuming that the uncertainty structure of the model was correct, and taking the estimated parameters as if they were true.

Given the operating model, length frequencies were simulated with age A ranging from 2 to 6. Each simulation was conducted for three sample sizes of fish length measurements (N): 15 000, 40 000, and 65 000. The vector $\{a_i\}$ was simulated in three ways:

(i) a random selection of ages from a continuous uniform distribution between 1 and 6 years, i.e. random distances between age classes;
(ii) a random selection of a_1 from a uniform distribution from 1 to 3 years, with $a_i = a_1 + (i-1)(0.5)$ for age groups $i = 2, \ldots, A$, a constant distance of 0.5 years between age classes;
(iii) as for (ii), but with a constant distance of 1.0 year between age classes.

These simulation arrangements emulate three biological ideas in simplified form: (i) random distances emulate the lack of a cyclic pattern in the occurrence of reproductive pulses; (ii) a constant distance of 0.5 years emulates two reproductive pulses each year; (iii) a constant distance of 1.0 year emulates one reproductive pulse per year. These three simulation arrangements correspond with ideas on reproduction published elsewhere (Cayré and Farrugio, 1986; Cayré *et al.*, 1986; Goldberg and Au, 1986; Matsuura, 1986; Pagavino, 1997).

The vector $\{g_i\}$ was built by randomly sampling A independent values $(u_1, \ldots, u_A)$ from a uniform distribution between 0 and 1, and standardizing these values to add up to 1:

$$g_i = \frac{u_i}{\sum_s u_s} \tag{9}$$

Finally, a Normal random variable with mean zero and s.d. σ_j was added to the simulated length frequency in length class j ($j = 1, \ldots, k$), emulating a sampling error. To make the shape of the length frequency more realistic, the standard deviation σ_j was defined as the square root of 10, 15, and 20% of the number of fish simulated in each length class, respectively.

Overall there are 135 different scenarios in a factorial design (5 values for A, 3 values for N, 3 mechanisms to define $\{a_i\}$, and 3 values of σ_j). For each scenario we simulated three replicates. Hence, we performed a total of 405 length frequency simulations to assess the model-selection criteria and the accuracy of the estimates (see details below).

The best fit and the reliability of estimates

The AIC (Akaike, 1974) and Deviance (McCullagh and Nelder, 1989) were used as criteria to select among models with different values of A. Choosing among any two models using the AIC is simply to select the model with the smallest AIC. In contrast, the Deviance function (DEV) applies only to nested models, by testing if the addition of more parameters significantly improves the fit (McCullagh and Nelder, 1989). Under suitable conditions, the reduction in Deviance attributable to the addition of new parameters can be accurately approximated by a χ^2 distribution. We used a 5% significance to decide if the reduction in Deviance represented a marked improvement of fit.

Four variations of the traditional AIC and DEV were evaluated:

(i) the global minimum AIC (AIC_G);
(ii) the global minimum AIC restricted to the condition that distance (in years) between estimated ages was $\geq 1/12$ years, which reflects the monthly scale of the data (AIC_GR);
(iii) the first local minimum AIC (AIC_F);
(iv) Deviance (DEV).

To compare the performances of the different model-selection criteria, three measures were defined:

(i) the error in determining the correct number of age classes;
(ii) the distance in years between the simulated age and the nearest estimated age class;
(iii) the sum of the absolute differences between the proportion of fish in each year class in the simulated and in the estimated models.

These three measures are represented by the quantities D, d, and T in Equations (10)–(12) below.

The estimated ages should be "near" the simulated ages in accurate models, especially where age classes contain many fish. Hence, let B and B′ denote the number of age classes in the operating and the estimating models, respectively. Similarly, let $\{a_1, \ldots, a_B\}$ and $\{a'_1, \ldots, a_B{}'\}$ denote the corresponding mean ages, and the vector $\{h_1, \ldots, h_B\}$ denote the proportion of fish in the age classes of the operating model. We defined

$$D = B' - B \tag{10}$$

and

$$d_i = \min_{j=1,\ldots,B'} |a_j{}' - a_i| \quad \text{for } i = 1, 2, \ldots, B. \tag{11}$$

The larger the absolute values of D, the poorer the quality of the estimation model. Acceptable models should result in estimates of a'_i close to a_i, at least for the most representative age classes. Therefore, the larger the value of h_i, the smaller should be the corresponding d_i.

In order to make comparisons between simulated and estimated proportions of fish in different age classes, the third measure had to be defined. For convenience, age classes of the simulation and estimation model were rounded to the nearest year and denoted by the term "year classes". For instance, a model with a vector $\{a_i\}$ with five age-class parameters $\{2.8, 3.5, 4.1, 4.4, 5.3\}$ would result in a vector of three year classes $\{3, 4, 5\}$. Let E and E′ denote the number of year classes in the operating and the estimating models, respectively. Similarly, $\{f_1, \ldots, f_E\}$ and $\{f'_1, \ldots, f'_{E'}\}$ denote the corresponding proportion of fish in these year classes. Then

$$T = \frac{\sum_k |f'_k - f_k|}{2} \quad \text{with } k = 1, \ldots, \max\{E, E'\} \tag{12}$$

The larger the values of T, the poorer the quality of the fit. The division by 2 simplifies the interpretation: if all fish are classified in a wrong year class, the summation in the numerator is 2, and T is 1 (i.e. an error of 100%).

One among the four model-selection criteria (AIC_G, AIC_GR, AIC_F, DEV) was chosen on the basis of the performance of the measurements. To verify which factor (e.g. noise, number of simulated age classes) affects the performance of the chosen model-selection criterion most, we undertook an analysis of variance using D and $\log(T + 0.03)$ as response variables. This transformation for T proved to be necessary to make the response variable nearly Normal.

Birthdate calculations

Once the LFA procedure and the selection criteria had been tested through simulation, the procedure was applied to available monthly length frequency data. The estimated catch-at-age parameters $\{a'_i\}$, together with the corresponding proportions $\{g_i{}'\}$, were the main results of the analysis. These estimates were used to back-calculate the estimated month of birth (the birthdate), as follows. If in a given month we estimated an age-class parameter, for instance $a'_i = 2.8$ years, this indicates a birthdate 34 months earlier for $100g'_i$% of that month's total catch. We made the simplifying assumption that all data were from instantaneous fishing in the middle of the month.

In order to determine whether there were seasonal variations in the time-series of monthly birthdate estimates, we first eliminated the effect of the variability between years by replacing monthly estimates by their normalized z-scores (Zar, 1984). A z-score of -0.5 for a given month would indicate that the estimate was 0.5 s.d. below the annual average. For each month we calculated the average z-score over all years, and also performed autocorrelation and spectral analysis of z-score time-series to evaluate whether there was an apparent cyclic pattern in the monthly birthdate estimates.

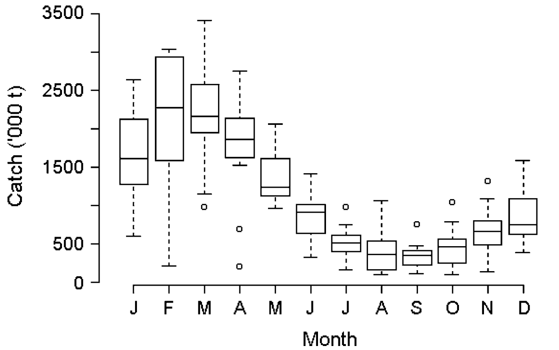

Figure 3. Monthly variation in skipjack tuna landings in Santa Catarina state pooled from 1995 to 2001 (source: Paiva, 1997; GEP/UNIVALI, 2002).

Results

Fishery and length frequency data

The Brazilian bait boat fleet of Santa Catarina state yields almost half the country's skipjack catch in the West Atlantic (Figure 2). Catches of this West Atlantic stock increased from the mid-1970s to the mid-1980s, when they peaked at 40 272 t, then decreased to oscillate around 28 000 t, apparently in dynamic stability (Figure 2). However, despite the apparent stability in the average annual catch, there has been strong seasonal variability in skipjack catches in the area of study. The main fishing season is January–March, the austral summer (Figure 3), when monthly catches are about four times larger than those made in winter (July–September).

The size frequency plots are highly variable. In most cases there were one or two modes, but there were other modes sometimes, mainly from May to September (Figure 4). Clearly too, the dynamics of the age groups sampled is high, some modes occurring in one month but missing from the next.

Criteria for model selection

When taken as a function of A (the number of age classes), model fitting with AIC and DEV worked well for most data

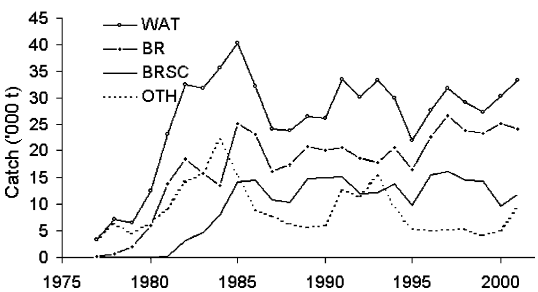

Figure 2. Historical landings of skipjack tuna in the West Atlantic: WAT, total catches of the West Atlantic stock; BR, Brazilian fleet catches; BRSC, catches of the Brazilian fleet of Santa Catarina state; OTH, catches of other fleets (source: Paiva, 1997; GEP/UNIVALI, 2002; ICCAT, 2003).

sets. In some cases, however, both criteria showed undesirable patterns (Figure 5). The results of the comparisons among all four criteria are shown in Figures 6–8. All criteria, except the AIC_G, tend to underestimate the true number of age classes (Figure 6), but calculations of median and mean D suggest that AIC_G is more biased than AIC_F and DEV (Table 2). Closer examination indicated that there were very large negative errors (i.e. underestimation) only when age classes with a very small number of fish (usually $<1\%$) showed up in the model simulation.

The minimum distance (d_i) between simulated and estimated ages and the proportion of fish at each simulated age is depicted in Figure 7. Most age classes containing a large proportion of the fish were estimated accurately using any of the four criteria. Values of d_i were large only for age classes with a small proportion of fish. Overall, about 75% of ages in the length frequencies were accurately estimated (precision of one month) employing any of the four criteria.

The proportions of fish classified in incorrect year classes were slightly different among criteria (Figure 8). In more than 45% of the fits with criteria AIC_F and DEV, no fish were classified in wrong year classes, whereas this proportion was about 40% for the other two criteria. In more than 70% of cases, the proportion of fish classified in wrong year classes was not $>15\%$ when the criteria AIC_F and DEV were considered.

In summary, the performances of the four criteria are not substantially different, but AIC_F was selected as best owing to the central tendency displayed by the measure of accuracy D (Table 2), and because of the slightly better performance for measure T.

Analyses of variance

While the factor "number of fish" in the length frequency is not significant (Table 3), the number of age classes has a negative effect on the accuracy (Figure 9b, e). Measure D is not close to 0, and measure T becomes large for length

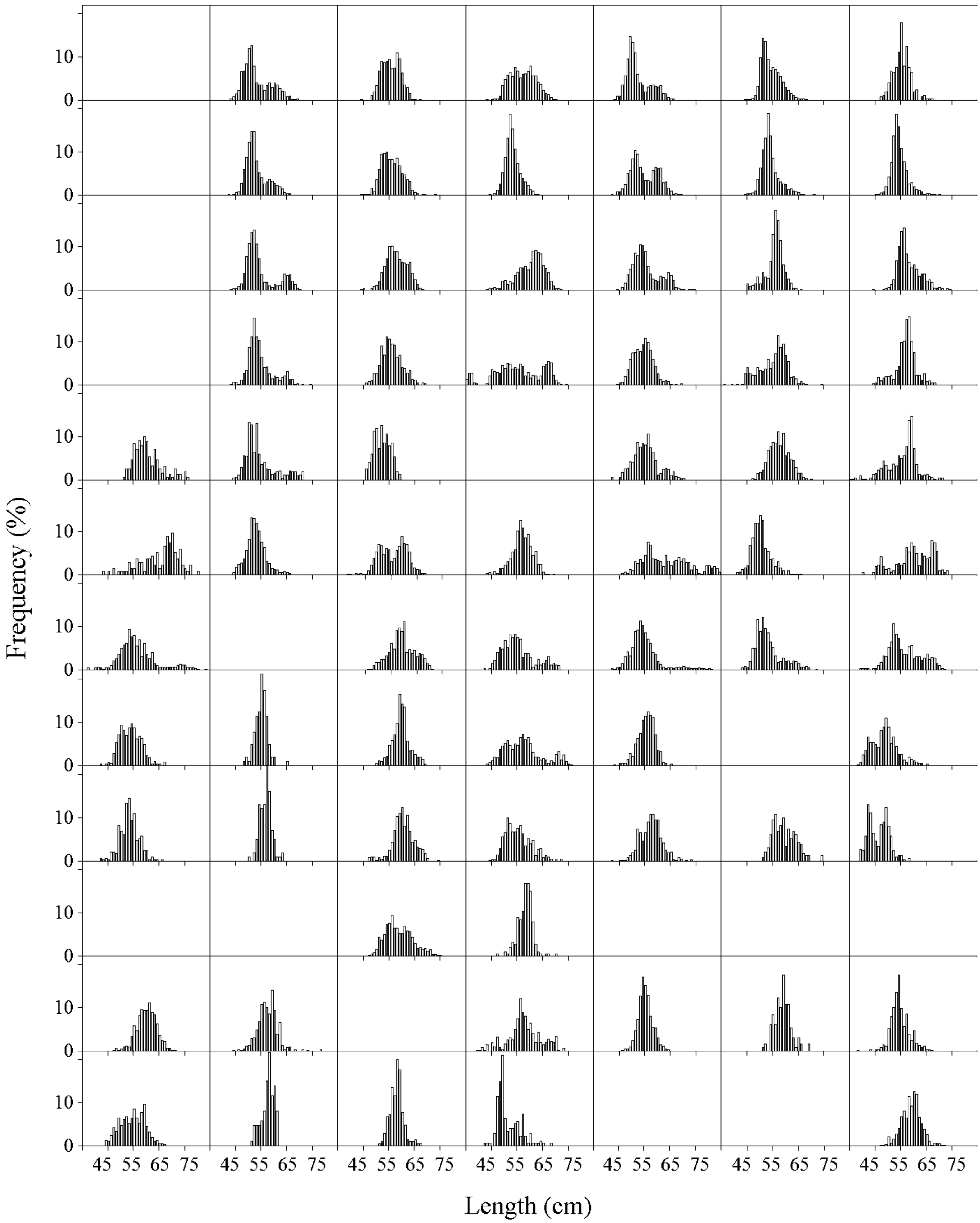

Figure 4. Length frequencies of skipjack tuna caught in the Southwest Atlantic. Months correspond to the rows, and years (from 1995 to 2001) to the columns. Empty squares depict missing length frequency samples.

frequencies in which the distance between age classes is 0.5 years (Figure 9c, f). Therefore, a small distance between age classes (in this case, 0.5 years) also has a negative effect on the accuracy.

The number of peaks in the length frequency distributions increases if noise (related to σ_j) is introduced, so the chance of detecting false age classes is greater when analysing such samples. Consequently, as noise increases, the

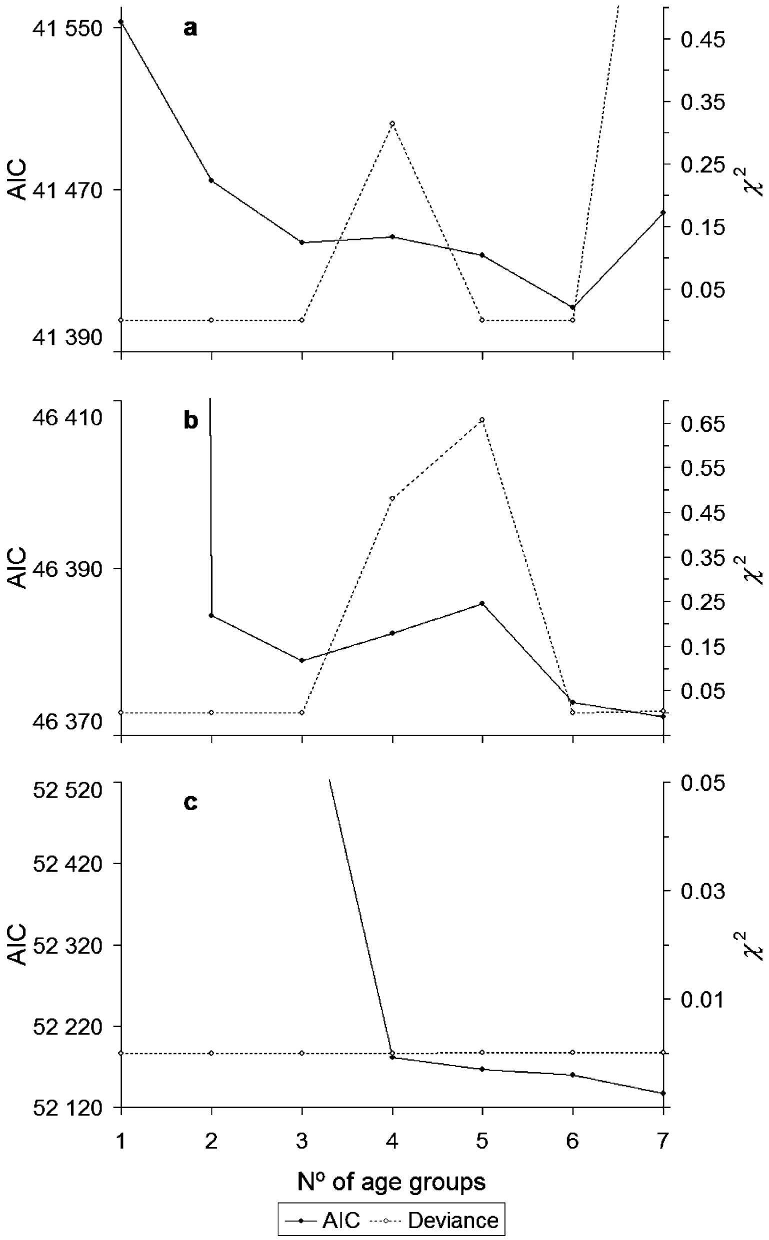

Figure 5. Three examples of atypical performances of the χ^2 p-value of AIC and Deviance, as a function of the number of age groups in the estimation model. (a) The correct number of age groups (A = 3) is associated with the first local minimum AIC; (b) the correct number of age groups (A = 3) is associated with the first local minimum AIC, but the global minimum is never reached; (c) the correct number of age groups (A = 5) is not at a minimum AIC. In all cases, the χ^2 p-value of Deviance oscillates or never reaches the significance level adopted (0.05).

proportion of fish classified in wrong year classes increases (Figure 9d). In contrast, analysis of samples with noise in the data did not result in large errors in the estimated number of age classes (Figure 9a). In that case, the tendency induced by the noise to overestimate was compensated for, because AIC_F favours underestimation of the number of age classes.

Birthdate and catch-at-age calculations

Skipjack caught from 1995 to 2001 were spawned between 1989 and 2000, most (89.5%) between 1993 and 1998 (Figure 10). Such dominance is explained by the short period during which skipjack are available to the fishery, only the 1993−1998 year classes being available in the data

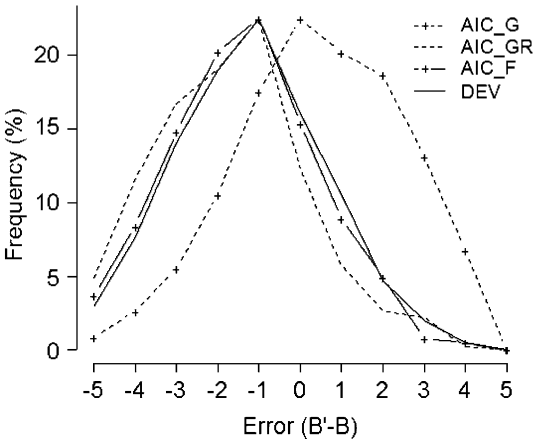

Figure 6. Distribution of the difference between the number of estimated and simulated age classes (D) according to four model-selection criteria: AIC_G, the global minimum AIC; AIC_GR, the global minimum AIC with restricted distance between estimated successive age classes of at least 1/12; AIC_F, the first local minimum AIC for increasing number of age classes; DEV, Deviance.

at their "strong" ages 2 and 3 (see below). Consequently, we considered just these six years in the birthdate analysis. Peak birthdates throughout the year were apparent in those six years (Figure 10), but spectral and autocorrelation analyses (not shown here) showed that there was no seasonal variation. This result is further supported by analysis of the z-score index. The variances of this index were high at the beginning and end of the year. Also, the average z-score is generally higher at the end of the year, although there is no statistical evidence for a strong 12-month cycle (Figure 11).

Skipjack are available to the fishery mainly from ages 1 to 4, though most catches (about 85%) are of ages 2 and 3. During the austral summer, much of the total catch is of young fish ($\approx$2 years old; Figure 12), but in autumn and spring, slightly older fish ($\approx$3 years old) dominate the catches.

The 1994 and 1997 year classes were seemingly relatively strong, as evidenced by the catches of 2-year-olds in 1996 and 1999 (Figure 12b). In particular, the strong recruitment of skipjack spawned in 1994 can be tracked

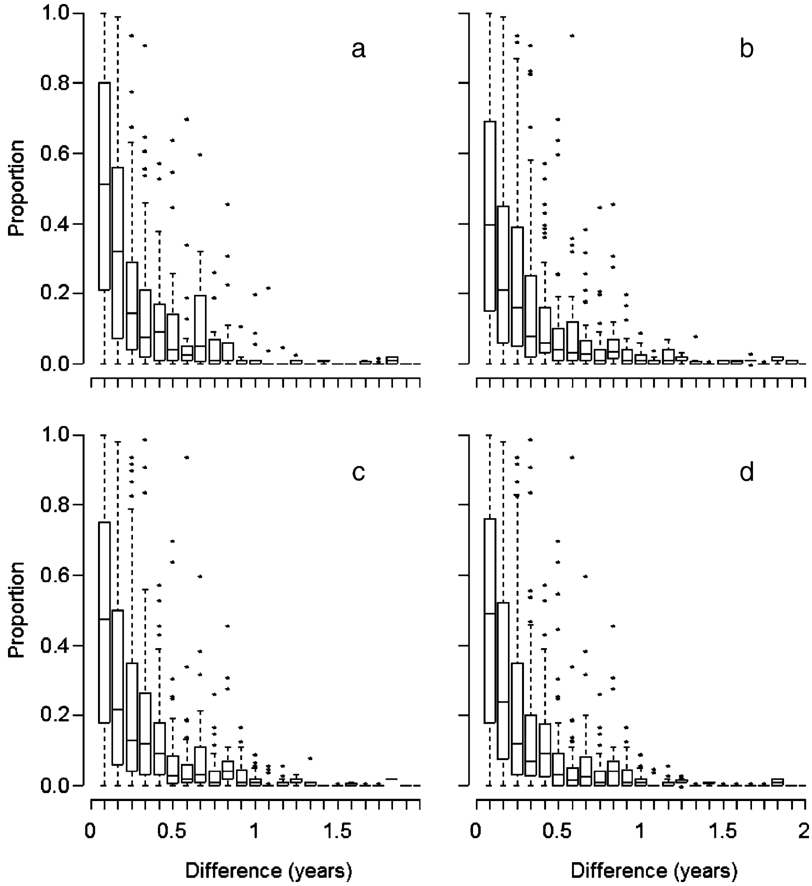

Figure 7. Comparison of the difference between simulated and the nearest estimated ages with the proportion of fish at the simulated age, according to four model-selection criteria: (a) the global minimum AIC; (b) the global minimum AIC with restricted distance between estimated successive age classes of at least 1/12; (c) the first local minimum AIC for increasing number of age classes; and (d) Deviance.

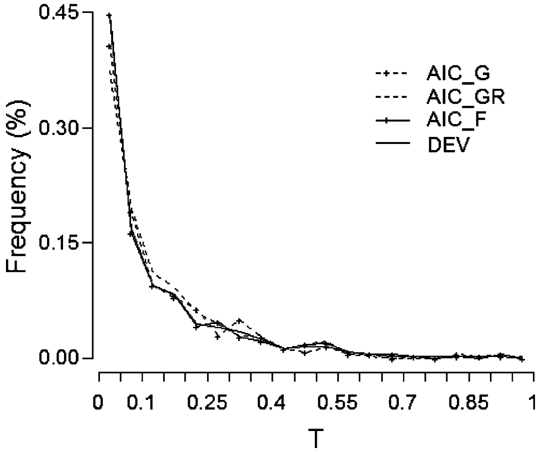

Figure 8. Frequency distribution of the proportion of fish classified in wrong year classes (T) for four model-selection criteria: AIC_G, the global minimum AIC; AIC_GR, the global minimum AIC with restricted distance between estimated successive age classes of at least 1/12; AIC_F, the first local minimum AIC for increasing number of age classes; DEV, Deviance.

through ages 2 to 4 from 1996 to 1998. The recruitment of fish spawned in 1997 was not so easily tracked, though its strength was still not negligible.

Discussion

Operating models are usually necessary to test the estimation procedures. In the example of skipjack, operating models proved useful for evaluating the different model-selection criteria and the accuracy of the estimates. Although AIC is a popular model-selection criterion, it can be biased for multinomial likelihoods (Hilborn and Mangel, 1997), and because our intention was to use a multinomial density distribution, Deviance was expected to perform better than traditional AIC (McCullagh and Nelder, 1989). The results of the analysis confirmed this expectation. Nevertheless, the performance of two modified AIC (i.e. AIC_GR and AIC_F) was similar to that of Deviance, and there is likely merit in the performance of these and other model-selection criteria (e.g. Bayesian Information Criterion – Schwarz, 1978) being evaluated in order to improve the LFA approach.

Table 2. Basic statistics for the difference (D) between the number of estimated and simulated age classes for different model-selection criteria: AIC_G, the global minimum AIC; AIC_GR, the global minimum AIC with restricted distance between estimated successive age classes of at least 1/12; AIC_F, the first local minimum AIC for increasing number of age classes; and DEV, Deviance.

Model-selection criterion	Median	Mean	Variance
AIC_G	1	1.452	4.011
AIC_GR	−1	−0.664	3.347
AIC_F	0	−0.370	3.254
DEV	0	−0.207	3.373

Table 3. Analysis of variance to examine the effects of different factors used to construct the operating model on the difference between the number of estimated and simulated age classes (D), and on a transformation of the proportion of fish classified in wrong year classes (T). The model-selection criterion was the first local minimum AIC (AIC_F).

Parameter	d.f.	Sum of squares	Mean square	F	p
D					
Sampling error	2	11.20	5.60	2.7632	0.06431
Number of fish	2	5.73	2.87	1.4145	0.24428
Number of age classes	4	440.22	110.06	54.3049	<2.2e−16
Distance between age classes	2	58.80	29.40	14.5069	8.3e−07
Residuals	394	798.49	2.03		
Log(T+0.03)					
Sampling error	2	5.658	2.829	3.8922	0.02119
Number of fish	2	1.016	0.508	0.6986	0.49788
Number of age classes	4	6.602	1.650	2.2707	0.06104
Distance between age classes	2	7.777	3.889	5.3499	0.00510
Residuals	394	286.388	0.727		

Selection of a representative growth model and its parameters is also important for LFA. As there is only one source of growth parameter estimates for the Southwest Atlantic (i.e. Vilela and Castello, 1991), this is the model used in this application to skipjack. A key point for the accuracy of the estimates obtained with LFA is the curvature of the growth model. If the curve is relatively flat, conversion of length to age is imprecise near the asymptotic length (Rosenberg and Pope, 1987). Although the model available in Vilela and Castello (1991) is somewhat flat, the accuracy of the estimates used here was still acceptable. In the few cases for which the estimation models were extremely inaccurate (T > 40; Figure 8), the poor performance was caused by, first, age classes simulated with small proportions (e.g. <1%) being embedded in age classes with larger proportions, and second, rounding errors attributable to the collapsing of age groups into year classes (the so-called "rounding effect"). The first cause introduced small errors because such undetectable embedded age classes had little weight in the analysis. However, the "rounding effect" introduced large errors. For example, if an estimated mean age of 2.45 years is compared with a simulated mean age of 2.52 years, they appear as highly inaccurate in Figure 8, because after rounding, they fall into year classes 2 and 3, respectively. Nevertheless, such inaccuracies affect catch-at-age (in year

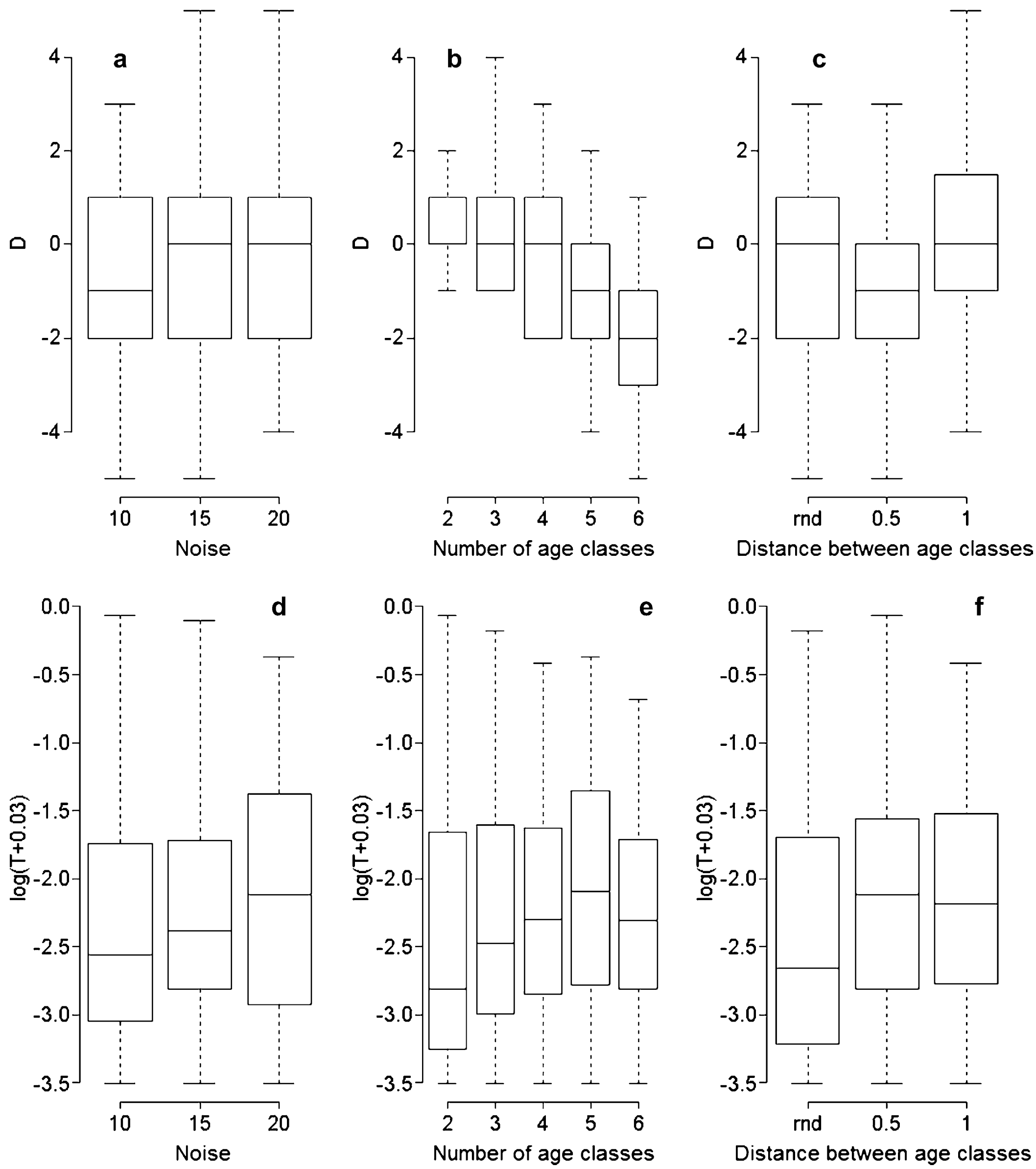

Figure 9. Box and whisker plot for the effects of the factors sampling error (noise), number of age classes, and distance between age classes in the operating model on the accuracy of estimates obtained using the first minimum AIC (AIC_F). D is the difference in estimated and simulated number of age classes, and T is the proportion of fish classified in wrong year classes.

classes) but not the estimate of birthdates, the main quantity of interest.

One reasonable assumption is that birthdates can be used to infer spawning dates, because times between the eggs floating free in the water and hatching are probably short. If the time lags are large, but constant throughout the year, the oscillations of birthdates would mirror the variability of spawning dates (with a fixed time lag). No matter which assumption holds, there is no evidence in the current analysis to support the hypothesis of a single main spawning season providing the fish entering the fishery in the Southwest Atlantic. It is likely that, given their highly migratory nature, skipjack derived from seasonal spawning in subtropical waters (Matsuura, 1982, 1986; Goldberg and Au, 1986) quickly mix with others spawned at any time and elsewhere in equatorial waters. Therefore, the seasonal subtropical spawning cycle does not translate into recruitment variability.

Continuous recruitment of juveniles to the adult stock does not necessarily translate into continuous recruitment to

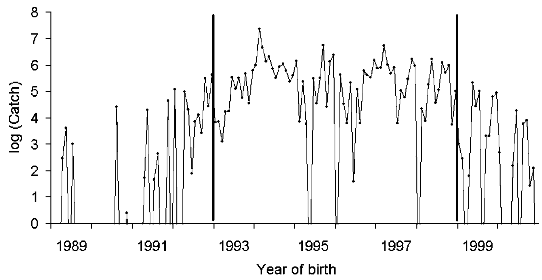

Figure 10. Birthdate and the number of skipjack tuna caught in the Southwest Atlantic from 1995 to 2001. Most catches (almost 90%) were of fish spawned between 1993 and 1998.

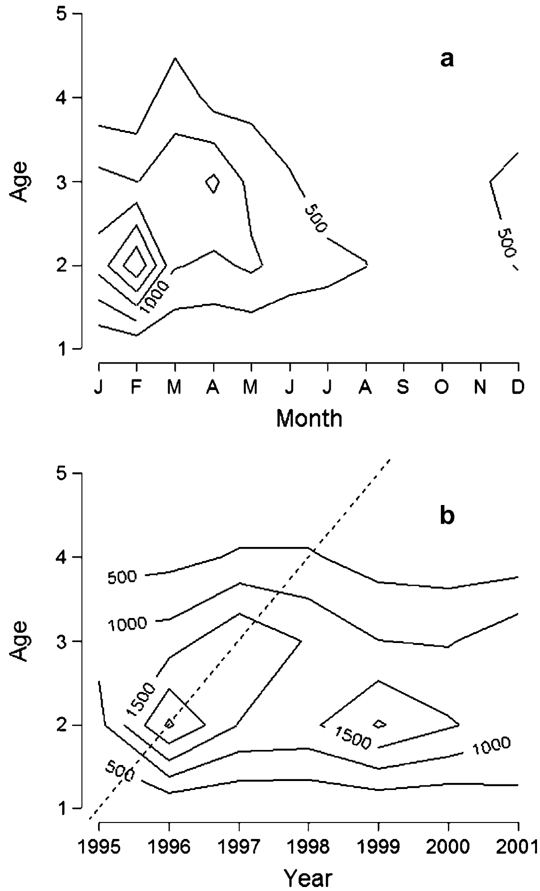

Figure 12. Catch-at-age (thousands) of skipjack compared with (a) birthdate month and (b) year the fish were caught in the Southwest Atlantic.

the fishery. Apparently, young fish join the adult population throughout the year, but become highly vulnerable only during the first summer after they reach 1.5 years of age. This increase in catchability during summer has been related to the oceanography (e.g. thermocline depth; Andrade, 2003).

To validate the results of recruitment strength obtained with LFA, the use of other methods would be desirable. Unfortunately, however, at present there is no other source of information on recruitment of skipjack in the Southwest Atlantic. Notwithstanding, although further comparisons are not possible, the results indicating a strong year class in 1994 were consistent throughout the analysis. The dominant age class in the catch clearly shifts from 2 to 3 and then to 4 years old after 1996 (Figure 12), so the increasing catches of the Brazilian fleet after 1995 (Figure 2) are clearly based on the successful recruitment of the 1994 year class.

The definition of just one West Atlantic stock is based on the conjecture that the fisheries for skipjack in the Caribbean Sea and Southwest Atlantic (south of 20°S) share the same spawning ground northeast of South America and the recruits generated by that spawning, despite the large geographical gap between the two main fishing grounds. If this is true, spawning events leading to strong year classes should affect catches equally, and fishery yields on both fishing grounds should be more in synchrony. However, the current results fail to demonstrate such a link between skipjack fished in the two areas: the strong year class of 1994 detected in the Southwest Atlantic seems not to have affected catches in the Caribbean Sea. This suggests that the unit West Atlantic stock assumption may be open to question. Analysis of data from the Caribbean Sea, perhaps using a procedure similar to that described here, could help to clarify how closely related the recruitment events and abundance oscillations of skipjack are in both fishing grounds of the West Atlantic.

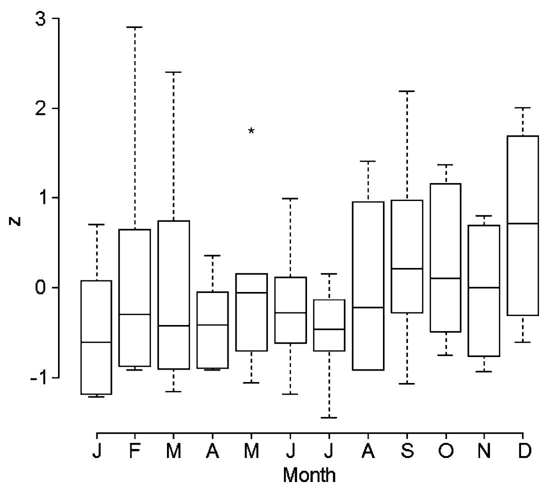

Figure 11. Box and whisker plot for the standardized z-score calculated for the number of fish born by month from 1993 to 1998.

Acknowledgements

We thank Pilar Palarés and two anonymous reviewers for their valued comments and suggestions that helped us focus the manuscript better. The first author is also grateful for the support he received from PICDT/CAPES.

References

Akaike, H. 1974. A new look at the statistical identification model. IEEE Transactions on Automatic Control, 19: 716–723.

Andrade, H. A. 2003. The relationship between the skipjack tuna (*Katsuwonus pelamis*) fishery and seasonal temperature variability in the southwestern Atlantic. Fisheries Oceanography, 12: 10–18.

Andrade, H. A., and Campos, R. O. 2002. Allometry coefficient variations of the length–weight relationship of skipjack tuna (*Katsuwonus pelamis*) caught in the southwest South Atlantic. Fisheries Research, 55: 307–312.

Cayré, P. M., and Farrugio, H. 1986. Biologie de la reproduction du listao (*Katsuwonus pelamis*) de l'Ocean Atlantique. *In* Proceedings of the ICCAT Conference on the International Skipjack Year Program, pp. 252–271. Ed. by P. E. K. Symons, P. M. Miyake, and G. T. Sakagawa. ICCAT, Madrid.

Cayré, P. M., Fonteneau, A., and Diouf, T. 1986. Eléments de biologie affectant la composition en taille des listaos (*Katsuwonus pelamis*) exploités dans l'Atlantique tropical oriental et leur effet sur l'analyse de la croissance de l'espèce par la méthode de Petersen. *In* Proceedings of the ICCAT Conference on the International Skipjack Year Program, pp. 326–334. Ed. by P. E. K. Symons, P. M. Miyake, and G. T. Sakagawa. ICCAT, Madrid.

Fournier, D. A., and Breen, P. A. 1983. Estimation of abalone mortality rates with growth analysis. Transactions of the American Fisheries Society, 112: 403–411.

Fournier, D. A., Hampton, J., and Sibert, J. R. 1998. MULTIFAN–CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. Canadian Journal of Fisheries and Aquatic Sciences, 55: 2105–2116.

Fournier, D. A., Sibert, J. R., Majkowski, J., and Hampton, J. 1990. MULTIFAN: a likelihood-based method for estimating growth parameters and age composition from multiple length frequency data sets illustrated using data for southern bluefin tuna (*Thunnus maccoyi*). Canadian Journal of Fisheries and Aquatic Sciences, 47: 301–317.

Fromentin, J. M., and Restrepo, V. 2001. Recruitment variability and environment: issues related to stock assessment of Atlantic tunas. ICCAT Collective Volume of Scientific Papers, 52: 1780–1797.

GEP/UNIVALI. 2002. Boletim estatístico da pesca industrial de Santa Catarina – Ano 2001. Itajaí, SC, UNIVALI/MAPA. 89 pp.

Goldberg, S. R., and Au, D. W. K. 1986. The spawning of skipjack tuna from southeastern Brazil as determined from histological examination of ovaries. *In* Proceedings of the ICCAT Conference on the International Skipjack Year Program, pp. 277–284. Ed. by P. E. K. Symons, P. M. Miyake, and G. T. Sakagawa. ICCAT, Madrid.

Hasselblad, V. 1966. Estimation of parameters for a mixture of normal distributions. Technometrics, 8: 431–444.

Hilborn, R., and Mangel, M. 1997. The Ecological Detective – Confronting Models with Data. Princeton University Press, New Jersey. 315 pp.

Ianelli, J. N. 1993. Studies on the population structure of skipjack tuna, *Katsuwonus pelamis*, in the central and eastern Pacific Ocean. PhD thesis, University of Washington.

ICCAT, 2000. Detailed report on skipjack tuna. Report of the ICCAT skipjack tuna stock assessment session. ICCAT Collective Volume of Scientific Papers, 51: 132–220.

ICCAT. 2003. Report of the Standing Committee on Research and Statistics of ICCAT. Executive Summaries of Species Status Reports, Madrid. 216 pp.

Johnson, S. J., and Quinn, T. J. 1987. Length frequency analysis of sablefish in the Gulf of Alaska. Juneau, Alaska Technical Report UAJ–SFS–8714.

Kikawa, S., and Nishikawa, Y. 1980. Distribution of larvae of yellowfin tuna and skipjack in the Atlantic Ocean (preliminary). ICCAT Collective Volume of Scientific Papers, 9: 138–142.

MacDonald, P. D. M. 1987. Analysis of length frequency distributions. *In* Age and Growth of Fish, pp. 265–346. Ed. by R. C. Summerfelt, and G. E. Hall. Iowa State University Press, Ames, Iowa. 544 pp.

MacDonald, P. D. M., and Pitcher, T. J. 1979. Age groups from size frequency data: a versatile and efficient method. Journal of the Fisheries Research Board of Canada, 36: 987–1001.

Matsumoto, W. M., Skillman, R. A., and Dizon, A. E. 1984. Synopsis of biological data on skipjack tuna, *Katsuwonus pelamis*. NOAA Technical Report, NMFS Circular, 451. 92 pp.

Matsuura, Y. 1982. Distribution and abundance of skipjack tuna (*Katsuwonus pelamis*) larvae in eastern Brazilian waters. Boletim do Instituto Oceanográfico, 31(2): 5–7.

Matsuura, Y. 1986. Distribution and abundance of skipjack larvae off the coasts of Brazil. *In* Proceedings of the ICCAT Conference on the International Skipjack Year Program, pp. 285–289. Ed. by P. E. K. Symons, P. M. Miyake, and G. T. Sakagawa. ICCAT, Madrid.

McCullagh, P., and Nelder, J. A. 1989. Generalized Linear Models. Chapman & Hall, New York. 513 pp.

McNew, R. W., and Summerfelt, R. C. 1978. Evaluation of a maximum likelihood estimator for analysis of length frequency distributions. Transactions of the American Fisheries Society, 107: 730–736.

Pagavino, M. 1997. Indice gonadal y crecimiento del atún Listado (*Katsuwonus pelamis*) del Mar Caribe. ICCAT Collective Volume of Scientific Papers, 46(4): 268–276.

Paiva, M. P. 1997. Recursos Pesqueiros Estuarinos e Marinhos do Brasil. Edições UFC, Fortaleza. 286 pp.

Pauly, D., and David, N. 1981. ELEFAN I, a BASIC program for the objective extraction of growth parameters from length-frequency data. Meeresforschung, 28: 205–211.

Pauly, D., and Morgan, G. R. 1987. Length-based methods in fisheries research. ICLARM Conference Proceedings, 13. 468 pp.

Quinn, T. J., and Deriso, R. B. 1999. Quantitative Fish Dynamics. Oxford University Press, New York. 542 pp.

Ricker, W. E. 1975. Computation and interpretation of biological statistics of fish populations. Bulletin of the Fisheries Research Board of Canada, 191. 382 pp.

Rosenberg, A. A., and Beddington, J. R. 1988. Length-based methods of fish stock assessment. *In* Fish Population Dynamics, pp. 83–103. Ed. by J. A. Gulland. John Wiley, Chichester, UK. 422 pp.

Rosenberg, A. A., and Pope, J. G. 1987. Comments on age-length vs. length-age relationships. *In* Length-Based Methods in Fisheries Research, pp. 305–309. Ed. by D. Pauly and G. R. Morgan. ICLARM Conference Proceedings, 13. 468 pp.

Schaefer, K. M. 2001. Assessment of skipjack tuna (*Katsuwounus pelamis*) spawning activity in the eastern Pacific Ocean. Fishery Bulletin US, 99: 343–350.

Schnute, J., and Fournier, D. A. 1980. A new approach to length frequency analysis: growth structure. Canadian Journal of Fisheries and Aquatic Sciences, 37: 1337–1351.

Schwarz, G. 1978. Estimating the dimension of a model. The Annals of Statistics, 6: 461–464.

Sparre, P., Ursin, E., and Venema, S. C. 1989. Introduction to tropical fish stock assessment. FAO Fisheries Technical Paper, 306/1. 429 pp.

Stéquert, B., Rodriguez, J. N., Cuissetc, B., and Le Mennc, F. 2001. Gonadosomatic index and seasonal variations of plasma sex steroids in skipjack tuna (*Katsuwonus pelamis*) and yellowfin tuna (*Thunnus albacares*) from the western Indian Ocean. Aquatic Living Resources, 14: 313–318.

Uchiyama, J. H., and Struhsaker, P. 1981. Age and growth of skipjack tuna, *Katsuwonus pelamis*, and yellowfin tuna, *Thunnus albacares*, as indicated by daily growth increments of sagittae. Fishery Bulletin US, 79: 151–162.

Vilela, M. J. A., and Castello, J. P. 1991. Estudio de la edad y del crecimiento del barrilete (*Katsuwonus pelamis*) en la region sur y sudeste de Brasil. Frente Maritmo, 9: 29–35.

Zar, J. H. 1984. Biostatistical Analysis. Prentice-Hall, New Jersey. 718 pp.