Northwest Atlantic Fisheries Organization

Serial No. N5906 **NAFO SCR Doc. 11/023**

SCIENTIFIC COUNCIL MEETING – JUNE 2011

On otoliths sampling

by

Mónica Mandado
Antonio Vázquez

Instituto de Investigaciones Marinas
mandado@iim.csic.es
avazquez@iim.csic.es

Summary

The best strategy for otoliths sampling and its optimum sample size were tested by Monte Carlo simulation. Every option was judged by its effects on VPA results. It was concluded that a stratified sampling is preferable to a random one. A stratified sampling of 20 otoliths by length class is the optimum for a species with 30-40 length classes. The effect of random mistakes in age determination was also analysed.

Introduction

NAFO (1980, 1999) defines three possible otoliths sampling strategies:

- Random sampling for age means that the sample is a random subsample of the length composition or it may be a separate small random sample of the catch taken specifically for ageing, with no attempt made to select fish by length groups.
- Supplemented random sampling for age implies that the basic age sample was taken at random and some effort made to supplement the basic sample with fish in the upper and lower parts of the length frequency distribution in order to broaden the length spectrum of the age-length key.
- Stratified sampling for age implies that a certain number of fish are selected from each length group represented in the catch length composition, and that the fish are selected at random within each length group.

It concluded that random age sampling is the least effective of the three types, and that stratified age sampling is the most effective and the most efficient. The same terminology will be used in this paper, but only random sampling and stratified sampling were tested.

Ketchen (1949) stated that a stratified otoliths sample for age determination, where the number of sampled fish in each length class is limited to 10 or 15, being less time consuming, is efficient to calculate numbers at age. The criterion he used was a χ^2 test on the agreement between observed and calculated numbers at age. Doubleday (1981) indicates that length stratification is preferred and the number required for each length interval can be calculated.

Three issues were considered in this paper: a) what sampling strategy is most efficient, b) how many otoliths must be taken to produce an adequate age-length key (ALK), and c) what is the effect of random mistakes in otoliths ageing.

In general, the criterion to establish a minimum sample size is linked to the precision required in the final results. Taking into account that otoliths are aged just to allow analytical analysis of the stock abundance, our approach is to use precision of VPA results as the only goodness criterion.

Methods

The Monte Carlo simulation as described by Vázquez and Mandado (2010) was used. The method allows to check the behaviour of any option when using analytical models: ADAPT (Gavaris 1988) or XSA (Darby and Flatman 1994). In each iteration the procedure defines a simulated "exact" case where abundance, fishing mortality, catch at age and survey indices at age satisfy the population dynamics equations. From these figures one input set of data is derived by introducing random variability on catch at age data and survey abundance at age indices. Every input data equals its simulated "exact" figure modified by a random factor with log-normal distribution:

$C_{a,y}$ – reported catch in number at age *a* and year *y*

$$C_{a,y} = spC_{a,y} \times \varepsilon \quad [1]$$

being:

$spC_{a,y}$ – simulated population's Catch in number at age *a* and year *y* (exact figure)
ε = logN (μ = 1, cvCA)
cvCA = parameter (coefficient of variation of commercial Catch at Age)

$I_{a,y}$ – observed survey index for age *a* and year *y*

$$I_{a,y} = spI_{a,y} \times \varepsilon \quad [2]$$

being:

$spI_{a,y}$ – simulated population's survey Index for age *a* and year *y* (exact figure)
ε = logN (μ = 1, cvSI)
cvSI = parameter (coefficient of variation of Survey Indices)

Partial recruitment to the simulated population also exhibits annual variability.

PR_a – Partial Recruitment at age *a*

$$PR_a = N\ (\mu = spPR_a,\ s.d. = sdPR\ parameter) \quad [3]$$

being:

$spPR_a$ – simulated population's Partial Recruitment at age *a* (exact figure)

Equations [1] and [2] were modified to make otoliths sampling an intermediate step in both catch at age data and survey indices. In the case of commercial catches additional intermediate steps include the following variables:

$FrecC_{l,a,y}$ – length frequency of catch at age, for length *l*, age *a*, and year *y*.

$$FrecC_{l,a,y} = C_{a,y} \times Frec_{l,a} \quad [4]$$

being:

$Frec_{l,a}$ = predefined length distribution for each age (Figure 1). $\sum_l Frec_{l,a}=1$, for each age.

$LFC_{l,y}$ – Length distribution of the catch, for length *l* and year *y*.

$$LFC_{l,y} = \sum_a FrecC_{l,a,y} \quad [5]$$

$ALK_{l,a,y}$ – Age-length key, built by random selection of length-age pairs with the probabilities set by $FrecC_{l,a,y}$. Two strategies for producing ALKs were considered, those denoted as "random" and "stratified", which difference how every age-length pair is selected from the "exact" figure (simulated otoliths). The "exact" figure was prepared as a set of 10000 age-length pairs where each pair frequency was the same as in the simulated catch at age ($FrecC_{l,a,y}$).

Every ALK was done by selecting so many age-length pairs from that set as required according to the following criteria[1].

- **Random selection**: age-length pairs were selected at random. The process only finish when a determined number of pairs has been produced. Total number was set at: 100, 200, 500, 1000 and 3000.
- **Stratified selection**: age-length pairs were selected at random but only recorded if the number of previous pairs for that length class were less than a predefined limit. The process finish when the number of revised possible pairs is ten times the attainable maximum (10 × number of length groups × limit). The limit for was set at: 5, 10, 20, 50 and 100 otoliths at each length class.

An additional procedure was introduced to simulate mistakes in age determination. Once an age-length pair was selected, a random process was followed to maintain that age or to change it to age+1 or age–1 (both cases being equally probable), with total mistakes reaching predefined levels between 0 and 50%. This procedure tries to mimic the ageing random mistakes, but not the systematic ones.

Finally, equation [1] for catch at age was replaced with:

$$C_{a,y} = \sum_l (LFC_{l,y} \times ALK_{l,a,y}) \qquad [6]$$

The ageing process through otoliths introduces a new component of variability in catch at age input data, which had originally the same value to every age-year combination (Equation [1]). Length sampling would also be an additional source of variability, but such variability was not considered here: equations [4] and [5] do not include any stochastic term, so length distribution of the catch ($LFC_{l,y}$) comes from the reported catch ($C_{a,y}$) directly.

A similar procedure was applied to survey indices, so for each survey:

$FrecS_{l,a,y}$ – length frequency of one survey for length *l*, age *a*, and year *y*.

$$FrecS_{l,a,y} = I_{a,y} \times Frec_{l,a}$$

being:

$Frec_{l,a}$ = the same predefined length distribution at age as used for the catch

$LFS_{l,y}$ – Length distribution in the survey, for length *l* and year *y*.

$$LFS_{l,y} = \sum_a FrecC_{l,a,y}$$

$ALK_{l,a,y}$ – Age-length key, built by random selection of length-age pairs with the probabilities set by $FrecS_{l,a,y}$.

Equation [2] for survey abundance indices at age was replaced with:

$$I_{a,y} = \sum_l (LFS_{l,y} \times ALK_{l,a,y}) \qquad [7]$$

Two independent surveys were simulated, as in Vázquez and Mandado (2010).

Each set of catch at age and survey indices at age was analysed with two different ADAPT formulations and one XSA. Vázquez and Mandado (2010) tested three different ADAPT formulations, and concluded that the one with greater number of parameters "produces solutions that are very far away of the simulated population"; that formulation also produces the highest inaccuracy indices in our analysis, so it was excluded of any further

[1] It was questioned if each of the 10000 cases were equally probable, which is the same as questioning the uniform distribution of the FORTRAN rand() function at a very fine scale. A chi-square (χ^2) test was done after distributing 1000 times 10000 values on the 10000 possible cases. The χ^2 distribution with so many degrees of freedom (df) satisfy:

$$\{x\} = \chi^2() \quad => \quad \{\sqrt{(2 \cdot x)}\} = N(\sqrt{(2 \cdot df - 1)}, \sigma^2 = 1) \qquad \text{(Abramowitz and Stegun 1972)}$$

The test confirms the goodness of the FORTRAN rand() function to produce 10000 equally probable cases.

consideration; the remaining two formulations provide quite similar results: ADAPT-10, with survivors of all year classes in the last year being parameters, and ADAPT-9, equal the former one but the oldest age being excluded as parameter.

Consistency of the retrospective was measured with the σ1 index (Vázquez and Mandado 2010).

σ1 – Compares survivors' abundance at age in retrospective peel_p with the same figure in peel_p-1, and calculate the squared root of mean of squared relative differences (peel_p is a retrospective case where the last p years were eliminated)

$$\sigma 1 = \sum_{p=1,5} \text{sqrt}\ [\sum_{a=1,na} (N^{p}_{a,yp} / N^{p-1}_{a,yp} - 1)^2 / na] / 5$$

being: a – ages from 1 to na
yp – survivors' year in peel_p

The agreement between the ADAPT or XSA results and the simulated population was measured by inaccuracy indices (they are named *bias indices* in the original paper). Among them, π3* and σSSB3* are good indicators of overall inaccuracy (Vázquez and Mandado 2010). They compare results on abundance or SSB in every year ($N_{a,y}$ or SSB_y) with the same figures in the simulated population ($spN_{a,y}$ or $spSSB_y$):

π3* – Mean squared root of squared logarithmic relative abundances for the last 15 years

$$\pi 3^* = \sum_{y=ny-14,ny} \text{sqrt}\ [\sum_{a=1,na} (\ln (N_{a,y} / spN_{a,y}))^2 / na] / 15)$$

being: y – year from ny–14 to ny
ny – survivors' year

σSSB3* – Squared root of mean squared relative differences in SSB for the last 15 years

$$\sigma SSB3^* = \text{sqrt}\ [\sum_{y= ny-14,ny} (SSB_y / spSSB_y - 1)^2 / 15]$$

Every Monte Carlo simulation comprises 1000 cases, each of them based on a new simulated population which was analysed with two ADAPTs and one XSA, including retrospective and inaccuracy indices for each of them. In doing so there are many cases where convergence is not achieved in either ADAPT or XSA. Non-convergence was particularly numerous with very low number of otoliths, or when the mistakes in age determination were too frequent; in both cases ALKs became too poor for consistent catch at age or survey indices input data. In order to allow the analysis when non-convergence was initially detected, the calculus routines were modified as follow:

ADAPT – The analysis starts with only those parameters (survivors) their calculation was feasible. However, an error occurs when some parameter approaches to cero along the Marquardt iteration. In those cases, the involved parameter is taken out of the analysis and its value is set to cero.

XSA – No F-shrinkage to the final year was initially assumed. When non-convergence occurs, F-shrinkage was set to two years. As in the ADAPT case, this problems mostly arrives from very low catch at some age in the last year.

In order to avoid cases were the analysis were not further completed due to lack of convergence in both ADAPT and XSA, indices with a deviation greater than 5 s.d. were excluded to calculate the final mean. Furthermore, even inaccuracy indices are independent of a retrospective analysis, indices were only considered if the retrospective output was correctly completed. It allows considering further relationship among indices.

The method was checked to verify that inaccuracy indices indicate complete coincidence between VPA results and simulated data when sources of variability in equations [1], [2] and [3] were cero, and when the ALK intermediate step was suppressed.

Results

Three issues were under consideration: otoliths' sampling strategy, sample size, and the percentage of mistakes in age determination. In order to make the analysis as simple as possible, other sources of variability were excluded, that is, cvCA in [1], cvSI in [2] and sdPR in [3] were all equal cero, so only the variability introduced by otoliths ageing was considered.

Figure 2 compares both sampling strategies with different sample sizes. It must be noted that abscissas scales are different for both sampling strategies. These graphics illustrates main features of both strategies as well as some characteristics of the method:

- On retrospective analysis – The $\sigma 1$ retrospective index decreases with increasing sample size.
- On inaccuracy indices – The $\pi 3^*$ and $\sigma SSB3^*$ inaccuracy indices decrease with increasing sample size (a lower value of inaccuracy index means higher accuracy).
- On precision – Stratified sampling produces lower retrospective and inaccuracy indices that random sampling, either with or without ageing mistakes, for equivalent overall sample size.
- On the required sample size - Stratified sampling reaches its almost asymptotic value at lower sample size than random sampling: roughly 660 (20 per length class) vs 1000 otoliths. In our case study, with roughly 40 length classes, 10 otoliths by length group in stratified sampling requires around 330 otoliths in total (not all length classes had frequency or completed the predefined number of otoliths).
- On ageing mistakes' effect – The $\pi 3^*$ inaccuracy index is the most sensible to ageing mistakes, but $\sigma 1$ retrospective and $\sigma SSB3^*$ inaccuracy indices are quite insensible.

Figures 3 and 4 illustrate the effect of ageing mistakes on both sampling strategies at different sample sizes. Main conclusions are:

- The $\sigma 1$ retrospective index is almost insensible to ageing mistakes.
- The $\pi 3^*$ inaccuracy index is more sensible than $\sigma SSB3^*$ inaccuracy index to ageing mistakes because the first one is based on numbers at age while SSB is the joint of many year-classes. When a mistake occurs at some age the decrease in abundance produced in that age is compensated by same increase in a close age, and the effect on SSB is partially minimized.
- Inaccuracy indices might indicate that disagreement between VPA results and the simulated population increases with ageing mistakes, and it occurs with the $\pi 3^*$ inaccuracy index, but it is not so clear for $\sigma SSB3^*$, particularly in the random strategy. The anomalous behaviour of the $\sigma SSB3^*$ inaccuracy index in the analysis of the random strategy is due, we think, to the high number cases (up to 10%) where the VPA was not accepted in the established conditions (which includes converged VPA and retrospective analysis completed), which resulted in skew results.
- Both figures remark the very obvious fact that disagreement between VPA results and the simulated population increases with the level of ageing mistakes and, what is remarkable, it cannot be compensated by increasing sample size.

Discussion

The described Monte Carlo simulation is a useful tool to analyse sampling strategies.

Stratified sampling ALK being preferable to random sampling has been well established in literature (i.e.: Ching-Ping 2009), but the method here describe allows to quantify goodness of the sampling strategy.

Even the stratified sampling strategy has the best behaviour when catch is high, it will become close or equal the random strategy when catch is low, so low that only some length classes could be completed. It means that the strategy *per se* does not guarantee the quality of the sampling, but enough number of otoliths well distributed along the whole length interval.

Ageing mistakes cannot be identified by a retrospective index such as $\sigma 1$.

References

Abramowitz, M. and I.A. Stegun (eds.) - 1972. Handbook of mathematical functions. Dover, NY, 1046 pp.

Ching-Ping, C.– 2009. Evaluation of the Sampling Efficiency of Three Otolith Sampling Methods for Commercial King Mackerel Fisheries. *Trans. Am. Fish. Soc.* 138:5, 990-999.

Darby C.D. and S. Flatman – 1994. Virtual Population Analysis: Version 3.1 (Windows/DOS) User Guide. *Information Technology Series, No. 1. Lowestoft: MAFF, Directorate of Fisheries Research,* 85 pp.

Doubleday, W.G.– 1981. Manual of Groundfish Surveys in the Northwest Atlantic. *NAFO Sci. Counc. Studies* 2.

Gavaris, S.– 1988. An adaptive framework for the estimation of population size. *Can. Atl. Fish. Adv. Comm. (CAFSAC) Res. Doc.* 88/29.

Ketchen, K. S.– 1949. Stratified subsampling for determining age distributions. *Trans. Am. Fish. Soc.* 79:205-212.

NAFO – 1980. Revised outline of NAFO Sampling Program. *NAFO SCS Doc.* 80/VI/20.

NAFO – 1999. Inventory of Sampling Data 1990-1994. *NAFO*.

Vázquez, A. and M. Mandado – 1010. Random Retrospective Pattern in Fish Stock Assessment. *NAFO SCR Doc.* 10/36.

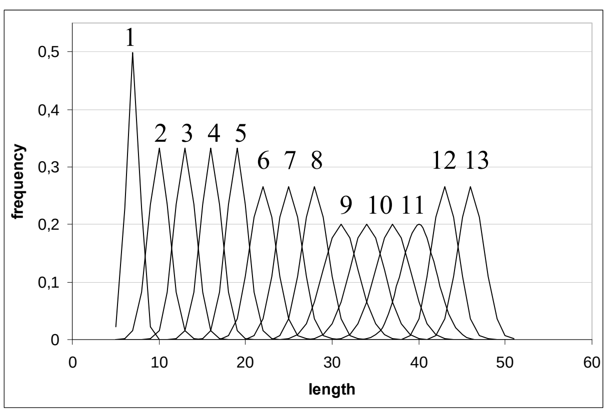

Figure 1 –Length frequencies distribution as assigned to each age.

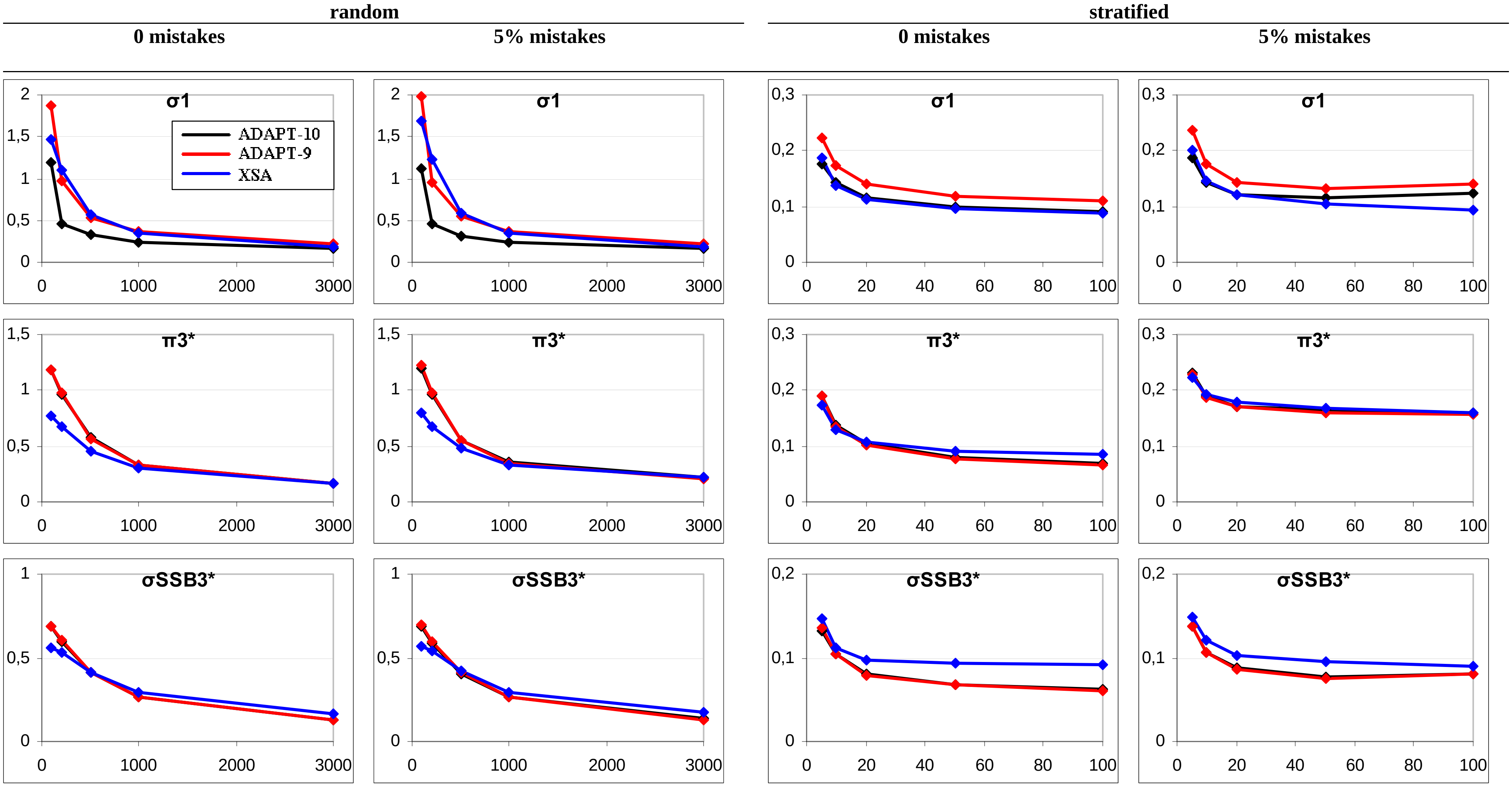

Figure 2 – Retrospective index (σ1) and inaccuracy indices (π3*, σSSB3*) vs sample size. Random/Stratified, no mistakes/5% mistakes

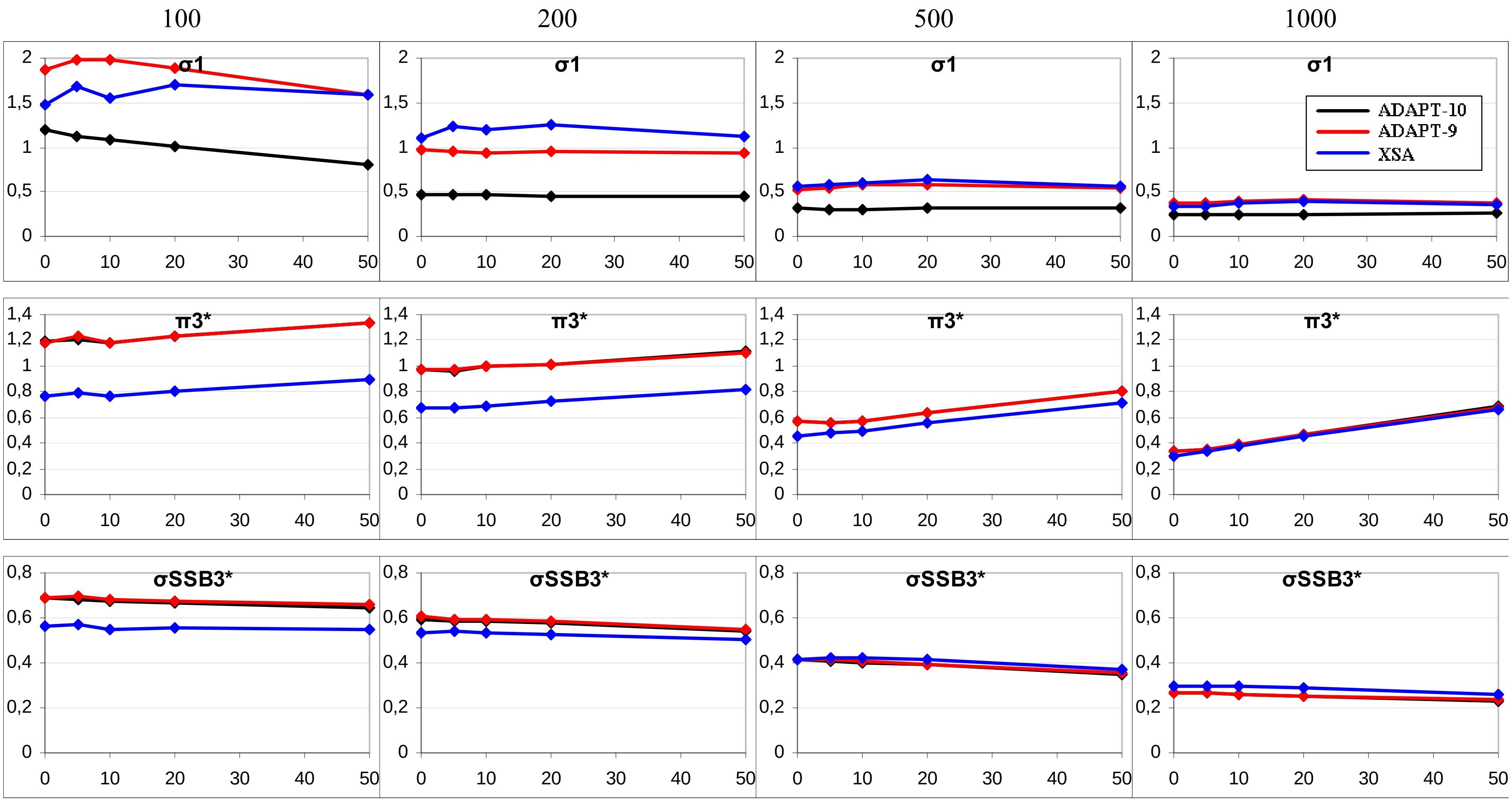

Figure 3 – Otoliths' random selection. Retrospective index (σ1) and inaccuracy indices (π3*, σSSB3*) vs mistakes in age determination (from 0 to 50%) for 4 levels of sample size: 100, 200, 500 and 1000 otoliths.

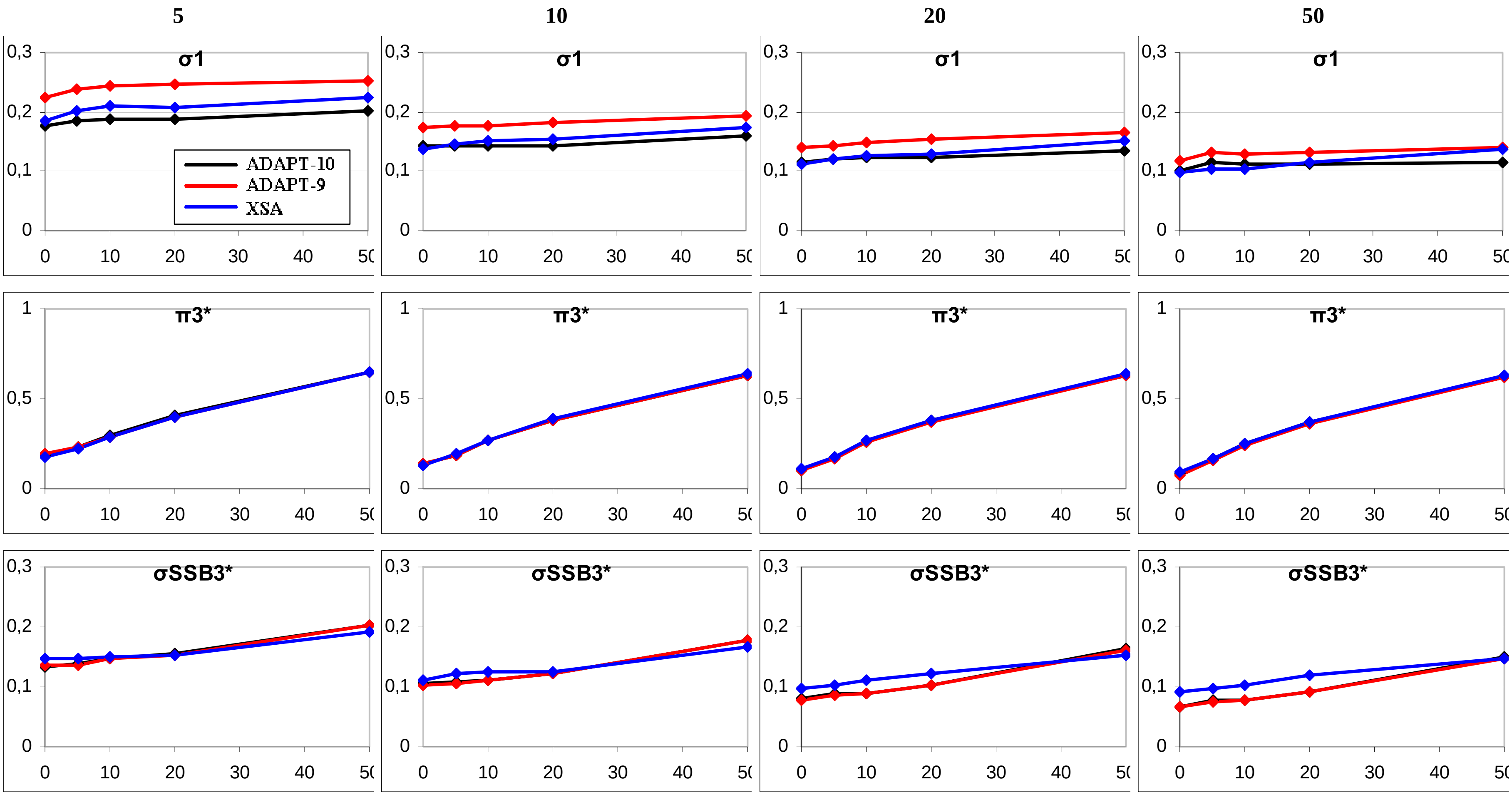

Figure 4 – Otoliths' stratified selection. Retrospective index (σ1) and inaccuracy indices (π3*, σSSB3*) vs mistakes in age determination (from 0 to 50%) for 4 levels of sample size: 5, 10, 20 and 50 otoliths per length class.