Northwest Atlantic Fisheries Organization

Serial No. N5002 **NAFO SCR Doc. 04/49**

SCIENTIFIC COUNCIL MEETING – JUNE 2004

Finding the Magical Minimum Sample Size: a Computer-intensive Approach to Minimize Re-aging Effort to Construct Age-length Keys for Yellowtail Flounder

by

K. S. Dwyer, M. Koen-Alonso and S. J. Walsh

Department of Fisheries & Oceans, Science, Oceans & Environment Branch
NW Atlantic Fisheries Centre, P.O. Box 5667
St. John's, NF, Canada, A1C 5X1

Abstract

Yellowtail flounder (*Limanda ferruginea*) is currently assessed using a stock-production model. The goal is to replace this model with an age-structured model; however validation studies show that the current method for ageing older fish is inaccurate and re-ageing the historical collection of otoliths in its entirety is not possible. Therefore our objective is to determine the minimum number of otoliths that need to be re-aged to generate a valid age-length key from a subsample of the full collection.

All otoliths (n = 1747) from the 1998 spring and fall research surveys were re-aged, and used to build all-year full data age-length matrices (FDALM). These FDALM's were considered "target keys" for which we implemented a subsampling routine to reproduce them from samples of smaller sizes. Random subsamples (without replication) of different sizes were taken from the corresponding full databases, and subsampled age-length matrices (SSALM) were produced, and compared with the corresponding FDALM using a bidimensional two-sample Kolmogorov-Smirnov test. We used the obtained p-values as a measure of the similarity between the keys. To assess variability, random sampling was performed 400 times at each subsample size, and the median and 95th percentile range of the p-values were obtained. We used plots of the median p-values against subsample size to determine the size that produces the desired similarity with the full sample (i.e. a median p-value close to one with small variability).

Our results indicate that small sample sizes (less than 10% of the full sample) generated non-significant differences between the SSALM and FDALM but with large variability, while subsample sizes around 60% of the full sample were similar enough to be considered adequate.

Introduction

Yellowtail flounder (*Limanda ferruginea*) (Storer, 1839) is a small-mouthed right-eyed pleuronectid found on the Grand Bank off Newfoundland (Div. 3LNO), where there is a commercial fishery for this species. These fish were traditionally aged for fisheries research using surface-read whole otoliths. Age determination of otoliths from recaptures of fish tagged in the early 1990's indicated that the traditional ageing technique was underestimating the ages of yellowtail flounder when compared with the time at liberty and studies were undertaken to find an accurate method for ageing yellowtail. Using validation studies such as tag-recapture analysis and bomb radiocarbon assays, it was found that thin-sectioned otoliths were the most accurate method for ageing yellowtail flounder (Dwyer *et al.*, 2003).

Yellowtail flounder is currently assessed using a stock-production model, which requires only biomass data. The goal is to replace this model with an age-disaggregated model; however, in order to attempt this, it is necessary to re-age historical archives of otoliths, which date back to the 1960s and number in the thousands (approximately 46 000). It becomes obvious that a substantial amount of effort and resources need to be allocated to re-ageing all previously collected samples. Since a significant amount of effort was also invested in building the current collection, it is reasonable to expect that any subsampling program should try to preserve as much as possible the information contained in the original full size sample.

Age-length keys (ALKs) are rectangular arrays which express in proportions the distribution of observations per age (columns) in each length-class (rows) (Hayes, 1993). An ALK is built from a matrix of number of observations per age by length (ALM for age-length matrix), and it is this ALM the one typically used for statistical comparisons (Haynes, 1993; Horbowy, 1998). These methods rely on multiple Chi-squared and/or Fisher's tests (Haynes, 1993; Horbowy, 1998), and they typically require the pruning of the matrices in order to compare meaningful arrays of identical size (i.e. where a zero value can be safely considered an informative datum) (e.g. Horbowy, 1998).

In spite of these standard procedures, we followed a different approach. By contrast with previous work (Haynes, 1993; Horbowy, 1998), our conceptual starting point was not the ALM. Many of the statistical difficulties in comparing ALMs are associated with the process of defining age and length classes (this is obvious for length, but it is still valid for age). Our approach does not require any age or length class definitions, and hence, the full sample can be used (i.e. there is no loss of data), and our arbitrary ways of classifying the data have no influence on the result. Furthermore, the comparison of two ALMs is done with a single statistical test, instead of many (typically as many as length classes are considered), and a single p-value can be used to make decisions about the similarity of the two ALMs. Here we use the p-values of the bidimensional Kolmogorov-Smirnov (2D K-S) test (Peacock ,1983; Fasano and Franceschini, 1987; Press *et al*., 1992) as a measure of the degree of similarity between two ALMs (full sample data versus subsampled data).

Therefore, a computer-intensive subsampling routine, along with the 2D K-S test, was considered in order to determine the minimum number of otoliths that need to be re-aged to construct adequate age-length keys for historical otoliths.

Materials and Methods

Age determination

Ages were obtained from all otoliths from yellowtail flounder (Div.3LNO) in 1998 (spring and fall DFO RV Surveys) which is the most recent year that ages had been initially obtained using the whole otolith method. These old ages were compared with the ages from the new thin-section otolith method. The ages obtained from the new method (n = 1747; range from 0 year to 30 years) were used to build all-year full data age-length matrices (FDALM) (Table 1), which were considered "target keys". Otoliths were read by both readers, and consensus ages were derived to minimize any possible reader effect in subsequent analyses. To avoid subjectivities, consensus ages were derived through the following algorithm: if ager 1 = ager 2 then ager 1, if ager 1>ager 2 in fall, then ager 2, if ager 1 > ager 2 in spring, then ager 1. This algorithm captures (and corrects) the two most common ageing mistakes. In fall, the main mistake is taking into account the ring on the edge which is not considered a full years' growth, and in the spring, it is opposite, the main mistake is not taking into account the fact that the newest ring may not be visible yet.

Whole left otoliths were originally read by the age readers by counting the annuli on the surface, occasionally using a grinding wheel to make the rings more visible. To re-age, the left otolith was embedded in wax and an Isomet low-speed saw was used to cut a transverse section (approximately 0.3-0.8 mm-thick) through the primordium, along the dorso-lateral plane of the otolith. Because of the brittle nature of smaller otoliths, those from fish less than 30 cm were cut transversely and each half of the otolith was examined.

Whole otoliths were placed in a black chamber containing ethyl alcohol and examined using a Nikon dissecting microscope under reflected light at 8-20X magnification according to the traditional methods of examination at NAFC. The number of winter-hyaline (translucent) zones (annuli) was recorded. Sectioned otoliths were also examined in alcohol using magnifications of 25-40X with reflected light.

Computer-intensive resampling scheme and statistical analyses

We started by assuming that the data used to build an ALM was a random sample from an underlying and unknown bivariate distribution. If we want to compare two univariate random samples and test if they were drawn from the same underlying one-dimensional statistical distribution, the classical two sample Kolmogorov-Smirnov test is an appropriate choice (Sokal and Rohlf, 1995; Conover, 1999; Zar, 1999). In the case of comparing two ALMs, we have exactly the same situation but with a bidimensional (2D) underlying distribution. There is a 2D version of the Kolmogorov-Smirnov (2D K-S) test (Peacock, 1983; Fasano and Franceschini, 1987), and there is available computer code to implement it (routine ks2d2s in Press *et al.*, 1992).

We used the p-values of the 2D K-S test as a measure of the degree of similarity between two ALMs. Larger p-values indicate a high degree of similarity, while small values indicate poor agreement, or even statistical differences (p-value<0.05).

Random sub-samples of different sizes were taken, without replication, from the full collection of already re-aged fish with known length (FDALM). Each of these sub-sampled age-length matrices (SSALM) was compared with the full collection using the 2D K-S test. For each (sub)sample size, we generated 400 independent random samples, and we plotted the median and 95th percentile range of the corresponding p-values as a function of sample size. Logically, as sample size increases, the p-value will approach to 1 because the sub-samples approach to the full collection, and its variability will also be reduced. However, the shape of this plot and the p-value variability at each sample size provide a simple way of visualizing the trade-offs implied when selecting a given sample size. As a starting point, we performed the analysis for the full 1998 combined surveys, and then discriminated by survey (spring and fall surveys separately). Most numerical procedures (i.e. resampling and 2D K-S tests) were implemented in Fortran 77. We used plots of the median p-values against subsample size to determine the size that produces the desired similarity with the full sample (i.e. a p-value close to one with small variability).

Because it was believed that there is no difference in ages derived from both ageing methods for fish up to 25 cm (Dwyer *et al.*, 2002), after which, differences in ageing became significant, we also compare old and new ages with for these fish (<26 cm) using linear regression analysis.

Results and Discussion

Age-length matrices are shown for combined spring and fall surveys (Table 1), spring survey (Fig. 2) and fall survey (Table 3).

Our results indicate that small sample sizes (less than 10% of the full sample) generated non-significant differences between the SSALM and FDALM but with large variability, while subsample sizes around 60% of the full sample were similar enough to be considered adequate (that is, the lower bound of the 95th percentile range had a p-value around 0.7-0.8). This was true for spring (Fig. 1), fall (Fig. 2) and combined ALMs (Fig. 3).

However, the comparison of the ALMs between spring and fall for 1998 were statistically different (2D K-S statistic d=0.1362 p-value=3.2E-06), indicating that combined ALMs should not be used.

The linear regression analysis confirmed that the slope of the regression between old and new ages was not significantly different than 1 for fish under 25 cm [slope=1.0037, 95% confidence interval: 0.9852-1.0223, p-value<0.0001, R^2 = 0.86, n=194] (Fig. 4). In the context of the subsampling scheme developed here, this result allows us to continue to use whole otoliths for those fish under 25cm included in the subsample. This reduces even further the actual number of otoliths that need to be re-aged. Therefore, based on our results, of the 1747 total otoliths re-aged from the 1998 surveys, we feel that approximately 1048 (60%) would actually need to be re-read in order to produce an ALM similar to the FDALM. Since 34% of the total 1747 otoliths were from fish under 25 cm, then it is likely that in actual fact, this number would be even less (approximately 695).

If we consider the differences between the ALMs from the fall and spring surveys, and apply the same rationale for each survey, the expected number of otoliths that will actually need to be re-aged will be around 314 for the spring survey, and 381 for the fall survey.

Finally, it is important to remember that these results are based on the surveys from a single year. Before generalizing these results it is important to assess their robustness and consistency. We believe that at least one more year needs to be fully re-aged and this analysis repeated before we can be confident in the percentage of otoliths needed to be re-aged that we obtained here. Furthermore, it would be preferable that this additional year to be as dissimilar in length composition from 1998 as possible, perhaps a sample from before 1980, when fishing pressure was not as high. If the results are the same, then the basis for generalizing these results will start to look tenable. We should understand that a substantial amount of resources will be necessary to repeat this exercise, and continue re-ageing the archived otoliths. This work is essential in order to ensure development of a VPA for yellowtail flounder and it must remain a priority.

References

Conover, W.J. 1999. Practical nonparametric statistics (Third Edition). Wiley Series in Probability and Statistics: Applied Probability and Statistics Section, John Wiley and Sons Inc., New York.

Dwyer, K. S., S.J. Walsh and S.A. Campana. Age determination, validation and growth of Grand Bank yellowtail flounder (*Limanda ferruginea*). ICES J. Mar. Sci. 60: 1123-1138.

Fasano, G., and A. Franceschini. 1987. A multidimensional version of the Kolmogorov-Smirnov test. Monthly Notices of the Royal Astronomical Society 225:155–170.

Hayes, D.B. 1993. A statistical method for evaluating differences between age-length keys with application to Georges Bank haddock, *Melanogrammus aeglefinus*. Fishery Bulletin 91:550-557.

Horbowy, J. 1998. Comparison of age-length keys of Baltic cod derived from Polish commercial and research data. Fisheries Research 36: 257-266.

Peacock, J.A. 1983. Monthly Notices of the Royal Astronomical Society 202:615–627.

Press, W.H., B.P.Flannery, S.A. Teukolsky, and W.T. Vetterling. 1992. Numerical Recipes in Fortran 77: The art of scientific computing (Second Edition). Cambridge University Press, New York.

Sokal, R.R., and F.J. Rohlf. 1995. Biometry (Third Edition). W.H. Freeman and Co., New York.

Zar, J.H. 1999. Biostatistical analysis (Fourth Edition). Prentice Hall, Upper Saddle River.

Table 1. Full data age-length (consensus ages) fir 1998 RV surveys spring and fall combined.

Length	Age 0	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8	Age 9	Age 10	Age 11	Age 12	Age 13	Age 14	Age 15	Age 16	Age 17	Age 18	Age 19	Age 20	Age 21	Age 22	Age 23	Age 24	Age 25	Ag
5.5	1																										
6.5		1																									
7.5		7																									
8.5		13	2																								
9.5		12	6																								
10.5		3	19																								
11.5			21	4																							
12.5			16	11																							
13.5			20	11																							
14.5			18	18	2																						
15.5			10	10	0																						
16.5			11	16	5																						
17.5			10	15	14	4																					
18.5			2	15	9	2																					
19.5			2	18	27	6																					
20.5				12	22	10	1																				
21.5				10	29	15																					
22.5				5	22	15																					
23.5				3	16	29	1																				
24.5				1	9	25																					
25.5					6	43	6	1		1																	
26.5					3	30	12																				
27.5						37	16	1																			
28.5					2	26	23																				
29.5					1	14	30	5							1												
30.5					1	13	35	8		3																	
31.5						7	25	9	2	3																	
32.5						1	37	16	3		1	1	1														
33.5						2	18	22	3	3	1					1											
34.5							15	23	11	3	1	1	3														
35.5							5	16	16	6	2	4	2														
36.5							4	8	22	11	7		1														
37.5							3	4	22	13	5		1	2			1			1							
38.5								5	10	18	6	8	4	2		1	1	1									
39.5								1	8	19	6	5	3	3	2			1	1								
40.5									2	11	15	13	8	3	2	1		2	1								
41.5							1	1	1	4	6	16	6	7	1	2	3				1						
42.5									3	7	5	11	12	8	5	2	1	4	1								
43.5										2	4	7	4	3	5	1	5		1	1		1					
44.5										3	4	6	9	4	3	3		2	2		1						
45.5											4	1	2	4	5	4	3		3	4							
46.5										1	1	3	4	3	6	4	3	3	2				1				
47.5											2	2	1	4	3	1	5	2						1			
48.5											1		3	1	3	2	1	2	4	1		1	1			1	
49.5														1	1		2			2		3					
50.5															1		1	1	1	1					1		
51.5														1					1		1						
52.5															1												
53.5																				1							
54.5																											
55.5																						1					
56.5																						1					
58.5																			1								
Total	1	36	137	149	168	279	232	120	103	108	71	78	64	46	39	22	26	18	18	11	3	7	2	1	1	1	

Table 2. Age length matrix (consensus ages) fir 1998 RV fall survey.

Length	Age 0	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8	Age 9	Age 10	Age 11	Age 12	Age 13	Age 14	Age 15	Age 16	Age 17	Age 18	Age 19	Age 20	Age 21	Age 22	Age 23	Age 26	
5.5	1																									1
6.5		1																								1
7.5		7																								7
8.5		13	1																							14
9.5		12	2																							14
10.5		3	9																							12
11.5			16																							16
12.5			15																							15
13.5			20	2																						22
14.5			18	1																						19
15.5			10	4																						14
16.5			11	6																						17
17.5			10	9	1																					20
18.5			2	14	1																					17
19.5			2	17	4																					23
20.5				11	12																					23
21.5				10	18																					28
22.5				5	15																					20
23.5				3	14	6																				23
24.5				1	9	7																				17
25.5					6	24																				30
26.5					2	19	1																			22
27.5						26	1																			27
28.5					2	21	6																			29
29.5					1	13	10																			24
30.5					1	13	20			1																35
31.5						7	13	2		1																23
32.5						1	28	4																		33
33.5						2	14	12	2		1					1										32
34.5							14	13	3	1		1	1													33
35.5							5	12	9		1	1														28
36.5							4	6	12	4	3															29
37.5							3	4	14	5	1			2												29
38.5								4	10	10	1	3	1	1												30
39.5								1	7	12	4		1	2	1											28
40.5									2	8	10	7	3		1			2								33
41.5							1	1	1	3	3	13	2	1			3									28
42.5										5	2	7	4	3	1			3								25
43.5										2	4	4	3	2	1	1	3		1			1		1	1	24
44.5										2	3	5	7	1	1	2			1							22
45.5											3	1	2		1	1			1	3						12
46.5										1	1	2	3	1	1	2	1	3	1				1			17
47.5											1	2	1	3	2		1	2								12
48.5											1		3	1			1		2	1						9
49.5														1	1					1		3				6
50.5															1			1	1	1						4
51.5																					1				1	2
52.5															1											1
53.5																		1		1						2
54.5																										0
55.5																										0
56.5																						1				1
58.5																			1							1
Total	1	36	116	83	86	139	120	59	60	55	39	46	31	18	12	7	9	12	8	7	1	5	1	1	2	954

Table 3. Age length matrix (consensus ages) for 1998 RV spring survey.

Length	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8	Age 9	Age 10	Age 11	Age 12	Age 13	Age 14	Age 15	Age 16	Age 17	Age 18	Age 19	Age 20	Age 21	Age 22	Age 24	Age 25	Age 30	
5.5																									0
6.5																									0
7.5																									0
8.5	1																								1
9.5	5																								5
10.5	11																								11
11.5	5	4																							9
12.5	1	11																							12
13.5		9																							9
14.5		17	2																						19
15.5		6																							6
16.5		10	5																						15
17.5		6	13	4																					23
18.5		1	8	2																					11
19.5		1	23	6																					30
20.5		1	10	10	1																				22
21.5			11	15																					26
22.5			7	15																					22
23.5			2	23	1																				26
24.5				18																					18
25.5				19	6	1		1																	27
26.5			1	11	11																				23
27.5				11	15	1																			27
28.5				5	17																				22
29.5				1	20	5							1												27
30.5					15	8		2																	25
31.5					12	7	2	2																	23
32.5					9	12	3		1	1	1														27
33.5					4	10	1	3																	18
34.5					1	10	8	2	1		2														24
35.5						4	7	6	1	3	2														23
36.5						2	10	7	4		1														24
37.5							8	8	4		1				1										22
38.5						1		8	5	5	3	1		1	1	1									26
39.5							1	7	2	5	2	1	1			1	1								21
40.5								3	5	6	5	3	1	1			1								25
41.5								1	3	3	4	6	1	2					1						21
42.5								2	3	4	8	5	4	2	1	1	1								31
43.5										3	1	1	4		2			1							12
44.5								1	1	1	2	3	2	1		2	1		1						15
45.5									1			4	4	3	3		2	1							18
46.5										1	1	2	5	2	2		1								14
47.5									1			1	1	1	4										8
48.5													3	2		2	2			1	1		1		12
49.5															2			1							3
50.5															1							1			2
51.5												1					1								2
52.5																								1	1
53.5																									0
54.5																									0
55.5																				1					1
56.5																									0
58.5																									0
Total	23	66	82	140	112	61	40	53	32	32	33	28	27	15	17	7	10	3	2	2	1	1	1	1	789

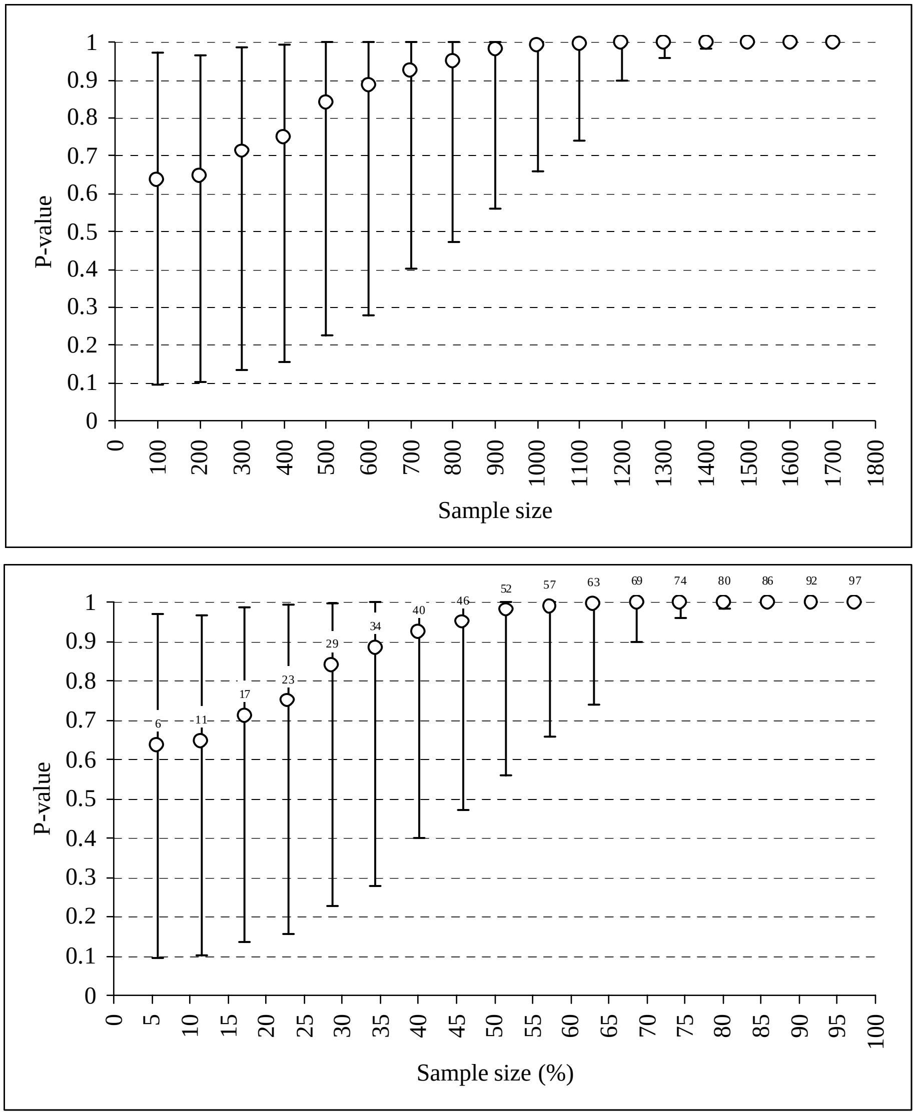


Fig.1. Comparison of FDALM with SSALM for the total 1998 sample (spring and fall data combined) using 2D Kolmogorov-Smirnov Test. Circles and bars correspond to the median and 95th percentile ranges at each subsample size. Sample replicates = 400. The small figures at each sample size indicate the exact percentage of the full sample.

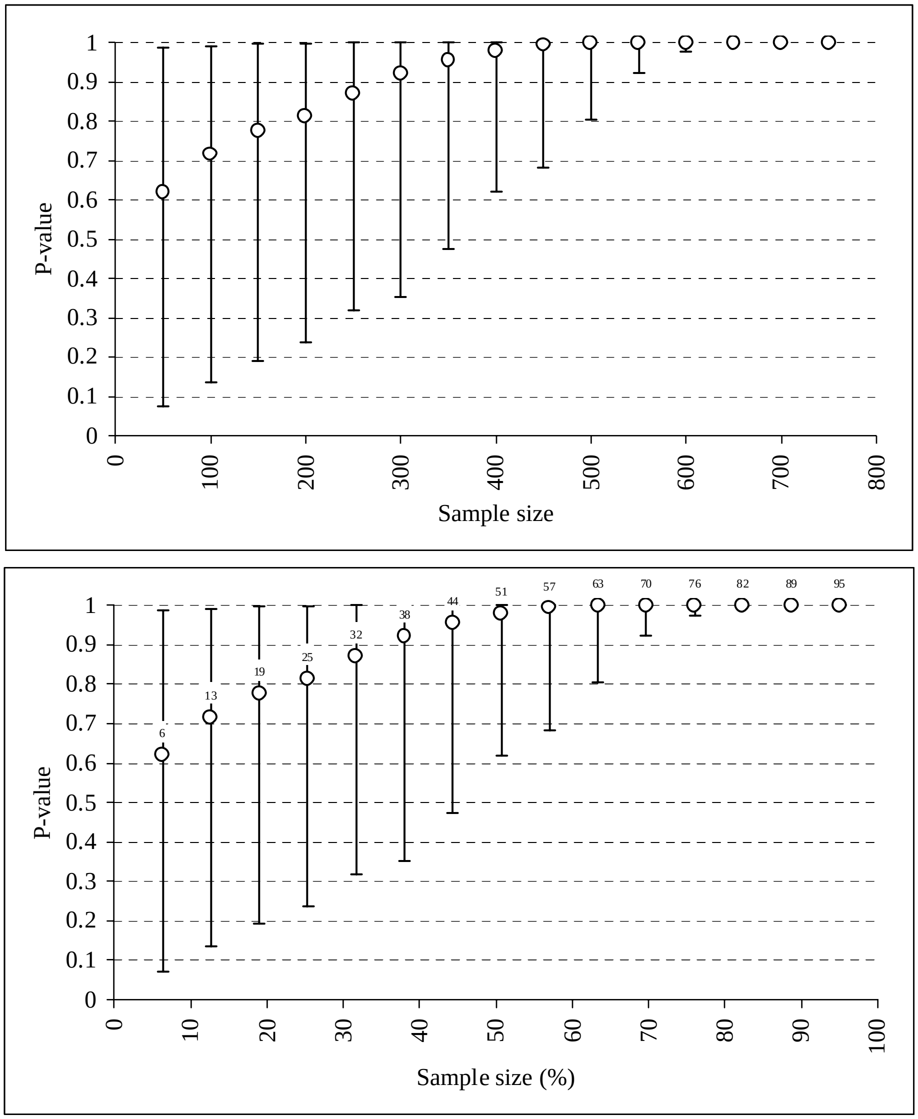
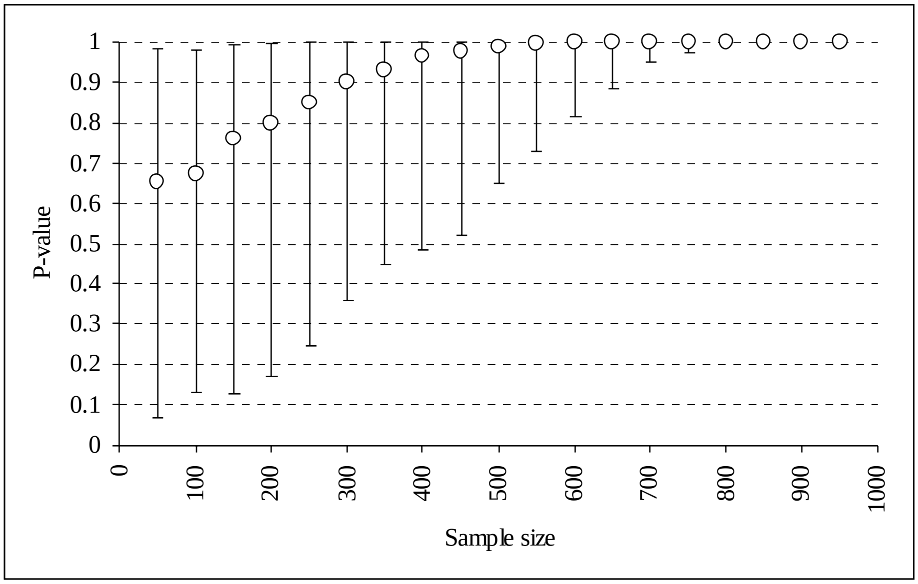


Fig. 2. Comparison of FDALM with SSALM for the Spring sample only using 2D Kolmogorov-Smirnov Test. Circles and bars correspond to the median and 95th percentile ranges at each subsample size. Subsample replicates = 400. The small figures at each subsample size indicate the exact percentage of the full sample.



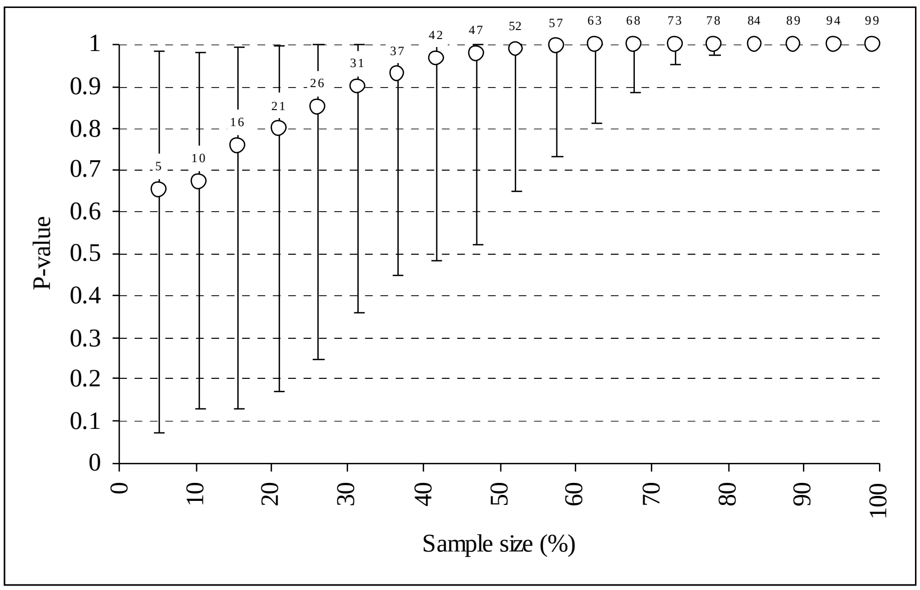


Fig. 3. Comparison of FDALM with SSALM for the Fall sample only using 2D Kolmogorov-Smirnov Test. Circles and bars correspond to the median and 95th percentile ranges at each subsample size. Subsample replicates = 400. The small figures at each subsample size indicate the exact percentage of the full sample.

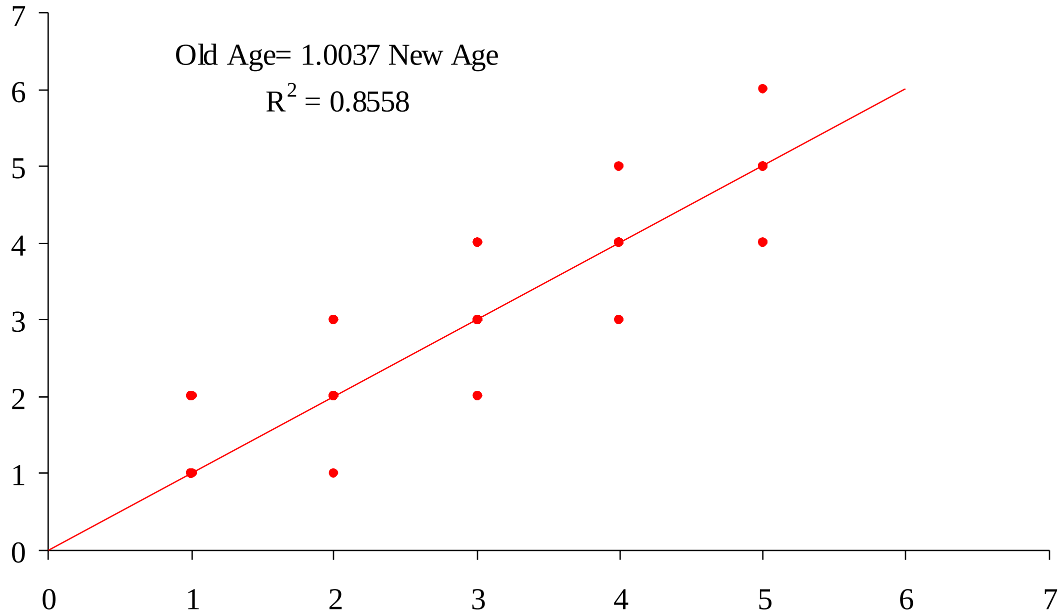


Fig. 4. Linear regression of old ages (whole otoliths) *versus* new ages (thin-sectioned otoliths) using consensus ages for fish <26 cm from 1998 spring and fall surveys combined.