



**Abstract**—Age determination is important for estimating productivity and status in fisheries stock assessments. Aging methods must balance advances in technology with continuity of data for long-term sampling programs. The long-term sampling program for Atlantic menhaden (*Brevoortia tyrannus*) and Gulf menhaden (*B. patronus*) has resulted in a database of ages that were determined by using an Eberbach projector, a now outdated technology. The objective of this study was to compare ages determined with an Eberbach projector to those determined with a more recent technology, the stereo microscope. Scales from Atlantic (total number of fish [N]=1317) and Gulf menhaden (N=1569) were aged by using both an Eberbach projector and a stereo microscope, and results were compared by using percent agreement, average percent error, Chang's average coefficient of variation, bias tests, and simultaneous multinomial confidence intervals. The results from these measures for comparing age estimates were generally within standard, expected levels. Some bias was detected between estimates from the use of the 2 devices but was likely due to the use of poorly preserved scales or images of low quality on the Eberbach projector. Our results indicate that the use of a microscope will help maintain continuity in age estimates over time for long-term monitoring and for stock assessments.

Manuscript submitted 2 July 2020.  
Manuscript accepted 8 March 2021.  
Fish. Bull. 119:21–32 (2021).  
Online publication date: 5 April 2021.  
doi: [10.7755/FB.119.1.4](https://doi.org/10.7755/FB.119.1.4)

The views and opinions expressed or implied in this article are those of the author (or authors) and do not necessarily reflect the position of the National Marine Fisheries Service, NOAA.

## Comparison of ages determined by using an Eberbach projector and a microscope to read scales from Atlantic menhaden (*Brevoortia tyrannus*) and Gulf menhaden (*B. patronus*)

Amy M. Schueller (contact author)

Amanda Rezek

Raymond M. Mroch III

Eric Fitzpatrick

Alicia Cheripka

Email address for contact author: [amy.schueller@noaa.gov](mailto:amy.schueller@noaa.gov)

Beaufort Laboratory  
Southeast Fisheries Science Center  
National Marine Fisheries Service  
101 Pivers Island Road  
Beaufort, North Carolina 28516

Determining the health of fish populations is of the utmost importance to ensure that fishery yields are sustainable over time. Knowledge of the life histories of species is integral to understanding the population dynamics and factors that may influence the sustainability of a fishery. Several life history statistics, including size estimates for fishable year classes, can be determined by knowing the age of individuals caught (June and Roithmayr, 1960; Campana, 2001). Collecting these age data can be done by grouping fish by length frequencies or by counting annuli on otoliths or scales (Secor et al., 1995; Campana and Thorrold, 2001). Determining age through the use of otoliths can sometimes be impractical because of the time and effort required to extract, prepare, and age the structure. The use of scale samples can yield age data, and the collection and preparation time required is significantly less for using scales than for using otoliths (Nicholson and Schaaf, 1978; VanderKooy<sup>1</sup>). Aging with scales

also does not require sacrificing the fish (Khan and Khan, 2009; Vilizzi, 2018).

Age data from collected scales and otoliths are used to calculate statistics such as productivity of a population, growth rate, and mortality rate (June and Roithmayr, 1960; Campana, 2001). These calculated variables form the basis of stock assessments, which provide information essential to the management of fisheries. Specifically, age data include catch composition and the age classes represented by survey data and can be used to track cohorts through the population over time and to estimate life history characteristics, such as growth rate, natural mortality, and fecundity (Beamish and McFarlane, 1995; Morison et al., 1998). All of these data are used to estimate parameters in stock assessments; therefore, the process of aging fish is critically important.

Atlantic menhaden (*Brevoortia tyrannus*) and Gulf menhaden (*B. patronus*)

<sup>1</sup> VanderKooy, S. (ed.). 2009. Gulf menhaden *Brevoortia patronus*. In A practical handbook for determining the ages of Gulf of

Mexico fishes, 2nd ed., p. 5-70–5-73. Gulf States Mar. Fish. Comm., Publ. 167. Gulf States Mar. Fish. Comm., Ocean Springs, MS. [Available from [website](#).]

are 2 commercially important species in the United States for which the stocks are routinely assessed (SEDAR, 2015, 2018). Both species are schooling clupeids pursued by a large, commercial purse seine fishery, called the *reduction fishery*. Additionally, a smaller bait fishery pursues Atlantic menhaden and captures about a quarter of the coastwide catch in U.S. waters of the Atlantic Ocean (SEDAR, 2015). Atlantic menhaden are caught along the Atlantic coast of the United States, mainly from Virginia to New Jersey. Gulf menhaden are taken in the Gulf of Mexico with most of the catch coming from the waters of Louisiana and Mississippi. These 2 fisheries are the largest by volume on their respective coasts and together compose the second-largest fishery by volume in the United States (NMFS<sup>2</sup>). Both species are important forage species and are eaten by other popular game fish species. The stock assessments for each species are reliant on the age data that have been provided from the Beaufort Laboratory of the NOAA Southeast Fisheries Science Center since 1955 for Atlantic menhaden and since 1964 for Gulf menhaden.

Aging of fish must evolve with time and technology, while continuity of data for long-term sampling programs is maintained. Traditionally, for both menhaden species, readings of scales to estimate ages were done at the Beaufort Laboratory by using an Eberbach<sup>3</sup> projector (Eberbach Corp., Ann Arbor, MI) that was built in the 1930s. However, the projector and its components are now obsolete, and replacement components are difficult to find. Given the challenges and risks of failure that arise from the use of obsolete equipment, alternative aging methods needed to be established. The objective of this study was to determine if ages for Atlantic and Gulf menhaden estimated by using a stereo microscope would be similar to those determined by using the Eberbach projector, with particular emphasis on precision or repeatability and on bias in age estimates between devices and methods. The aim of this study was to provide an alternative, more accessible method for aging menhaden in the future, a method that will decrease the likelihood of system failure as well as allow digital storage of sample images and information.

## Materials and methods

Scales for this study were obtained from a long-term monitoring program that regularly samples landings of both the reduction and bait fisheries, in a 2-stage cluster design at various processing plants and dock locations along the Atlantic and Gulf coasts, geographically and temporally covering the entirety of the combined fishery (SEDAR, 2015, 2018). Samples used for this study were collected in 2013 and 2017 for Atlantic menhaden and in 2005 and

2017 for Gulf menhaden. Port agents retrieved a bucket of fish from the top of a randomly selected vessel's fish-hold dockside to characterize the fish caught in the final set of the trip. The agent verified catch data and then selected 10 fish at random from the bucket and recorded fork length (in millimeters) and weight (in grams). From the 10 fish, scales were removed from an area centered in line with the foremost part of the dorsal fin along the lateral line. For each individual fish, 6–10 scales were cleaned and mounted between 2 glass microscope slides. Each scale sample was labeled by processing plant (if from the reduction fishery) or state (if from the bait fishery) and by year and was assigned a scale number or ID code. All scale samples were then used to estimate fish ages with an Eberbach projector and with a stereo microscope (Olympus SZX16, Olympus Corp., Tokyo, Japan).

Age was determined by counting the number of continuous annuli between the focus and edge of the scale and by considering marginal growth as well as size, capture date, and each species' conventional birth date, a standard practice in some systems for the estimation of ages of fish sampled from commercial catches over a series of years (*production aging*) and in the Beaufort Laboratory's menhaden program since the 1960s (June and Roithmayr, 1960; Nicholson and Schaaf, 1978; Morison et al., 1998; VanderKooy<sup>1</sup>). The additional information was important to help place fish in the proper cohort (O'Sullivan<sup>4</sup>), given the protracted spawning seasons of these species, and to provide better, more timely age estimates (Campana, 2001).

For age estimation done with the Eberbach projector, mounted scale samples were placed under the projector's objective (40× magnification), and the best scale was chosen for aging. Ages were recorded on a data sheet. If all scales were unreadable or if data were missing for a fish, the sample for that fish was excluded.

The same scale samples used in aging with the Eberbach projector were observed by using a stereo microscope (7.5–16.0× magnification). An image of the best scale on each slide was taken by using image analysis software (cellSens, vers. 1.17, Olympus Corp.). Time between readings with the Eberbach projector and with the microscope for the primary reader (reader 1) varied from a few weeks to 3 years. Time between readings was the result of equipment and reader availability, sample selection intervals, and elapsed time between first readings during standard production aging for the long-term sampling program and the second readings for this study.

All scale samples were read on both instruments, and some were read by 2 readers. For Atlantic menhaden, samples were from fish caught in 2013 and 2017 and from the reference collection for this species at the Beaufort Laboratory. For Gulf menhaden, samples were from fish caught in 2005 and 2017 and from the reference collection for this species. These reference collections are sets of quality scale

<sup>2</sup> NMFS (National Marine Fisheries Service). 2018. Fisheries of the United States, 2017. NOAA, Natl. Mar. Fish. Serv., Curr. Fish. Stat. 2017, 142 p. [Available from [website](#).]

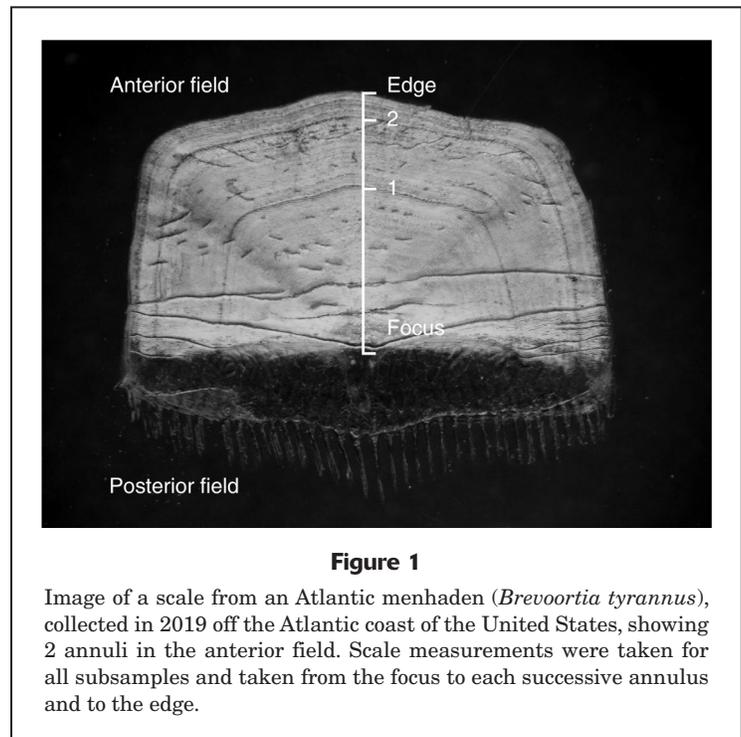
<sup>3</sup> Mention of trade names or commercial companies is for identification purposes only and does not imply endorsement by the National Marine Fisheries Service, NOAA.

<sup>4</sup> O'Sullivan, S. 2007. Fisheries long term monitoring program—fish age estimation review, 22 p. Dep. Prim. Ind. Fish., Brisbane, Australia. [Available from [website](#).]

samples with agreed upon ages from experienced readers, ages that are representative of all ages and the monthly seasonality typical in the fishery. The reference collections are used to assess aging consistency between readers and among estimates from a reader over time. The second reader (reader 2) estimated ages with scales on the Eberbach projector but only for Atlantic menhaden sampled in 2013, for Gulf menhaden sampled in 2005, and for both reference collections (SEFSC<sup>5</sup>). From 1969 through 2015, the same reader (reader 2) estimated ages of menhaden by using scales, and those estimates from the database at the Beaufort Laboratory (SEFSC<sup>5</sup>) compose the historical record of ages that was used in this study. Comparisons of ages determined with an Eberbach projector were made between reader 1 and reader 2 for each species for the sample years that could be compared.

A subsample of 200 scales each from the full set of scale samples from Atlantic and Gulf menhaden was used to estimate ages a second time by reader 1 with both the Eberbach projector and the microscope. The best scale for each sample was selected, and its position on the slide was recorded so that the same scale could be used for aging on both devices. Subsampling provided 2 age readings on each device for each scale, allowing comparison of aging errors for each instrument and between them. This subsampling allowed errors to be classified as random reading errors versus errors due to the effect of the device used. On average, for reader 1, 10–19 months separated the first and second readings for scale samples from Atlantic and Gulf menhaden on the Eberbach projector and 6 months separated readings with the microscope.

Measurements of scales were taken for all subsamples (Fig. 1). A sonic digitizer pen fitted on the Eberbach projector was used for measuring distance from the focus of the scale to each annulus and to the edge margin. Measurements were recorded in Eberbach units and later converted to millimeters. The same distances were measured in millimeters on the calibrated image taken with the microscope, by using the image analysis software. On each image, an origin was placed on the scale's focus. The point tool was used to mark each annulus from the point of origin up to and including the edge. A new embedded image was created that included the scale image and the recorded measurements, and measurements were exported for analysis to Microsoft Excel 2016 (Microsoft Corp., Redmond, WA). In addition, while the Eberbach projector was being used for aging, a ruled blue card was placed on top of the projector's screen to manually measure each annulus and edge measurement of a scale image, as was done during the early years of data collection by



**Figure 1**

Image of a scale from an Atlantic menhaden (*Brevoortia tyrannus*), collected in 2019 off the Atlantic coast of the United States, showing 2 annuli in the anterior field. Scale measurements were taken for all subsamples and taken from the focus to each successive annulus and to the edge.

the Beaufort Laboratory. A comparison of measurements for each scale was used to verify if the same annuli were counted on each device. This process allowed comparison of measurements taken by hand, with a sonic digitizer pen, and with image analysis software, temporally across the age data from the Beaufort Laboratory for both species of menhaden.

Both the full sample and the subsample of scales were chosen to reflect the seasonality of the fishery by month and the spatial spread of the fishery by location of the processing plants. In addition, samples were selected so that the length distribution of fish from which the samples were taken reflected the length distribution of all fish sampled from landings during a given year. Samples were chosen such that an adequate number of fish with younger and older ages were selected and such that the age composition matched that of the fishery samples taken in a given year.

Some scale samples were more than a decade old, and the quality of scales was inconsistent because of storage conditions. Although some scales remained in a quality good enough to age, others had debris or had degraded with time. Scales that were deemed unreadable by reader 1 on either of the instruments or that were missing or lost were excluded before analysis. First, we compared the number of excluded scales between readers and between devices. Second, we compared age readings for those scales that were legible. Finally, we compared successive annulus measurements between the Eberbach projector and the microscope.

Basic plots of ages by reader and device were created for each of the comparisons listed in the previous paragraph

<sup>5</sup> SEFSC (Southeast Fisheries Science Center). 2018. Unpubl. data. Menhaden Biostatistical Database. Beaufort Lab., Southeast Fish. Sci. Cent., Natl. Mar. Fish. Serv., NOAA, 101 Pivers Island Rd., Beaufort, NC 28516.

for both the full sample and subsamples. Plots of mean age with 95% confidence intervals were created, with each plot containing a 1:1 line, and are commonly used to examine age bias (Campana, 2001). We also produced plots of sample sizes by age. We explored the data to determine if the location of the processing plant, area, or month affected differences in aging for Atlantic and Gulf menhaden by comparing the distribution of the samples used in this study to the distribution of samples used in production aging of the long-term sampling program for processing plant, area, and month.

Several statistical tests were used to determine if the age estimates were significantly different between the 2 devices, between readers, and within readings of the same device and reader. The significance level ( $\alpha$ ) of 0.05 was used. Multiple approaches were needed to assess error, agreement, bias, precision, and symmetry (Campana et al., 1995); therefore, multiple tests are described and justified here.

Average percent error (APE; Beamish and Fournier, 1981; Campana et al., 1995), Chang's average coefficient of variation (ACV; Chang, 1982), and percent agreement (PA) were calculated to determine precision and agreement within the age data (Campana, 2001). All of these statistics are commonly used to compare age estimates. The APE and ACV can be artificially inflated by bias (Campana, 2001) and are related. In general use, APEs <5% are acceptable, as are ACVs of <5%; Campana (2001) calculated an average APE of 5.5% and an average ACV of 7.6% across studies. Values higher than those averages may indicate that structures are difficult to interpret or that readers lack training (Morison et al., 1998).

Three tests were used to determine symmetry within the paired age estimates between readers, by the same reader on the same device, and by the same reader on different devices: Bowker's (Bowker, 1948; Hoening et al., 1995), Evans and Hoening's (Evans and Hoening, 1998), and McNemar's (McNemar, 1947; Hettmansperger and Mckean, 1973) tests. These tests were fairly robust in simulation testing (McBride, 2015), but Evans and Hoening's test performed best. Bowker's test can detect bias with low APE and is unpooled, but results from this test have a higher incidence of type II errors (McBride, 2015). Evans and Hoening's test can detect bias at higher APEs and is pooled on the diagonal (McBride, 2015). Evans and Hoening (1998) found that their test was more powerful than Bowker's test and more general than McNemar's test. Finally, McNemar's test is a pooled test sensitive to small differences on one side of the diagonal (McBride, 2015). Each of these tests provide different information on precision and are not meant to corroborate results of the other tests.

Simultaneous multinomial confidence intervals were estimated to determine if the multinomial distributions differed significantly between age readings from the different instruments (Sison and Glaz, 1995; Zar, 1999). The intent of the use of these intervals was to determine if the age estimates and other information that would be provided to a stock assessment would be similar between the devices used for aging. Ultimately, the end products to

which the age estimates contribute are the stock assessments for Atlantic and Gulf menhaden. Therefore, if the overall age compositions produced for those assessments do not differ by the instrument used, either device could be used to successfully provide age compositions. The multinomial confidence intervals were calculated by using the MultinomialCI package (vers. 1.1; Villacorta, 2019) in R (vers. 3.5.3; R Core Team, 2019), and 95% confidence intervals were provided for each age.

Measurements of distances on scales made with the 3 methods (microscope, Eberbach projector, and blue cards) for subsamples were examined separately for the Atlantic and Gulf menhaden for each distance (e.g., focus to first annulus), excluding the final measurement from the focus to the edge of the scale. Box plots were generated to compare across measurement types. A repeated measures analysis of variance was used to determine if the distance measurements among the 3 methods for each increment were significantly different, given an  $\alpha$  of 0.05.

## Results

Total sample sizes for Atlantic and Gulf menhaden collected from landings were 1317 and 1569, respectively (Table 1). After exclusion of samples, 1119 and 1307 samples remained for Atlantic and Gulf menhaden, respectively. Atlantic menhaden samples were from sampling in 2013 ( $n=605$ ) and 2017 ( $n=418$ ) and from a reference collection ( $n=96$ ). Gulf menhaden samples were from sampling in 2005 ( $n=759$ ) and 2017 ( $n=499$ ) and from a reference collection ( $n=49$ ). Samples were excluded at a greater rate when examined with the Eberbach projector than with the microscope (Table 1). Scales of Atlantic menhaden were excluded 14% of the time when observed with the Eberbach projector and 8% of the time when examined with the microscope. Scales of Gulf menhaden were excluded 12% of the time when observed with the Eberbach projector and 10% of the time when examined with the microscope. A higher percentage of scale subsamples ( $n=200$ ) were excluded when they were examined with the Eberbach projector than with the microscope (Table 2). For both Atlantic and Gulf menhaden, neither location of processing plant nor area nor month were more indicative of differences in age than the other variables, with an even distribution of differences across those variables; therefore, we do not provide those specific results here.

Both the APE and ACV for age data from the full sample sets for each species were below the level considered a benchmark for aging of fish (5%; Morison et al., 1998). The APEs of the age estimates for all of the samples of Atlantic menhaden and Gulf menhaden were 2.7% and 3.5%, respectively (Table 1). The APEs for age estimates for samples in the reference collections were as low as 0.8% for Atlantic menhaden and as 0.0% for Gulf menhaden. The APEs for age data on Atlantic menhaden were 3.5% for samples taken in 2013 and 1.8% for samples collected in 2017. The APEs for age data on Gulf menhaden

**Table 1**

Precision and agreement statistics for ages estimated by examining scales on an Eberbach projector and a stereo microscope for Atlantic menhaden (*Brevoortia tyrannus*) and Gulf menhaden (*B. patronus*) sampled from landings along the Atlantic and Gulf coasts of the United States. Scales were from Atlantic menhaden collected in 2013 and 2017, Gulf menhaden collected in 2005 and 2017, and reference collections for each species. The statistical tests used to determine precision and agreement of age estimates included average percent error (APE), average coefficient of variation (ACV), and percent agreement. Total number of fish for which scales were used to determine ages (N), sample size used for analysis after exclusion of scales (*n*), and percentage of scale samples that were excluded are presented for each data set.

Data set	N	<i>n</i>	Samples excluded (%)		APE	ACV	Percent agreement
			Projector	Microscope			
Atlantic menhaden							
All	1317	1119	14%	8%	2.7%	3.8%	88.9%
2013	730	605	16%	8%	3.5%	5.0%	86.3%
2017	490	418	12%	9%	1.8%	2.6%	91.2%
Reference	100	96	4%	4%	0.8%	1.2%	95.8%
Gulf menhaden							
All	1569	1307	12%	10%	3.5%	5.0%	89.7%
2005	920	759	12%	9%	4.4%	6.2%	86.2%
2017	600	499	12%	11%	2.5%	3.6%	94.0%
Reference	49	49	0%	0%	0.0%	0.0%	100.0%

**Table 2**

Precision and agreement statistics for ages estimated by examining scale subsamples on an Eberbach projector and a stereo microscope for Atlantic menhaden (*Brevoortia tyrannus*) and Gulf menhaden (*B. patronus*) collected from landings in 2013 and 2017 along the Atlantic coast of the United States and in 2005 and 2017 along the Gulf coast, respectively. The statistical tests used to determine precision and agreement of age estimates included average percent error (APE), average coefficient of variation (ACV), and percent agreement. Reading 1 is the first reading of the sample set on both the projector and microscope, and reading 2 is the second reading on both instruments. Values in the "Projector" and "Microscope" rows are from comparisons of the 2 readings of the scale subsamples on the same device. *n*=sample size used for analysis.

Data set	<i>n</i>	APE	ACV	Percent agreement
Atlantic menhaden				
Reading 1	186	2.9%	4.2%	89.3%
Reading 2	185	3.1%	4.4%	89.2%
Projector	176	3.1%	4.4%	88.6%
Microscope	176	3.7%	5.2%	86.4%
Gulf menhaden				
Reading 1	177	9.7%	13.8%	83.6%
Reading 2	173	0.7%	1.0%	97.1%
Projector	157	8.7%	12.3%	84.7%
Microscope	157	4.1%	5.8%	92.4%

were 4.4% for samples taken in 2005 and 2.5% for samples collected in 2017. The ACV for age estimates for all of the samples of Atlantic menhaden was 3.8%, and the ACV for age estimates for all of the samples of Gulf menhaden was 5.0% (Table 1). The ACVs for age estimates for samples in the reference collections were as low as 1.2% for Atlantic menhaden and 0.0% for Gulf menhaden. The ACVs for age data on Atlantic menhaden were 5.0% for samples taken in 2013 and 2.6% for the samples collected in 2017. The ACVs for age data on Gulf menhaden were 6.2% for samples taken in 2005 and 3.6% for samples collected in 2017. Age estimates for the more recent samples, those taken in 2017, for each species had better APE and ACV than the age estimates for the older samples from 2013 and 2005 (Table 1).

Average percent errors for the subsamples were below 5% for all comparisons of age data except for 2 comparisons of estimates for Gulf menhaden (Table 2). Specifically, when the first readings of a scale sample were compared between devices, the APE was 9.7%, and when the first and second readings with the Eberbach projector were compared, the APE was 8.7%. The ACVs for age estimates for the subsamples followed the same patterns as the APEs, given that the ACV is a fixed proportion larger than APE when comparisons are between 2 measurements.

Percent agreement was 88.9% for readings of scales from Atlantic menhaden overall and 89.7% for readings of scales from Gulf menhaden overall (Table 1). Percent agreement varied by year of sampling with the readings of

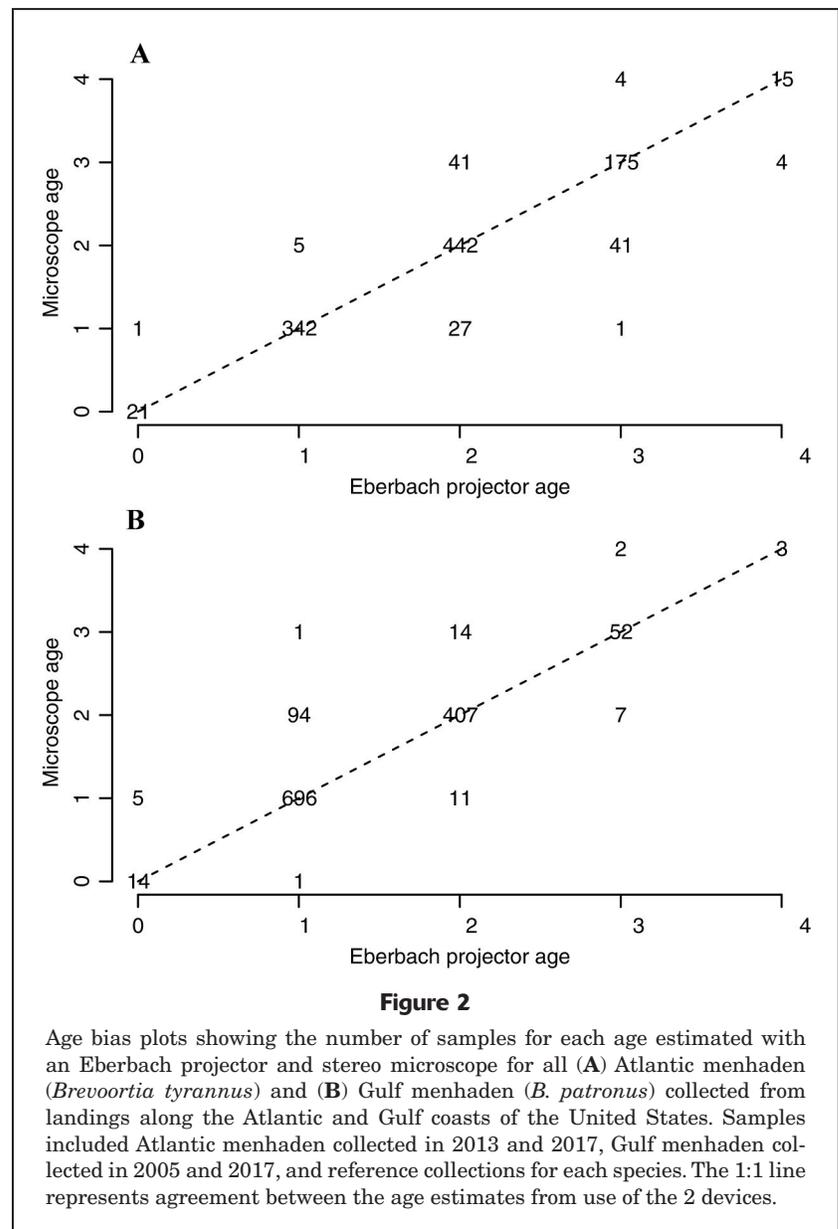
scales from the reference collections having the best agreement with 95.8% for Atlantic menhaden and 100.0% for Gulf menhaden. The readings for more recent samples, those taken in 2017, had better PA than the readings of the older samples from 2013 and 2005. Percent agreement in the age estimates for scale subsamples ranged from 83.6% to 97.1% (Table 2). For age data on subsamples of Atlantic menhaden, PA was consistently >88%. For age data on subsamples of the Gulf menhaden, PA was more variable, with comparisons of the second readings done with the Eberbach projector and microscope and comparisons of the first and second readings done with the microscope having PA above 90%.

For all scale samples examined in our study, the proportion of scales for each of the determined ages reflected the proportion of samples for each observed age in the long-term sampling program for the Atlantic and Gulf menhaden fisheries (Fig. 2; for age compositions used in the respective stock assessments, see SEDAR, 2018, 2020 and [Supplementary Figures 1–6](#)), and the age bias plots indicate that 95% confidence intervals for mean ages were small ([Suppl. Figs. 7–14](#)). Differences in aging were greatest at the older ages (>3 years) and at age 0; however, sample sizes for these ages were smaller than those for the ages of 1 year and 2 years. Patterns in the number of samples for each of the ages determined for subsamples of Atlantic and Gulf menhaden on the same device are similar to patterns for the full sample, with small numbers of deviations in age that are both above and below the 1:1 line in age bias plots, except for the ages of 3 and 4 years estimated with the Eberbach projector for Atlantic menhaden ([Suppl. Fig. 15](#)).

For samples from the reference collections and samples taken in 2017 for both Atlantic and Gulf menhaden, no bias in age estimates was found in comparisons between the 2 pieces of equipment—across all of the tests of symmetry (Table 3). For the samples of Gulf menhaden from 2005, however, the results of all 3 tests of symmetry indicate significant bias. A significant bias in age estimates for samples of Atlantic menhaden from 2013 is indicated only in results from the Bowker's test. For all samples of Atlantic menhaden in total, results of the Bowker's test indicate some bias, likely influenced heavily by the data set for samples from 2013, and the results of the other tests do not indicate significant bias in age data between the Eberbach projector and microscope. For

all samples of Gulf menhaden in total, results of all 3 tests of symmetry indicate significant bias in age estimates, influenced heavily by the data set for samples from 2005. For the subsamples, results from all tests for both species indicate significant bias in ages estimated with the Eberbach projector; results of these tests for both species indicate that bias was not significant for ages determined with the microscope. In comparisons of ages for the subsamples between estimates made with the Eberbach projector and those made with the microscope, results of some tests indicate significant bias, but results from other tests do not, outcomes that are likely influenced by the bias in ages determined with the Eberbach projector.

Overall simultaneous multinomial confidence intervals were slightly different for the ages of 1 year and 2 years



**Table 3**

Results from tests of symmetry in ages estimated by examining scales on an Eberbach projector and a stereo microscope for Atlantic menhaden (*Brevoortia tyrannus*) and Gulf menhaden (*B. patronus*) sampled along the Atlantic and Gulf coasts of the United States. Scales were from Atlantic menhaden collected in 2013 and 2017, Gulf menhaden collected in 2005 and 2017, and reference collections for each species. Reading 1 is the first reading of the sample set on both the projector and microscope, and reading 2 is the second reading on both instruments. Values are from comparisons of the 2 readings of scales in full samples and in subsamples on both devices and from comparisons of ages for subsamples determined on the same instrument. *P*-values for test results are provided in parentheses. *n*=sample size used for analysis. NA=not applicable.

Data set	<i>n</i>	Bowker's	Evans and Hoenig's	McNemar's
<b>Atlantic menhaden</b>				
All	1119	17.1 (<0.01)	4.6 (0.10)	3.9 (0.05)
2013	605	17.8 (<0.01)	2.8 (0.25)	2.0 (0.15)
2017	418	2.3 (0.51)	1.3 (0.25)	1.3 (0.25)
Reference	96	1.0 (0.32)	1.0 (0.32)	1.0 (0.32)
Subsample: reading 1	186	3.2 (0.68)	1.5 (0.48)	0.2 (0.65)
Subsample: reading 2	185	16.8 (<0.01)	16.2 (<0.01)	16.2 (<0.01)
Subsample: projector	176	10.3 (0.07)	9.9 (0.01)	7.2 (0.01)
Subsample: microscope	176	4.5 (0.60)	1.6 (0.44)	1.5 (0.22)
<b>Gulf menhaden</b>				
All	1307	73.7 (<0.01)	69.8 (<0.01)	69.7 (<0.01)
2005	759	74.4 (<0.01)	72.1 (<0.01)	72.1 (<0.01)
2017	499	4.4 (0.22)	3.3 (0.07)	3.3 (0.07)
Reference	49	0.0 (1.00)	0.0 (1.00)	NA (NA)
Subsample: reading 1	177	19.8 (0.01)	12.6 (< 0.01)	12.5 (<0.01)
Subsample: reading 2	173	5.0 (0.08)	5.0 (0.03)	5.0 (0.03)
Subsample: projector	157	11.5 (0.02)	6.0 (0.01)	6.0 (0.01)
Subsample: microscope	157	5.7 (0.13)	1.3 (0.25)	1.3 (0.25)

but not for the other ages for both the Atlantic and Gulf menhaden (Fig. 3). The mean ages for samples taken in 2017 and for samples from the reference collections were within the 95% confidence intervals for both species (Suppl. Figs. 16 and 17). However, the mean ages for older samples that were collected in 2005 and 2013 were significantly different.

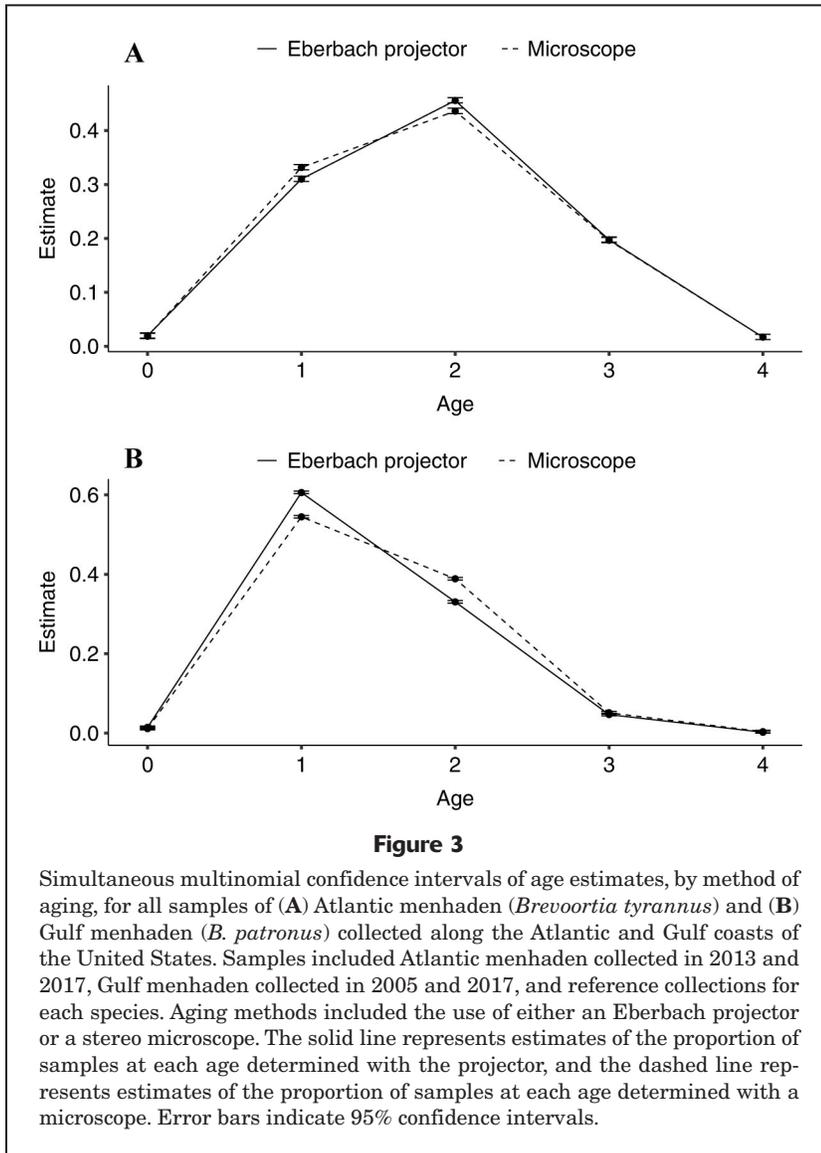
Mean annuli distances for Atlantic and Gulf menhaden increased as counts of annuli increased (Suppl. Table 1). Scale distances measured at each annulus with a blue card were longer than those measured with the Eberbach projector and those measured with the microscope (Suppl. Fig. 18). Values were significantly different between the methods used, for the measurements from the focus to the second annulus and from the focus to the third annulus for Atlantic menhaden and for the measurement from the focus to the second annulus for Gulf menhaden (Suppl. Table 1).

Readers 1 and 2 generally agreed on ages determined for samples from both reference collections (Suppl. Table 2), and APEs and ACVs were low for their age estimates. Average percent error was less than 5% for all comparisons, except for the comparison of ages for the samples

of Gulf menhaden taken in 2005, for which the APE was 12.0%. The PA for ages estimated for samples of Gulf menhaden taken in 2005, at 64.5%, was lower than the PAs for age data for the other samples, between 82.2% and 100.0%.

## Discussion

Quality assurance (QA) and quality control (QC) are required for integrity of data collected or calculated as part of long-term monitoring programs and for the provision of age estimates for fish sampled from commercial catches over a sequence of years (i.e., data from production aging) to be used in stock assessments. Quality assurance and QC are addressed differently in what is called *research aging*, which is geared toward a validation method and typically uses smaller sample sizes within a shorter time interval in comparison with production aging (Morison et al., 1998). Because it spans long time periods and large numbers of samples, production aging requires QC to ensure consistency among readings of each reader and between readers and methods for checking bias and minimizing errors to be



appropriate and applied in a timely manner. Quality assurance and QC help to determine whether perceived changes in age estimates are from variation in aging precision or from real changes in fish populations (Morison et al., 1998). Although this study served as QA and QC for the Beaufort Laboratory's menhaden program, its main purpose was to assess the level of continuity in the data if age readings were done on a microscope rather than with the Eberbach projector.

The menhaden program of the Beaufort Laboratory has been run without appreciable change for more than half a century, and well over a million samples have been amassed with a remarkable continuity in collection methods and output. Because of the shift in aging methods that was necessitated by critical program updates, we needed to determine whether outputs from the use of the new method would be consistent with those from the use of the previous method. Results from our study indicate that,

although there was some bias in the age estimates for Gulf menhaden and in estimates between readings made by using the Eberbach projector, the precision of age estimates made with the microscope, in comparison with those made with the Eberbach projector, is sufficient to allow the menhaden program to update the technology used for aging these 2 menhaden species. Continued aging for stock assessments should yield the same quality of data throughout the time series.

Sample sizes in our study were larger for each species than in many other studies that have considered aging precision. Experts have suggested a sample size of 200 fish, 100 from a reference collection and 100 from recent production samples, is sufficient to identify significant differences when testing for aging errors or inconsistencies (Campana, 2001). The large sample size in this study could have influenced the results of the bias tests and measures of aging consistency. McBride (2015) used a sample size of 5 fish to produce 5 age estimates per age class, a sample size that like those in most studies is much smaller than the sample sizes used in our study. For example, other studies have used sample sizes of 85–180 per species when comparing aging structures (Khan and Khan, 2009; Khan et al., 2015; Baudouin et al., 2016; Kumbar and Lad, 2016), 14–390 per species when comparing ages between readers (Wakefield et al., 2017; Khan et al., 2019), and 30–182 per species when comparing both aging methods and ages between readers (Gürsoy et al.,

2005; Goldman and Musick, 2006; Herbst and Marsden, 2011). Given these smaller sample sizes, comparing age estimates from our study to those from the other studies has been difficult.

How large should sample sizes be to determine if bias occurs? If we consider only data from the reference collections of each species, the conclusions would be different than those from consideration of data from the full sample. For example, the PA is much higher for estimates based on the reference collections alone than for those based on the full sample, and the APE and ACV are much lower. The ages from the reference collections do not indicate significant bias for either species. Because of the smaller sample sizes, use of the reference collection for each species led to different conclusions than those from the use of the full sample; however, the sample sizes of the reference collections are similar to the sample sizes that have been used in other aging studies.

Percent agreement values were all above 86% among age estimates for samples overall between the microscope and Eberbach projector and were all above 83% between first and second readings within an instrument. Our results are similar to those from McBride (2015), who reported PA of around 80% with no bias and low variability. Values of PA were high for reference collections of Atlantic and Gulf menhaden (95.8% and 100.0%, respectively). A workshop on aging Atlantic menhaden that used the same reference collection calculated PA values of 58.2–86.7% for comparisons of microfiche with an Eberbach projector between 10 state laboratories and of 91.8% for comparisons on the Eberbach projector between reader 2 and another reader (ASMFC<sup>6</sup>). These values are similar to the PA of 91.7% between reader 1 and reader 2 in this study. An exchange of 100 fishery-independent scales between reader 1 and an aging technician at the University of Southern Mississippi, aged on the same Eberbach projector, yielded a PA of 79.0% (Leaf<sup>7</sup>). The PA values from our study were also similar to or higher than those from studies for other species, with PA values of 39.8–100.0% when precision of estimates for multiple aging structures were compared (Khan and Khan, 2009; Khan et al., 2015; Kumbar and Lad, 2016) and of 68.1–92.4% when estimated precision was compared between readers (Goldman and Musick, 2006; Khan et al., 2019).

Overall, comparisons of age estimates made with the microscope and the Eberbach projector resulted in APEs under 4.4%, which is similar to or less than published APEs of about 8.0% (Campana, 2001; McBride, 2015), including the mean APEs for estimates from a simulation by McBride (2015) that indicated no bias and that decreased with increasing precision. Average percent errors from other studies have ranged from 0.7% to 16.1% when age estimates were compared among multiple aging structures or between readers (Gürsoy et al., 2005; Khan and Khan, 2009; Khan et al., 2015; Khan et al., 2019). Average percent errors in age readings were best for the reference collections, which had better quality, more legible scales than the scales taken from fish sampled for this study. An APE of 8.2% was calculated at a workshop during which aging was done for a reference collection of Gulf menhaden ( $n=30$ ) and during which microscope-based age estimates from a group of 6 state laboratories were compared with ages estimated with the Eberbach projector provided by reader 1 (VanderKooy<sup>8</sup>). Age estimates for scale subsamples from Gulf menhaden in our study had APEs of 9.7% when compared across devices and of 8.7% when compared among readings by a reader on the Eberbach projector, values that are higher than the benchmark identified in Campana (2001).

High APEs may indicate difficulty in maintaining consistency in the ages determined for Gulf menhaden with the Eberbach projector because of sample readability and quality. One third of the subsamples from Gulf menhaden were from 2005, a year in which scales had reduced quality due to improper storage. Although the use of both pieces of equipment yielded similar numbers of unreadable scales for reader 1, 55% of samples aged on the Eberbach projector, versus 35% of samples aged on the microscope, were noted to be fuzzy, dirty, or unreadable. Comparisons of subsamples of scales from Atlantic menhaden may have been affected by samples from the bait fishery (20% of the samples were from the bait fishery); samples from the bait fishery can be harder to read because of differences in fishing and sampling practices. Specifically, handling time on the vessel and in storage is greater for samples from the bait fishery than for samples from the reduction fishery, and fish are caught at different times of the year. Also, staff members that mount scales from fish sampled from the bait fishery typically have less experience.

The increase in APE values when comparing between the first and second readings for reader 1 could be due to greater experience. In 2017, when the first reading was done, reader 1 had been estimating ages of menhaden for 2 years, long enough to become proficient but still early enough that another year's experience would likely result in some level of improvement at the time of the second reading in 2018. However, although the APEs for the first readings were higher, they are on par with values of around 10% that have been observed for harder to read samples and for new readers in other studies (Potts<sup>9</sup>; McBride, 2015).

Values of ACV in this study were comparable to or less than values in simulation analyses (McBride, 2015), a workshop for aging Atlantic menhaden (ASMFC<sup>6</sup>), and studies on other species. The highest ACV value in this study was for the samples of Gulf menhaden taken in 2005 (6.2%), and this value is within the range of values reported in McBride (2015) on the basis of simulating age data with no bias across multiple precision levels but is higher than the value of 5% found to be accepted across data from several laboratories reviewed by Campana (2001). Analysis during a workshop on aging Atlantic menhaden resulted in ACVs that are generally above 5% among data from several laboratories but that do not exceed 15% (ASMFC<sup>6</sup>). The workshop participants computed an ACV of 2.1% (ASMFC<sup>6</sup>), a value that is similar to the ACV of 2.2% computed in this study for the same reference collection. Values of ACV for age data for both Atlantic and Gulf menhaden across all comparisons are similar to the values of 1.0–22.8% from other studies across aging structures and between readers (Khan and Khan, 2009; Herbst and Marsden, 2011; Khan et al., 2015; Khan et al., 2019). Values of ACV from this study indicate that the use of the microscope and the use of the Eberbach

<sup>6</sup> ASMFC (Atlantic States Marine Fisheries Commission). 2015. 2015 Atlantic menhaden ageing workshop report, 47 p. Atl. States Mar. Fish. Comm., Washington, D.C. [Available from [website](#).]

<sup>7</sup> Leaf, R. 2018. Personal commun. Univ. South. Miss. 703 East Beach Dr., Ocean Springs, MS 39564.

<sup>8</sup> VanderKooy, S. 2019. Personal comm. Gulf States Mar. Fish. Comm. 2404 Government St., Ocean Springs, MS 39564.

<sup>9</sup> Potts, J. 2014. Personal commun. Natl. Mar. Fish. Serv. 101 Pivers Isl. Rd., Beaufort, NC 28516.

projector are comparable in providing ages for Atlantic and Gulf menhaden.

Our study used 3 bias tests to determine consistency in age readings, which were not consistent across all tests for Atlantic and Gulf menhaden. Bias in age estimates in our study is higher for some years than for others and higher for older samples than for more recent samples. For Atlantic menhaden, the only significant bias test was the Bowker's test, which can be sensitive to one pair of age estimates with a large difference (McBride, 2015). Bias in the age estimates for scale subsamples indicates that use of the Eberbach projector results in less consistency in age readings, and the lack of bias in the readings of subsamples done with the microscope indicates that the use of the microscope lends well to consistency. During a workshop on aging Atlantic menhaden conducted in 2015, comparisons of ages between reader 2 and 10 laboratories yielded inter-laboratory and intra-laboratory *P*-values of 0–1 depending upon the test (ASMFC<sup>6</sup>).

Tests of symmetry for age data are often variable, with not all tests providing the same indication of bias or lack thereof. Results of some studies indicate no bias across tests (Goldman and Musick, 2006; Harry et al., 2011), but other studies have variable results pertaining to bias depending upon the test, reader, or age structure used (Harry et al., 2013; Baudouin et al., 2016; Stewart and Ogle, 2016). McBride (2015) used 100 samples across 20 ages with 5 samples per age and found that bias was subject to varying rates of type I and type II errors depending upon the underlying simulations. Given the sample size used by McBride (2015), the question becomes how sample size might affect those results. Campana et al. (1995) suggested that bias could be ignored when bias was low. Given the low numbers of age-4 fish and the expectation that the bias may have primarily crept in through ages estimated with the Eberbach projector, results of this study indicate that using the microscope for age estimation of menhaden species has a lower chance of introducing bias to the age estimates than using the Eberbach projector. Also, the main use of the age data from the long-term sampling program is to provide age compositions to stock assessments, and the simultaneous multinomial confidence intervals indicate that the composition data will not be affected by a switch to aging on the microscope.

Scale distance measurements changed a small degree as the menhaden program of the Beaufort Laboratory switched from blue cards to the sonic digitizer pen to the microscope; many of the differences were not significant. With each update of the method, measurements may have become more precise. For example, for the method that includes the use of blue cards, the cards are lined up with the scale's focus on a slanted screen below waist height, and annuli are ticked off with a pencil. Consistently placing the blue card ruler slightly below the focus that would have been selected with the sonic digitizer pen could lead to consistent, slightly larger measurements. Updating to the sonic digitizer pen decades ago allowed pen-point precision for measurement selection. Further updating of the method to the use of a microscope has

allowed a clearer and full view of the scale at eye level, use of a straight guideline for annulus selection, more precise selection of measurement points, and use of imaging software to view the measurements. The scale distance measurements indicate that similar structures are being measured to indicate true annuli. These data are not used in stock assessments. If these data are used for future research projects, the decreases in annulus measurements with changes in method should be taken into account.

In general, age estimates for Gulf menhaden had lower agreement, and the scale samples of this species were more difficult to read than those of Atlantic menhaden. Several factors, including scale size, consistency in sampler over time, and environmental conditions, could have contributed to these differences. First, individual Atlantic menhaden and their scales tend to be larger in size than Gulf menhaden and their scales. Larger sizes could make it easier for the port sampler to take samples from the proper location on the fish, where scales are more regular in shape and, therefore, more readable. Larger scales could result in more contrast and distance between perceived annuli, making them easier to read. Second, the same full-time employee has been the sampler of Atlantic menhaden since 1992. In the Gulf of Mexico, budgetary constraints have resulted in regular turnover of contract personnel that are selected, hired, and trained on an annual basis. Personnel differences would result in a consistently higher level of expertise in sampling at the port for the fishery in the Atlantic Ocean than that for the fishery in the Gulf of Mexico. Third, Atlantic menhaden caught in the fishery are exposed to environmental conditions that are more seasonally variable than those to which Gulf menhaden are exposed, given the differences in latitude included in their habitat range. Such differences in the habitats of other species have been reported to result in more contrast between rapid- and slow-growth periods (Hoxmeier et al., 2001; Green et al., 2009). The environmental differences could result in relatively higher-contrast scale circuli in Atlantic menhaden than the circuli of Gulf menhaden, therefore making them easier to age.

Overall, on the basis of the results of this study, we have determined that the use of scales on a microscope is an acceptable method for aging menhaden. Age estimates from use of the Eberbach projector were repeatable on the microscope, allowing continuity in the age data that have been maintained by the Beaufort Laboratory for decades. The microscope has many advantages. First, stage, light, and magnification options offer multiple ways to view samples more clearly while allowing for ergonomic considerations. Second, image analysis software used with a microscope allows more accurate straight-line measurements in units of the International System of Units. Third, because most laboratories have microscopes, samples can be viewed in a more uniform manner across facilities, making collaborations and comparisons easier. Fourth, data are more accessible and analysis of data is quicker because the data are no

longer stored in an outdated format. Finally, notes can be added to data entries in the database, along with sample information, and digital images of scales can be captured for easy reference and for training purposes within and across laboratories.

Moving forward, reference collections will be a primary QA and QC tool for training and maintaining aging standards (Eltink et al.<sup>10</sup>; Campana, 2001; Morison et al., 2005). New digital reference collections for Atlantic and Gulf menhaden will be composed of average quality scales with an annotated version to be shared with other aging laboratories for training purposes. A digital collection has longevity and ensures that all aging scales are prepared by the same method (Campana, 2001). Samples for the digital reference collection should represent multiple years and have 10 samples, with varying edge margins, from each of the age classes for each month of landings sampled (Eltink et al.<sup>10</sup>). Campana (2001) suggests that a reference collection composed of 200–500 samples be used for quality control, although a training set of 100 samples is sufficient. For exchanges, a reference collection of 400 samples has been suggested, with possibly 150–200 subsamples sent at a time (Eltink et al.<sup>10</sup>; Campana, 2001). A subsample of the samples used in this study could be assembled as a control collection because they are of unknown ages and are true representatives of the catch. A subsample of this collection should be aged again yearly by the current reader to reassess consistency, precision, and bias of age estimates and could be used to train another reader (Eltink et al.<sup>10</sup>; Campana, 2001). Collections should be updated at regular intervals.

Finally, the standards for determining continuity in age determination and for comparing between methods need further investigation. Sample sizes used in many studies are smaller than those in our study, and work to identify the sample size that is sufficient to answer questions of such examinations is greatly needed. McBride (2015), through simulation, tested PA, APE, ACV, and bias of age estimates by using different types of bias and sample sizes of 5 samples per age class with 20 ages. The behavior of these tests with differences in sample sizes across ages, meaning a more natural sample size distribution given common sampling protocols, would help to elucidate whether ages are consistent across methods.

## Acknowledgments

We thank J. Potts, G. Nessler, and the anonymous reviewers for helpful comments and suggestions on earlier drafts. This project was funded through the Improve a Stock Assessment Program of the Office of Science and Technology, National Marine Fisheries Service.

<sup>10</sup> Eltink, A. T. G. W., A. W. Newton, C. Morgado, M. T. G. Santamaria, and J. Modin. 2000. Guidelines and tools for age reading (PDF document version 1.0 October 2000). Eur. Fish Ageing Netw., EFAN Rep. 3-2000, 54 p. Eur. Fish Ageing Netw., Arendal, Norway.

## Literature cited

- Baudouin, M., M. Marengo, A. Pere, J.-M. Culioli, M.-C. Santoni, B. Marchand, and E. D. H. Durieux.  
2016. Comparison of otolith and scale readings for age and growth estimation of common dentex *Dentex dentex*. *J. Fish Biol.* 88:760–766. [Crossref](#)
- Beamish, R. J., and D. A. Fournier.  
1981. A method for comparing the precision of a set of age determinations. *Can. J. Fish. Aquat. Sci.* 38:982–983. [Crossref](#)
- Beamish, R. J., and G. A. McFarlane.  
1995. A discussion of the importance of aging errors, and an application to walleye pollock: the world's largest fishery. *In* Recent developments in fish otolith research (D. H. Secor, J. M. Dean, and S. E. Campana, eds.), p. 545–565. Univ. South Carolina Press, Columbia, SC.
- Bowker, A. H.  
1948. A test for symmetry in contingency tables. *J. Am. Stat. Assoc.* 43:572–574. [Crossref](#)
- Campana, S. E.  
2001. Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. *J. Fish Biol.* 59:197–242. [Crossref](#)
- Campana, S. E., and S. R. Thorrold.  
2001. Otoliths, increments, and elements: keys to a comprehensive understanding of fish populations? *Can. J. Fish. Aquat. Sci.* 58:30–38. [Crossref](#)
- Campana, S. E., M. C. Annand, and J. I. McMillan.  
1995. Graphical and statistical methods for determining the consistency of age determinations. *Trans. Am. Fish. Soc.* 124:131–138.
- Chang, W. Y. B.  
1982. A statistical method for evaluating the reproducibility of age determination. *Can. J. Fish. Aquat. Sci.* 39:1208–1210. [Crossref](#)
- Evans, G. T., and J. M. Hoenig.  
1998. Testing and viewing symmetry in contingency tables, with application to readers of fish ages. *Biometrics* 54:620–629. [Crossref](#)
- Goldman, K. J., and J. A. Musick.  
2006. Growth and maturity of salmon sharks (*Lamna ditropis*) in the eastern and western North Pacific, and comments on back-calculation methods. *Fish. Bull.* 104:278–292.
- Green, B. S., B. D. Mapstone, G. Carlos, and G. A. Begg (eds.).  
2009. Tropical fish otoliths: information for assessment, management and ecology, 313 p. Springer, Dordrecht, Netherlands.
- Gürsoy, C., A. S. Tarkan, Ö. Gaygusuz, and H. Acipinar.  
2005. A comparison of ageing techniques to improve precision of age estimation from fish scales. *Ege J. Fish. Aquat. Sci.* 22:423–425.
- Harry, A. V., W. G. Macbeth, A. N. Gutteridge, and C. A. Simpfendorfer.  
2011. The life histories of endangered hammerhead sharks (Carcharhiniformes, Sphyrnidae) from the east coast of Australia. *J. Fish Biol.* 78:2026–2051. [Crossref](#)
- Harry, A. V., A. J. Tobin, and C. A. Simpfendorfer.  
2013. Age, growth and reproductive biology of the spot-tail shark, *Carcharhinus sorrah*, and the Australian blacktip shark, *C. tilstoni*, from the Great Barrier Reef World Heritage Area, north-eastern Australia. *Mar. Freshw. Res.* 64:277–293. [Crossref](#)
- Herbst, S. J., and J. E. Marsden.  
2011. Comparison of precision and bias of scale, fin ray, and otolith age estimates for lake whitefish (*Coregonus clupeaformis*) in Lake Champlain. *J. Gt. Lakes Res.* 37:386–389. [Crossref](#)

- Hettmansperger, T. P., and J. W. Mckean.  
1973. On testing for significant change in cxc tables. *Commun. Stat.* 2:551–560. [Crossref](#)
- Hoenig, J. M., M. J. Morgan, and C. A. Brown.  
1995. Analysing differences between two age determination methods by tests of symmetry. *Can. J. Fish. Aquat. Sci.* 52:364–368. [Crossref](#)
- Hoxmeier, R. J. H., D. D. Aday, and D. H. Wahl.  
2001. Factors influencing precision of age estimation from scales and otoliths of bluegills in Illinois reservoirs. *North Am. J. Fish. Manage.* 21:374–380. [Crossref](#)
- June, F. C., and C. M. Roithmayr.  
1960. Determining age of Atlantic menhaden from their scales. *Fish. Bull.* 171:323–342.
- Khan, M. A., and S. Khan.  
2009. Comparison of age estimates from scale, opercular bone, otolith, vertebrae and dorsal fin ray in *Labeo rohita* (Hamilton), *Catla catla* (Hamilton), and *Channa marulius* (Hamilton). *Fish. Res.* 100:255–259. [Crossref](#)
- Khan, S., M. A. Khan, K. Miyan, and F. A. Lone.  
2015. Precision of age estimates from different ageing structures in selected freshwater teleosts. *J. Environ. Biol.* 36:507–512.
- Khan, M. A., S. Khan, Ankita, I. Ahmed, and M. Nadeem.  
2019. Age and growth of spotted snakehead, *Channa punctata* from the Ganga River. *J. Ichthyol.* 59:197–204. [Crossref](#)
- Kumbar, S. M., and S. B. Lad.  
2016. Estimation of age and longevity of freshwater fish *Salmophasia balookee* from otoliths, scales and vertebrae. *J. Environ. Biol.* 37:943–947.
- McBride, R. S.  
2015. Diagnosis of paired age agreement: a simulation of accuracy and precision effects. *ICES J. Mar. Sci.* 72:2149–2167. [Crossref](#)
- McNemar, Q.  
1947. Note on sampling error of the difference between correlated proportions or percentages. *Psychometrika* 12:153–157. [Crossref](#)
- Morison, A. K., S. G. Robertson, and D. C. Smith.  
1998. An integrated system for production fish ageing: image analysis and quality assurance. *North Am. J. Fish. Manage.* 18:587–598. [Crossref](#)
- Morison, A. K., J. Burnett, W. J. McCurdy, and E. Moksness.  
2005. Quality issues in the use of otoliths for fish age estimation. *Mar. Freshw. Res.* 56:773–782. [Crossref](#)
- Nicholson, W. R., and W. E. Schaaf.  
1978. Aging of Gulf menhaden, *Brevoortia patronus*. *Fish. Bull.* 76:315–322.
- R Core Team.  
2019. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Available from [website](#), accessed May 2019.]
- Secor, D. H., T. M. Trice, and H. T. Hornick.  
1995. Validation of otolith-based ageing and a comparison of otolith and scale-based ageing in mark-recaptured Chesapeake Bay striped bass, *Morone saxatilis*. *Fish. Bull.* 93:186–190.
- SEDAR (Southeast Data Assessment and Review).  
2015. SEDAR 40—Atlantic menhaden stock assessment report, 643 p. SEDAR, North Charleston, SC. [Available from [website](#).]  
2018. SEDAR 63—Gulf menhaden stock assessment report, 352 p. SEDAR, North Charleston, SC. [Available from [website](#).]  
2020. SEDAR 69—Atlantic menhaden benchmark stock assessment report, 691 p. SEDAR, North Charleston, SC. [Available from [website](#).]
- Sison, C. P., and J. Glaz.  
1995. Simultaneous confidence intervals and sample size determination for multinomial proportions. *J. Am. Stat. Assoc.* 90:366–369. [Crossref](#)
- Stewart, T. R., and D. H. Ogle.  
2016. Age, growth, and size of Lake Superior pygmy whitefish (*Prosopium coulterii*). *Am. Midl. Nat.* 175:24–36. [Crossref](#)
- Vilizzi, L.  
2018. Age determination in common carp *Cyprinus carpio*: history, relative utility of ageing structures, precision and accuracy. *Rev. Fish Biol. Fish.* 28:461–484. [Crossref](#)
- Villacorta, P. J.  
2019. MultinomialCI: simultaneous confidence intervals for multinomial proportions according to the method by Sison and Glaz. R package, vers. 1.1. [Available from [website](#), accessed January 2020.]
- Wakefield, C. B., J. M. O'Malley, A. J. Williams, B. M. Taylor, R. S. Nichols, T. Halafihi, R. L. Humphreys Jr., J. Kaltavara, S. J. Nichol, and S. J. Newman.  
2017. Ageing bias and precision for deep-water snappers: evaluating nascent otolith preparation methods using novel multivariate comparisons among readers and growth parameter estimates. *ICES J. Mar. Sci.* 74:193–203. [Crossref](#)
- Zar, J. H.  
1999. *Biostatistical analysis*, 4th ed., 663 p. Prentice Hall, Upper Saddle River, NJ.