

COMPUTING AND SELECTING AGEING ERRORS TO INCLUDE IN STOCK ASSESSMENT MODELS OF PACIFIC SARDINE (*SARDINOPS SAGAX*)

EMMANIS DORVAL*,
JENNIFER D. McDANIEL
Fisheries Resources Division
Southwest Fisheries Science Center
National Marine Fisheries Service, NOAA
8901 La Jolla Shores Drive
La Jolla, CA 92037-1508 USA
*Corresponding author:
ph: (858) 546-7079, fax: (858) 546-7003
Emmanis.Dorval@noaa.gov

DIANNA L. PORZIO
California Department of Fish and Wildlife
4665 Lampson Ave, Suite C
Los Alamitos, CA 90720

ROBERTO FÉLIX-URAGA
Instituto Politécnico Nacional-CICIMAR
Playa Palo de Santa Rita, Apdo. Postal 592
23096 La Paz, B.C.S. Mexico

VANESSA HODES
Pacific Biological Station
Fisheries and Oceans Canada
Nainamo, BC Canada V9T 6N7

SANDRA ROSENFELD
Washington Department of Fish and Wildlife
600 Capitol Way North
Olympia, WA 90501

ABSTRACT

From 2007 to 2010, Pacific sardine stock assessments relied on traditional methods to compute and include ageing errors in the integrated assessment model, Stock Synthesis (SS). Traditional methods assumed that all age readers were unbiased and estimated ageing imprecisions by averaging across all fish that were assigned a given age a by one or more readers. In this study, we used the Ageing Error Matrix (Agemat) model to compute ageing imprecisions, based on classification matrices that quantified the probability of a fish of true age a to be assigned an age a or some other age a' , $P(a' | a)$. Using sardine samples collected from Mexico to Canada and aged in five laboratories, we compared three Agemat models, assuming that: (1) the most experienced reader from each laboratory was unbiased (model A); (2) no bias but different standard deviation (SD) at age among readers (model B); and (3) no bias, but similar standard deviation at age among readers (model C). We evaluated the performance of this model using the Akaike information criterion corrected for finite sample sizes. Sardine ages ranged from 0 to 8, with increasing reader SD with age. Model C performed better than models A and B, across all data sets and laboratories, and thus was recommended for including ageing imprecisions in sardine assessment models. However, the observed differences in SD across ages and readers called for a better standardization of ageing protocols among laboratories and for applying new methods to reduce potential bias in estimating the oldest age classes.

INTRODUCTION

Ageing errors can influence the process of estimating demographic parameters for fish populations, the performance of assessment models, and ultimately management measures derived from these analyses. Estimation of biological parameters, such as maturity-at-age, length-at-age, and weight-at-age may be biased in the presence of these errors. Fishing data such as catch-at-age and catch-per-unit effort indices may be affected as well. Further, in assessment models, these errors may smooth out estimates of recruitment and total allowable catch allocated

to fisheries (Reeves 2003). Consequently, ageing errors can significantly mask important stock-recruit relationships and potentially the effects of environmental factors on year-class strength (Fournier and Archibald 1982; Richards et al. 1992).

New statistical models that can take account of both bias and precision in estimating ageing errors have been developed in recent years (e.g., Richards et al. 1992; Punt et al. 2008). These newer models can estimate the true age distribution of a population based on multiple age-readings of individual fish, particularly when fish were aged from a validated ageing method. In these models, age-reading errors are represented using classification matrices that quantify the probability of a fish of true age a to be assigned an age a or some other age a' , $P(a' | a)$. Parameters of various functions can be estimated from these models to determine the relationship between true age and estimated age. Because these statistical models are based on the maximum likelihood method, they allow for considerable flexibility in the relationship between true age and the expectation and imprecision of the estimated age (Richards et al. 1992; Punt et al. 2008). However, because all assessment models are not uniformly parameterized or structured, it can be difficult to select and include ageing errors in some models. For example, integrated assessment programs have been used to include ageing errors per reader in modeling fish stocks harvested off southern Australia (Punt et al. 2008), whereas Stock Synthesis (SS, Methot 2000) does not support data from multiple readers. In SS only one vector of ageing error can be input in the model. Thus, it is important to develop approaches to compute and include errors in SS models when multiple readers participate in the age production. To address this issue, we will use as a model species, the Pacific sardine (*Sardinops sagax*), a trans-boundary fish that is exploited from Mexico to Canada and whose age determination for stock assessment required the involvement of several readers within and across laboratories.

Since the 1990s Pacific sardine stocks have been assessed using age-structured models (Deriso et al. 1996;

Conser et al. 2004; Hill et al. 2007, 2009). Although many of these models could include age-reading errors, a systematic estimation of these errors has never been conducted for sardine samples collected from Mexico to Canada. Butler et al. 1996 used traditional methods (i.e., Beamish and Fournier 1981; Chang 1982) to assess age-reading imprecisions for fish collected during the 1994 Daily Egg Production Method (DEPM) survey, however these estimates could not be applied to age-data time series used in past assessment models. Hill et al. 2007, 2009 also used traditional methods to compute the mean standard deviation-at-age (*SDa*) for all readers that participated in a 2004 tri-national sardine ageing workshop (i.e., involving age readers from Mexico, the USA, and Canada). These estimates were included in the assessment models, although they represented a snapshot in time and did not account for differences in age estimation between fisheries or laboratories. Traditional methods generally focused on computing either precision (i.e., Beamish and Fournier 1981; Chang 1982) or bias (Campana et al. 1995; Morison et al. 1998), but not both. Therefore, these methods are not appropriate to develop age-reading error matrices for use in stock assessment models (Punt et al. 2008).

The otolith is the preferred hard part used to age Pacific sardines collected along the North American Pacific coast. A methodology for determining age of Pacific sardines from whole, un-sectioned otoliths was established by Yaremko 1996 and is currently used in ageing laboratories in Mexico, the USA and Canada. The Yaremko 1996 method summarized and integrated techniques developed during the historical sardine fishery in the 1940s and 1950s (Phillips 1948; Walford and Mosher 1950; Mosher and Eckles 1954), and during the recovery of the stock in the 1980s and 1990s (Butler 1987; Barnes and Forman 1994). McFarlane et al. 2010 proposed a preliminary method to age fish older than 1 year collected off British Columbia (BC), which consisted of polishing otoliths on their distal and proximal (sulcus) sides. Comparing their method to the otolith surface ageing of Yaremko 1996, McFarlane et al. 2010 found that the polished otolith method could improve the identification of the first and the second annuli. In addition, fish aged from the polished otolith method were found to be 1 to 3 years older than when aged from whole otoliths.

However, the polished otolith method is not currently being used for age production, because the approach needs further evaluation for sardine collected throughout their range. Although McFarlane et al. 2010 assessed potential bias in older fish relative to the Yaremko 1996 method, these comparisons were not based on known age fish or validated ages. Therefore, there were considerable uncertainties in the age differences detected by

McFarlane et al. 2010 between the two approaches. Further, older clupeids are often the most difficult to age using a standard-zone counting method (Rogers and Ward 2007), as their otolith thickness increases and the marginal increments are narrower compared to younger fish. As most fish collected in BC are older than 2 years old, it is not known whether the polished method would produce significantly different ages from the whole otolith method if applied to younger fish caught off California and Mexico. Finally, because the assessment of the northern sardine stock relied on the ageing contribution from Mexico, Canada, and the USA, and given the preliminary nature of McFarlane et al. 2010 study, the Yaremko 1996 otolith ageing method remains the primary means used among all ageing laboratories.

The primary objective of this study was to estimate and compare age-reading errors for Pacific sardines that were collected from Mexico to Canada and that were aged in five different ageing laboratories: (1) The Centro Interdisciplinario de Ciencias Marinas (CICIMAR, Baja California Sur, Mexico); (2) the California Department of Fish and Wildlife (CDFW, CA, U.S.); (3) the Southwest Fisheries Science Center (SWFSC, CA, U.S.); (4) the Washington Department of Fish and Wildlife (WDFW, WA, U.S.); and (5) the Pacific Biological Station (PBS) of the Department of Fisheries and Ocean (DFO, BC, Canada). A second objective of this study was to develop a quantitative approach to select and include ageing error matrices in the sardine stock assessment models that use the SS framework (Hill et al. 2011). These objectives were based on recommendations from the Pacific sardine 2009 Stock Assessment Review Panel. This review panel required more systematic age-reading comparisons among Pacific sardine ageing laboratories and better estimation and integration of age-reading errors in future assessment models.

MATERIALS AND METHODS

Sample Collection

Pacific sardines were collected from the DEPM survey and from port sampling of commercial fishery landings from Mexico to Canada. DEPM samples were collected during the 2004–11 April surveys from San Diego to San Francisco (CA). Port sampling data were collected using various designs (Hill et al. 2009) but were assumed to be representative of four major fisheries: the Mexican fishery, from Ensenada to Magdalena Bay (Mex, Mexico); the California fishery (CA, including the southern and central California fisheries); the Pacific Northwest fishery (PNW, including Oregon, Washington), and the British Columbia (BC, Canada) fishery. The DEPM survey and the port samplings are respectively detailed in Lo et al. 2005 and Hill et al. 2009.

Otolith age-reading

Pacific sardines were aged from unpolished whole otolith pairs by readers located at five ageing laboratories: CICIMAR, CDFW, SWFSC, WDFW, and PBS. All laboratories used the conventional technique of otolith age-reading described in Yaremko 1996, although with slight variations. The method is straightforward and generally recommends that: (1) the age reader immerses the otolith in distilled water for about three minutes; and (2) the age reader counts the number of annuli observed on the distal side of the otolith using a light microscope. An annulus is defined as the interface between an inner translucent growth increment and the successive outer opaque growth increment (Fitch 1951; Yaremko 1996). The method assumed a July 1 birth date for all individual fish hatched in U.S. waters within a calendar year. Although the spawning peak is typically in April, Pacific sardines continue to spawn throughout the summer months. The July 1 birth date assumption coincides with the recruitment of age-0 fish into the USA fishery, and thus this birth date accounts for fish born off California in the spring and off Oregon and Washington in summer. Age assignment by readers is based on the capture date and the interpretation of the most distal pair of growth increments:

1. Fish caught in the first semester of a calendar year have not yet reached their July 1 birth date; therefore their most distal pair of opaque and translucent increments should not be counted, even if exhibiting the early beginning of a second opaque increment (Yaremko 1996).
2. Fish caught in the second semester of a calendar year have completed a year since their last birth date; therefore their age is equal to the number of annuli counted in their otolith.
3. The marginal increment is categorized as opaque or translucent, wide or narrow, allowing a confidence rating to be assigned to the age determination.

Since the 1990s, all five ageing laboratories have used this method for Pacific sardine collected along the Pacific coast. The CDFW and SWFSC laboratories strictly followed the procedures outlined by Yaremko 1996, thus all otoliths were submerged in distilled water and observed using reflected light on a black background at the 24X resolution. However, the WDFW ageing protocol slightly modified the method, by immersing the otoliths in full-strength alcohol. Fish caught in the Pacific Northwest are generally older than those caught in California, and it appears that alcohol may enhance the readability of their otoliths under the 24X resolution. At the CICIMAR ageing laboratory, the method was modified by fixing whole otoliths on glass slides (sulcus side down) using the synthetic resin Cytoseal, and by counting the inner

annual increments at 16X resolution and the marginal increments at the 40X resolution. Note that the recommended 24X resolution was not selected because the light microscopes that are currently used at CICIMAR provided only 2 options, the 16X and the 40X resolutions. Further, at the CICIMAR laboratory, no birth date was assumed because of a long protracted spawning season (i.e., November to April) followed by a minor season that peaks in July off the coast of Bahia Magdalena. Thus the ages reported from fish collected in Magdalena Bay were similar to the number of annuli counted in otoliths. Finally, the PBS ageing laboratory followed the method of Yaremko 1996, but assumed a January birth date for all fish collected off British Columbia. Otoliths were submerged in distilled water and observed with a dissecting scope up to a maximum magnification of 50X, using reflected light against a dark background. Magnification and brightness were adjusted as necessary to clarify annuli and determine the presence of check marks. As for the CDFW, SWFSC, and WDFW ageing laboratories, ages reported by the PBS readers accounted for both capture and birth dates.

Age-reading data

Age-reading data from each fishery and survey were organized in data sets, which were defined as sets of otoliths that were aged by the same group of readers. A total of 10 data sets were used in this study, with total sample sizes (N) varying from 145 to 959 otoliths (table 1). In each data set, fish that were assigned a similar age combination among readers were grouped, and the absolute frequency of observed samples (n) for each unique age combination was computed. Each one of these unique age combinations among readers was input in the model as one observation associated with its absolute frequency (n). Thirteen age readers participated in this study, and each reader was provided with a unique identification number (ID). The number of readers per data set is presented in Table 1. All readers participating in this study were certified by their respective laboratory, but with varying degree of experience.

For the Mexican fishery, sardines were collected in Magdalena Bay during the 2005 fishing season. All fish collected in this bay were aged by a single reader (reader 13) at the CICIMAR ageing laboratory. Otoliths were first read in December 2006 and then re-read in June 2011.

In the California fishery, sardine samples were collected from port landings in southern California (San Pedro to Santa Barbara) and central California (Monterey Bay region) from 2005 to 2011. Depending on the year of collection, three to five CDFW readers participated in the age estimation process (table 1). Data sets were built based on time of collection (one to two years)

TABLE 1

Number of Pacific sardine otoliths (N) and age-readings (R) per otolith, data set, laboratory, and for each fishery or survey.

Ageing Laboratory	Fishery/Survey	Data set	Collection Year	Reader ID	R	N
CICIMAR	Mex	1	2005	13	2	240
		2	2005	1,2,3	3	219
		3	2007	2,4,5,6	4	148
CDFW	CA	4	2008–2009	2,4,5,6,7	5	507
		5	2008–2009	2,5,6,7	4	145
		6	2010–2011	2,5,6	3	266
		7	2009	8,9	2	711
WDFG	PNW	8	2007	10,11	3	283
PBS	BC	9	2004, 2006	1,12	2	360
CDFW-SWFSC	DEPM	10	2005–2011	2,12	2	959

using only completely reported age-reading combinations (i.e., observations containing one or more missing values were discarded). For the 2008–09 period, two different data sets (4 and 5) were used in the modeling process, to account for turnover among readers.

In the PNW, sardine samples were collected off Oregon during the summer of 2009. All otoliths were aged by readers 8 and 9 from the WDFW (table 1) in March and April of 2010. In model A, reader 8 was most experienced and was assumed to be unbiased.

British Columbia fishery samples were collected from July to September of 2007. Whole otoliths from BC were first aged separately by readers 10 and 11 (table 1). Then, each otolith was re-aged simultaneously by both readers to estimate a best, resolved age (RA). Taking into account this ageing protocol, we assumed that the resolved age was more likely to be unbiased.

Sardine samples from DEPM cruises were collected during the spring season from 2004 to 2011. Age determinations were done by readers 1 and 2 from CDFW and reader 12 from the SWFSC. There were no overlapping age-readings between readers 1 and 2; therefore, two different data sets were built for the DEPM survey (table 1).

Ageing error estimation

We used the Ageing Error Matrix (Agemat) model developed by Punt et al. 2008 to estimate age-reading error matrices by reader. The model computed ageing-error matrices based on otoliths that had been aged multiple times by one or more readers, while assuming that: (1) ageing bias depends on reader and the true age of a fish; (2) the age-reading error standard deviation depends on reader and true age; and (3) age-reading error is normally distributed around the expected age. Hence, the probability to assign an age a' to a fish of true age a was computed following Equation 1:

$$P^i(a' | a, \phi) = \int_a^{a'+1} \frac{1}{\sqrt{2\pi\sigma_i^2(\phi)}} \exp \left[\frac{-(a'-b_i^i(\phi))^2}{2(\sigma_i^i(\phi))^2} \right] da', \quad (1)$$

where b_i^i is the expected age when reader i determines the age of a fish of true age a , σ_i^i is the standard deviation for reader i and for fish whose true age is a , and ϕ is the vector of parameters that determines the age-reading error matrices. The values for these parameters were estimated by maximizing the following likelihood function, assuming there was some set of J otoliths that was read by all readers:

$$L(A | \beta, \phi) = \prod_{j=1}^J \sum_{a=L}^H \beta_a = \prod_{i=1}^I P^i(a_{i,j} | a, \phi) \quad (2)$$

where $a_{i,j}$ is the age assigned by reader i to the j^{th} otolith; L and H are respectively the minimum and the maximum ages, and A is the entire data set of age-readings. The β s are nuisance parameters that can be interpreted as the relative frequency of fish of true age a in the sample.

For the purpose of this study we were mostly interested in estimating the SDs for the different fisheries and the DEPM survey. We defined three model scenarios, based on different assumptions about SDs among readers: (1) model A assumed that age-readings from the most experienced reader for one or two data sets were unbiased, and accordingly SDs for all readers were estimated based on this assumption; (2) model B assumed no bias in estimated ages for all readers, but SDs were different among readers; (3) model C assumed no bias in estimated ages for all readers, but the readers had similar SDs. Firstly, we compared the fit of each model to the age-reading data graphically, by plotting the predicted frequency (n) for a given age combination among readers versus the observed frequency for each model by data set. Secondly, we used the Akaike information criterion with a correction for finite sample sizes (AICc) to compare the fits among the three models and ultimately to select the best model, determining whether there was enough evidence to support the assumption of equality of SDs among readers for the data sets considered in a given model. Note that model C was the only model that could be used in the current SS version, i.e., 3. The validity of using this model was tested against models A

and B that are commonly used to introduce errors in assessment models.

We assumed that the functional form of random ageing error precisions followed either Eq. 3 or 4.

$$\sigma_a = \sigma_L + (\sigma_H - \sigma_L) \frac{1 - \exp(-\delta(a-1))}{1 - \exp(-\delta(a_{max}-1))} \quad (3)$$

where, σ_L and σ_H are respectively the standard deviation of the minimum and the maximum age in a given data set, and δ is a parameter that determines the extent of linearity between age and the age-reading standard deviation.

$$CV_a = CV_L + (CV_H - CV_L) \frac{1 - \exp(-\delta(a-1))}{1 - \exp(-\delta(a_{max}-1))} \quad (4)$$

where CV_L and CV_H are respectively the coefficient of variation of the minimum and the maximum age in a given data set (Richards et al. 1992; Punt et al. 2008).

For the DEPM survey, the PNW and BC fisheries all three models were used. In these cases, the most experienced readers (DEPM and PNW) and the resolved ages (BC) were assumed to be unbiased in model A, whereas the functional form for ageing bias for all other readers was assumed to follow Equation 5:

$$E_a = E_L + (E_H - E_L) \frac{1 - \exp(-\beta(a-1))}{1 - \exp(-\beta(a_{max}-1))} \quad (5)$$

where E_a is the expected age of a fish of age a ; E_L and E_H are respectively the minimum and the maximum ages in a given data set; a_{max} is a pre-specified maximum age; and β is a parameter that determines the extent of linearity between age and the expected age.

For all model runs, the maximum expected age for sardines was set to be 15, based on current maximum age used in assessment models (Hill et al. 2011). Based on length frequencies, scale, and otolith ageing analyses, previous authors have reported that Pacific sardine can live longer than 12 years, thus by convention age 15 was established to be the maximum of the probable 11–15 age group (e.g., Marr 1960; Wolf 1961; Murphy 1967). Further, the maximum SD allowed in model runs was 100. Very high SD s such as 100 would result from insufficient observed samples combined with inexperience in ageing older fish. This occurred once in model B for readers 2 and 7 and for the 2008–09 CA samples and in model A for reader 9 for the 2009 PNW samples. In these cases, SD was not plotted for these readers. Finally, to better describe the variability of the age-reading data, we conducted pairwise comparisons of age combinations among readers for each data set, by establishing 4 levels of agreement in age assignment between pair of readers: (1) low, <65%; moderate, 65% to 74%; high 75% to 84%; and very high, $\geq 85\%$.

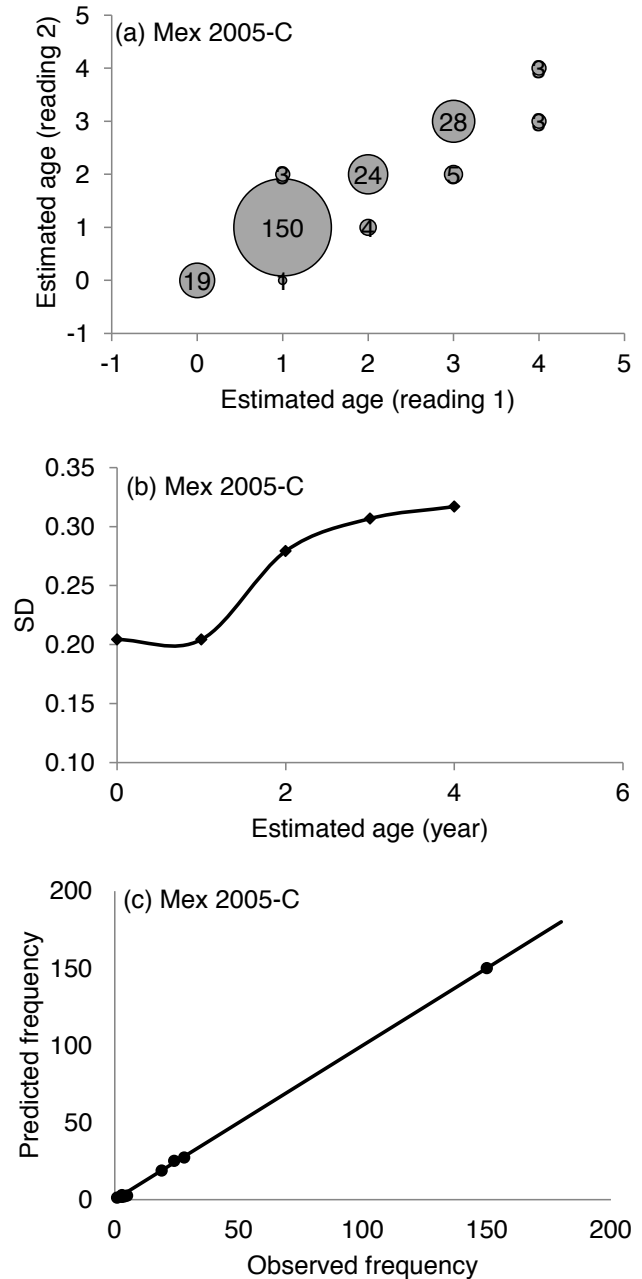


Figure 1. Age and standard deviation estimated for fish collected in Magdalena Bay, Mexico in 2005: (a) agreement plot for pairwise comparison of ages estimated by 1 CICIMAR reader from 2 readings. Number and bubble size represent the frequency of observed age combinations; (b) Standard deviation (SD) estimated from model C; and (c) diagnostic plot for the fit of model C to the data.

RESULTS

Because both age readings for the Mexican fishery were performed by one reader, no ageing bias was assumed. Agreement between the two readings was very high from age 0 to age age 3 ($\geq 85\%$), but low for age 4 (50%) (fig. 1a). As expected, SD increased with age, varying from 0.20 at age 0 to 0.30 at age 4 (fig. 1b). Compared to other laboratories, SD s estimated in Mexico

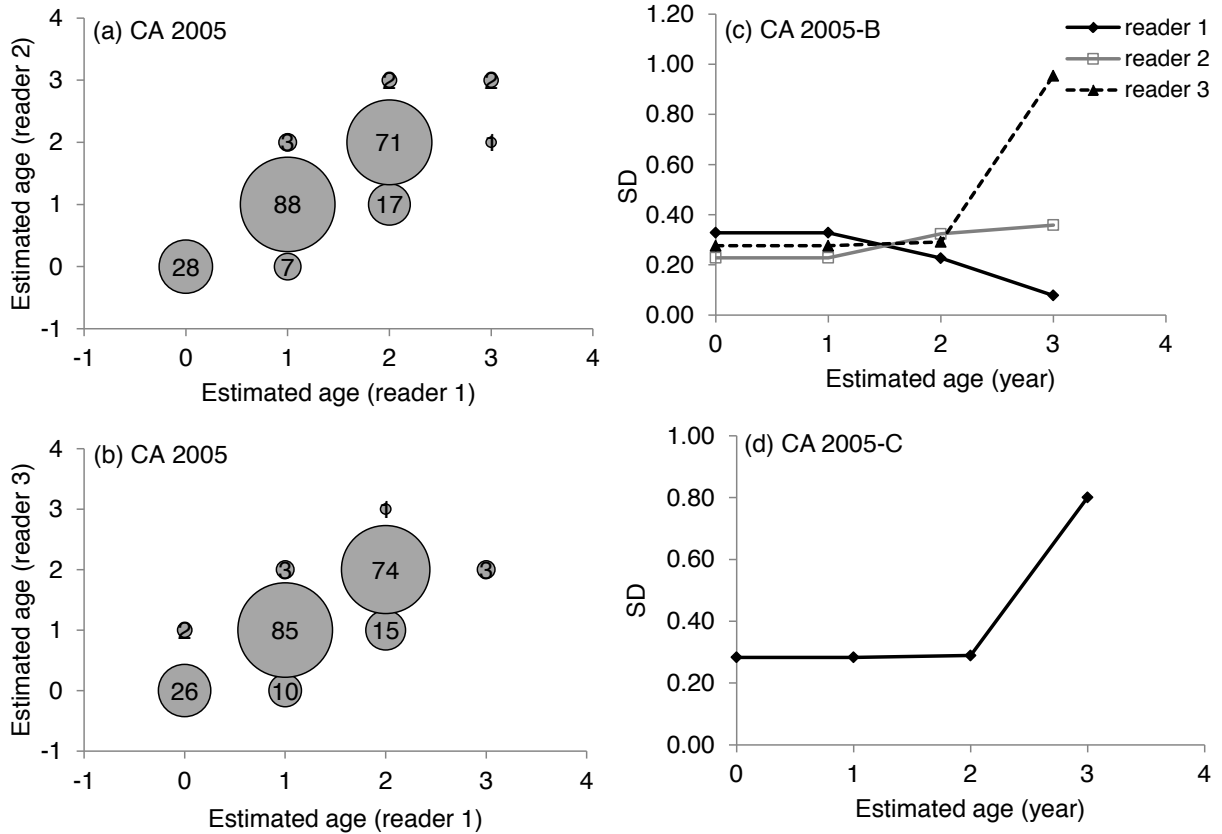


Figure 2. Age and standard deviation estimated for fish collected off California in 2005: (a–b) agreement plots for pairwise comparison of ages estimated by readers 1, 2, and 3. Number and bubble size represent the frequency of observed age combinations; (c) Standard deviation (SD) estimated from model B; and (d) SD estimated from model C.

TABLE 2
 Comparison of models A, B, and C per fishery or per survey and by collection year, based on the Akaike information criterion corrected for finite sample sizes (AIC_c). $Eff N$ is the effective sample size input or estimated by each model, and p is the number of parameters estimated from each model.

Fishery/ Survey	Model	Estimation	Bias Data set	N	Input $Eff N$	Model $Eff N$	$Eff N$ Ratio	Total Likelihood	p	AIC_c
ENS	B	No	1	240	240	1333.71	5.56	322.78		
	B	No	2	219	219	299.16	1.37	416.50	13	15.71
	C	No	2	219	219	253.68	1.16	423.87	7	2.43
	B	No	3	148	148	235.32	1.59	315.99	17	27.20
	C	No	3	148	148	76.79	0.52	337.98	8	5.39
CA	B	No	4	507	100	69.50	0.69	862.47	27	49.27
			5	145	100	22.11	0.22			
	C	No	4	507	100	75.32	0.75	855.52	15	19.11
			5	145	100	25.02	0.25			
	B	No	6	266	160	159.21	1.00	342.14	15	21.66
PNW	C	No	6	266	160	146.29	0.91	346.32	9	7.51
	A	Yes	7	711	700	2992.13	4.27	1476.56	16	18.20
	B	No	7	711	700	254.03	0.36	1494.12	13	11.91
BC	C	No	7	711	700	206.54	0.30	1502.38	10	5.69
	A	Yes	8	283	260	221.32	0.85	798.38	21	32.52
	B	No	8	283	260	47.64	0.18	878.85	15	18.41
DEPM	C	No	8	283	260	69.21	0.27	839.69	9	5.25
	A	Yes	9	360	70	69.85	1.00	2464.35	28	42.31
			10	959	800	801.89	1.00			
	B	No	9	360	70	63.51	0.91	2666.45	22	29.42
			10	959	800	25.34	0.03			
	C	No	9	360	70	70.70	1.01	2782.58	16	16.78
			10	959	800	17.88	0.02			

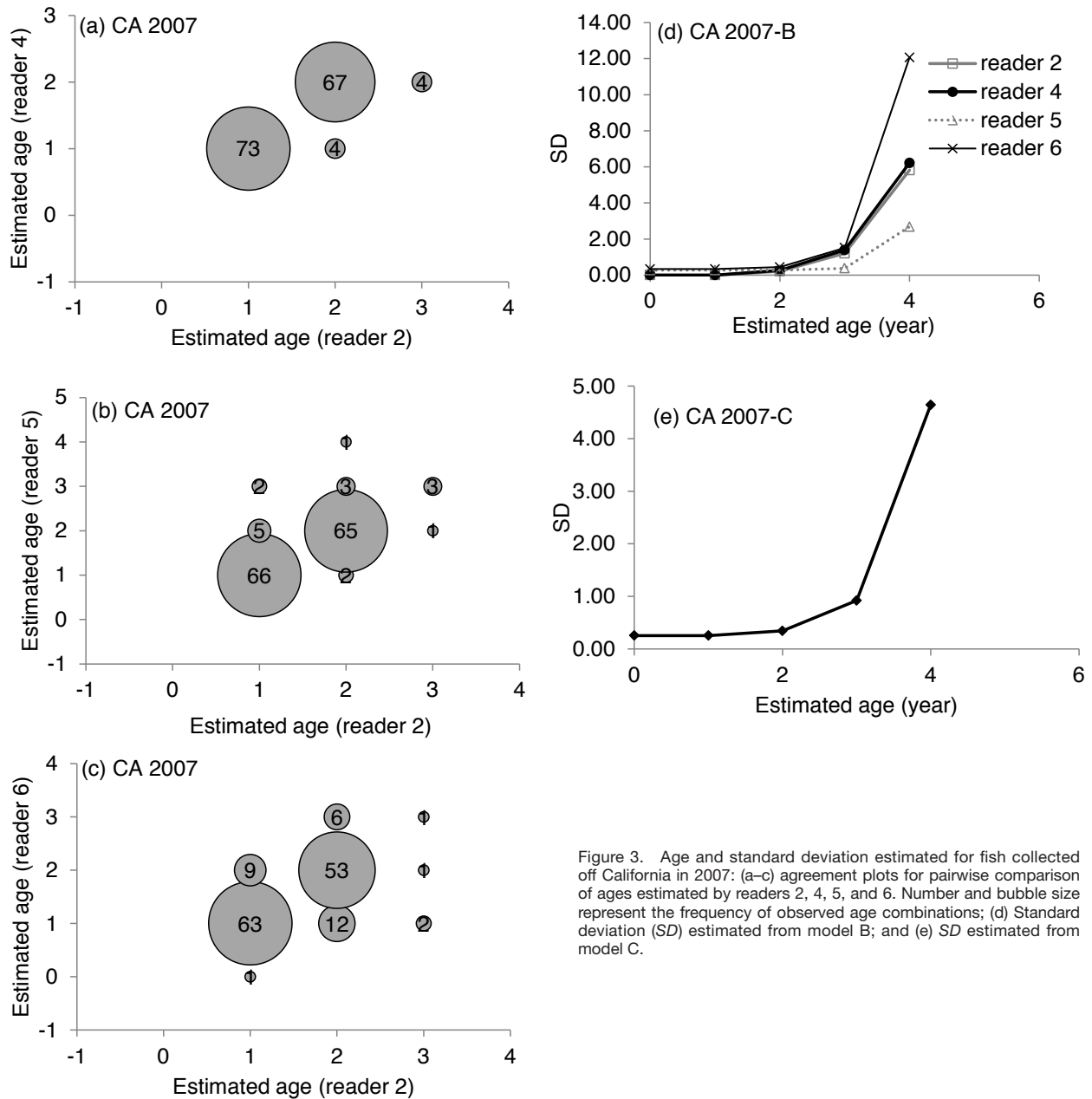


Figure 3. Age and standard deviation estimated for fish collected off California in 2007: (a–c) agreement plots for pairwise comparison of ages estimated by readers 2, 4, 5, and 6. Number and bubble size represent the frequency of observed age combinations; (d) Standard deviation (*SD*) estimated from model B; and (e) *SD* estimated from model C.

were generally smaller for ages 3 and 4. The goodness of fit of model C to the age-reading data is shown in Figure 1c, which plots the observed frequency of each age combination between the two readings in Figure 1a (e.g., 150 fish were assigned age 1 by both readings, whereas 5 were assigned age 3 at the first reading, but age 2 at the second) against the frequency predicted by the Agetat model. These data showed no evidence of over-dispersion, because most of the points lay on the 1:1 line plot (fig. 1c). Thus, we concluded that the fit of model C to the Mexican age-reading data was adequate. Note that for simplifying the presentation of the results, the diagnostic plots will not be shown for the other age-

reading data sets, although the quality of model fits to these data will be reported below.

Age-reading imprecisions were estimated for the CA fishery by reader and period of sample collection (table 2). In the 2005 and 2007 periods, there was very high agreement ($\geq 85\%$) between readers in assigning ages 0, 1, and 2 to sardine samples, leading to similar *SDs* among readers from age 0 to age 2 (figs. 2, 3). However, agreement varied from low to very high among readers in estimating age 3, resulting in considerable variability in the *SDs* at that age (figs. 2, 3). For the 2008–09 period, readers 4 and 7 highly agreed with reader 2 in assigning age 0 (92%), moderately in assigning age 1

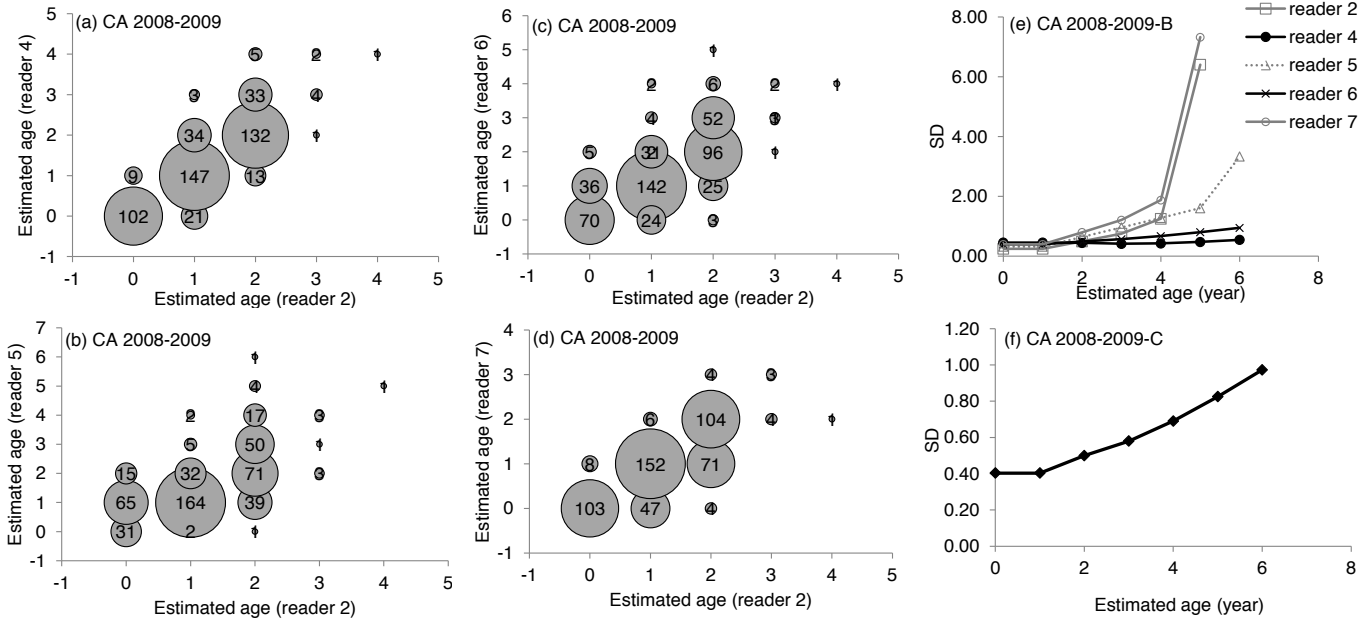


Figure 4. Age and standard deviation estimated for fish collected off California in 2008-09: (a-d) agreement plots for pairwise comparison of ages estimated by readers 2, 4, 5, 6, and 7. Number and bubble size represent the frequency of observed age combinations; (e) Standard deviation (SD) estimated from model B; and (f) SD estimated from model C.

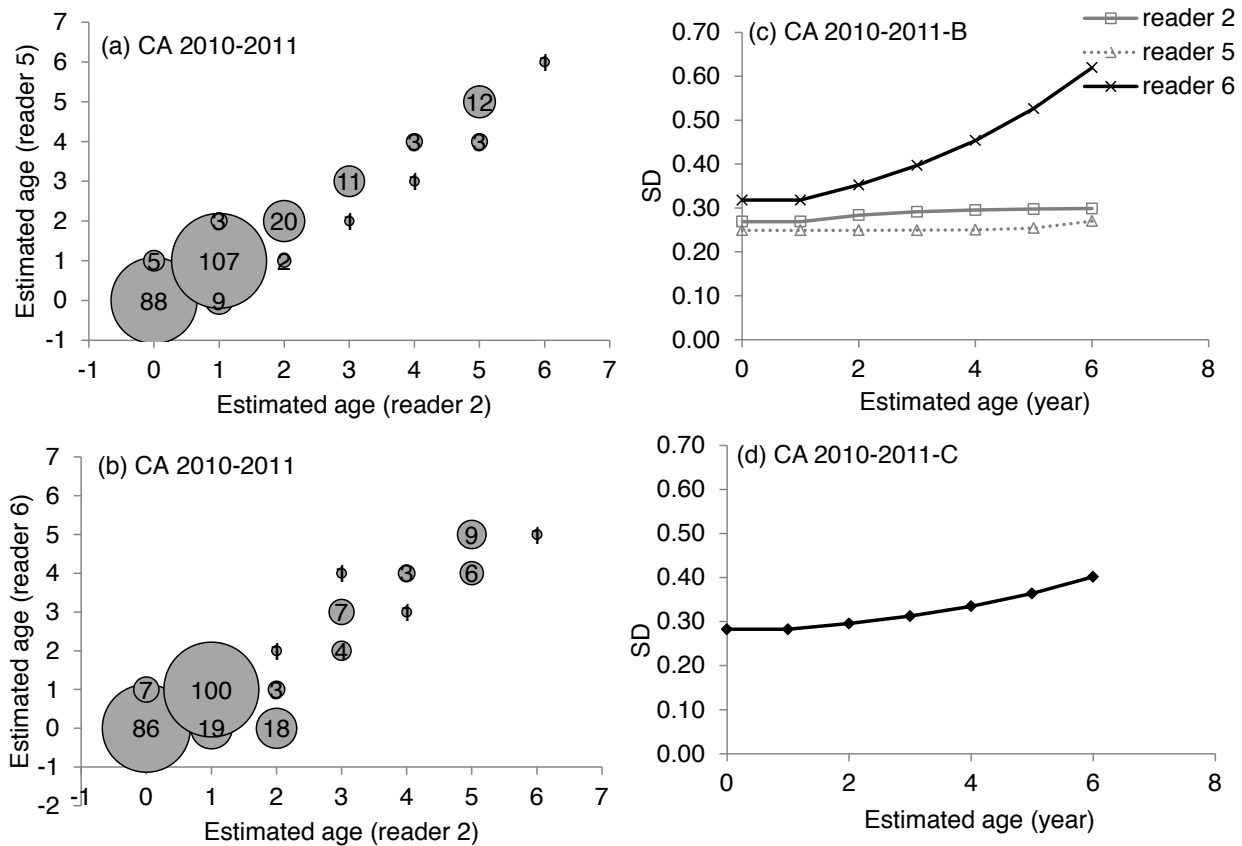


Figure 5. Age and standard deviation estimated for fish collected off California in 2010-11: (a-b) agreement plots for pairwise comparison of ages estimated by readers 2, 5, and 6. Number and bubble size represent the frequency of observed age combinations; (c) Standard deviation (SD) estimated from model B; and (d) SD estimated from model C.

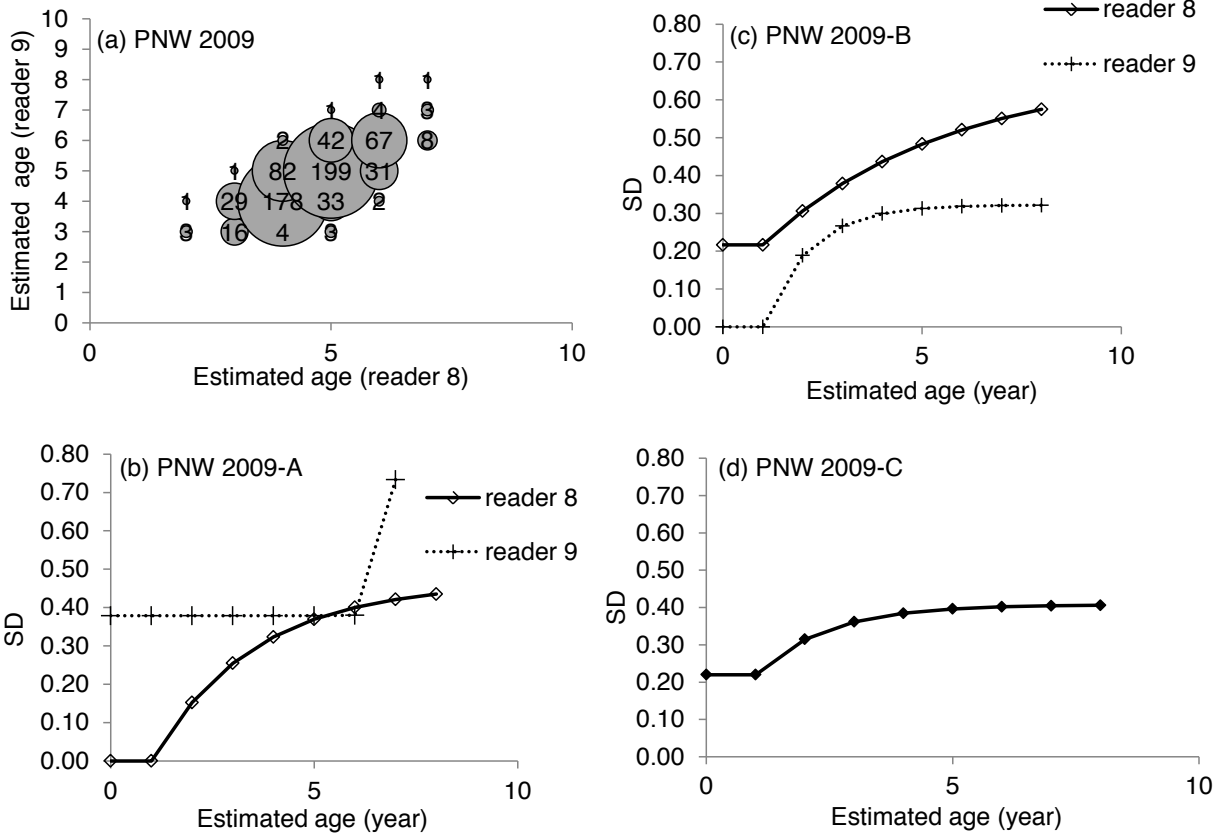


Figure 6. Age and standard deviation estimated for fish collected off the U.S. Pacific Northwest (PNW) fishery in 2009: (a) agreement plots for pairwise comparison of ages estimated by readers 8 and 9. Number and bubble size represent the frequency of observed age combinations; (b) Standard deviation (SD) estimated from model A; (c) SD estimated from model B; and (d) SD estimated from model C.

(72–74%), but their agreement in estimating age 3 was generally low, i.e., <65% (figs. 4a, 4d). Reader 5 highly agreed with reader 2 in estimating age 1 (80%), but for all other ages agreement was low between these two readers (fig. 4b). In contrast reader 6 had low agreement with reader 2 across all ages (fig. 4c). As a result SDs estimated for the 2008–09 period were similar from age 0 to 2 but differed from age 3 to 6 (fig. 4e). In 2010 and 2011, agreement between pairs of readers was very high for age 0 and 1 ($\geq 85\%$), thus SD estimates were closely similar at these ages (fig. 5). Further, reader 2 had very high agreement with reader 5 in assigning age 2, but not with reader 6. Reading agreement from ages 3 to 5 was generally low, resulting in different SDs among readers (fig. 5). Finally, the fits of models B and C showed no over-dispersion for fish aged during the 2005, 2007, and 2010–11 periods for the CA fishery. However, for the 2008–09 period, both models B and C showed considerable over-dispersion in their fits to the age-reading data sets. Changing the assumption on the functional form of the random ageing error precision could not improve these fits. Finally, for each one of the time periods considered, model C that assumed equality of SD among readers had a lower $AICc$ than model B (table 2; figs. 2d,

3e, 4f, 5d), indicating that model C generally performed better than model B.

Standard deviations at age for fish collected in the PNW were estimated for all three models. Agreement between readers was moderate in assigning ages 4 and 5, but low in estimating all other ages, i.e., <65% (fig. 6a). In model A, the unbiased reader (8) had the lowest SD at age, except for ages 5 and 6 (fig. 6b). In contrast, model B showed the opposite pattern, with the highest SD at age for the unbiased reader, across all ages (fig. 6c). The fit of model A to the PNW age-reading data showed no evidence of over-dispersion and thus was considered to be adequate. Although both models B and C showed reasonable fits to the age-reading data, these models underfitted the most frequent age combination. Despite this misfit, the model C had a lower $AICc$ value than models A and B (table 2, fig. 6d).

Ageing errors were computed for all three models, A, B, and C, for fish collected from the BC fishery. Agreement between the RA and reader 10 was very high (90%) in assigning age 3, moderate in assigning ages 4 and 6 (60–70%), but low in assigning ages 5, 7, and 8 (<65%) (fig. 7a). In contrast, reader 11 highly agreed with the RA in assigning ages 3 and 4 (80%), moder-

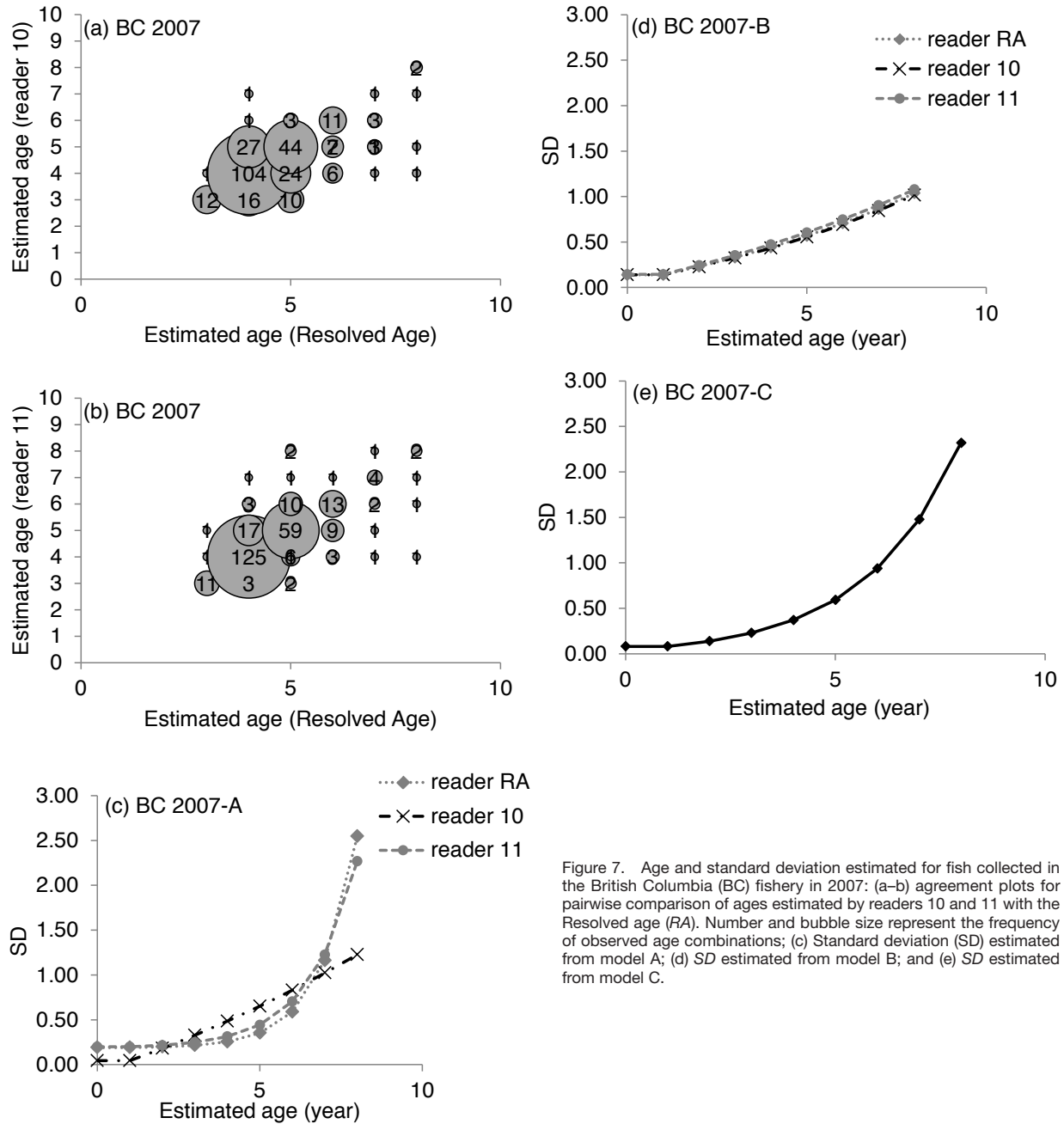


Figure 7. Age and standard deviation estimated for fish collected in the British Columbia (BC) fishery in 2007: (a–b) agreement plots for pairwise comparison of ages estimated by readers 10 and 11 with the Resolved age (RA). Number and bubble size represent the frequency of observed age combinations; (c) Standard deviation (SD) estimated from model A; (d) SD estimated from model B; and (e) SD estimated from model C.

ately in assigning age 6, but little in assigning ages 7 and 8 (40%–50%) (fig. 7b). There was little difference among the three readings in SD estimated from model A, except for age 8 (fig. 7c). Model B showed closely similar SDs among readers 10, 11, and the best/resolved age estimates (fig. 7d). Model A generally fit the age-reading data better than models B and C, but all three model fits showed little over-dispersion. Finally, model C had a lower AICc than models A and B (table 2, fig. 7e), indicating that model C performed better than the other two models.

For the DEPM survey, all three models were used to estimate SD at age and by reader. Agreement between

the pair of readers was generally low across all estimated ages (figs. 8a, 8b). Consequently, there were considerable differences in SD estimated for all three readers. In model A, the unbiased reader (1) had lower estimates of SD at age than the other two (fig. 8c). In model B, the SWFSC reader had higher SDs for ages 0 to 2 than the CDFW readers; whereas for ages 5 to 6, estimates of SD at age for the SWFSC reader were consistently lower (fig. 8d). Model A fitted both data sets well; whereas model B and model C fitted well to data set 9 but poorly to the data set 10. Changing the assumption on the functional form of the random ageing error precision could

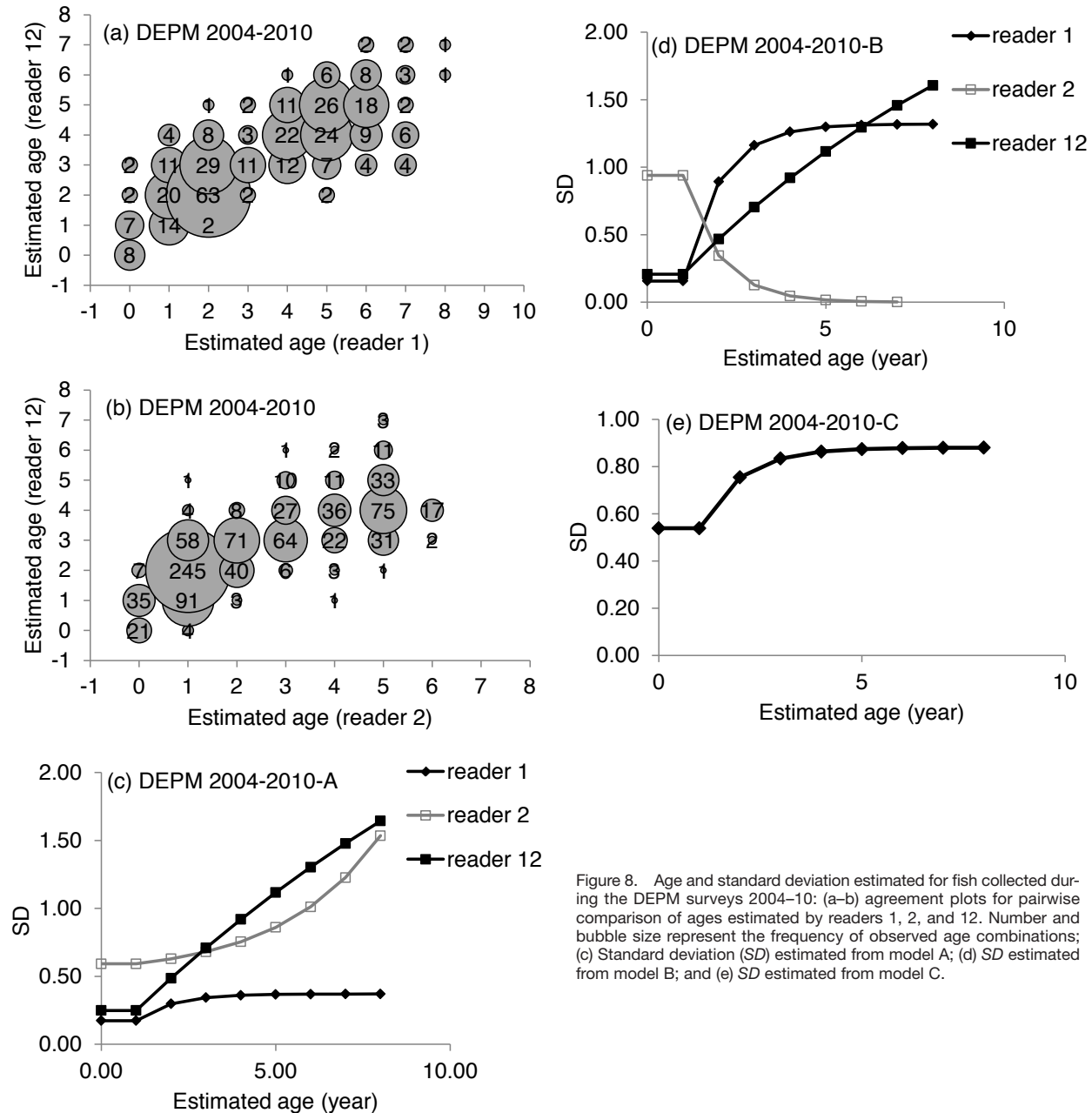


Figure 8. Age and standard deviation estimated for fish collected during the DEPM surveys 2004–10: (a–b) agreement plots for pairwise comparison of ages estimated by readers 1, 2, and 12. Number and bubble size represent the frequency of observed age combinations; (c) Standard deviation (SD) estimated from model A; (d) SD estimated from model B; and (e) SD estimated from model C.

not improve these fits. Nevertheless, model C had a lower AIC_c than models A and B (table 2, fig. 8e).

DISCUSSION

Ageing imprecisions estimated for Pacific sardines collected along the Pacific coast showed considerable differences among ages, readers, and laboratories. Estimated ages ranged from 0 to 8, and thus were much below the maximum expected age of sardine, 15 years (Hill et al. 2011). Regardless of laboratory, SD increased with age, indicating higher imprecision in counting annuli in older fish otoliths. Except for the Pacific Biologi-

cal Station, SD estimated for the oldest fish (ages 4–8) showed greater differences among readers than those estimates for the youngest individuals (ages 0–3). The bulk of the sardine population off California is made up of fish younger than 4 years (Hill et al. 2011), thus the likelihood of sampling older fish is lower than in BC, where older fish migrate during the fall (Lo et al. 2011; Demer et al. 2012). Thus, besides the difficulty of ageing older fish, these results partly reflect reader experience in ageing older sardines. Further, the Yaremko 1996 method was not consistently applied across all laboratories, with significant modifications in the use of media

and microscope resolutions. Although the impact of the use of resin or alcohol instead of water on age determination has not yet been evaluated, the change of magnifications could significantly affect the magnitude of ageing errors observed among laboratories. Except for the DEPM survey, a limitation of this study was that no common sets of otoliths were aged by all readers. Hence, testing the effect of medium quality and levels of magnification across laboratories could not be achieved. Therefore, we recommend establishing an otolith exchange program where all readers from all laboratories age fish from the same set of otoliths and use standardized ageing protocols. This program would be a more valuable approach to compare ageing imprecision across ages, readers, and laboratories. Nevertheless, the results of this study highlighted the need to improve and to standardize the application of the ageing method across laboratories.

By comparing the three hypothesized models, we showed the model that assumed similar *SD* among readers performed better than those that assumed different *SDs* among readers, across all age-reading data sets. This model had lower *AICc*, even in cases where the diagnostic plots showed considerable over-dispersion in model fitting to the age-reading data. These results suggested that assuming equal *SD* between readers within a laboratory was quantitatively a reasonable assumption. This assumption appeared to allow for better estimation of the variance contained in the data. Except for the DEPM survey, all data sets were produced by readers belonging to the same laboratory. Within a laboratory, readers are typically trained similarly and often by the most experienced reader and thus would apply closely similar techniques during the ageing process. However, assuming similar *SDs* across laboratories may be problematic, as shown by the results from the DEPM age-reading data sets. Both CDFW readers had lower standard deviations at age 0 and 1 than the reader from the SWFSC, but their *SDs* for ages 4 to 8 were higher (see model B, fig. 8d). These systematic differences in *SD* reflected both the degree of experience and consistency in applying the method within each laboratory. Therefore, for age-reading data sets produced by readers within the same laboratory, the assumption of equality of standard deviation can be used to select and include ageing imprecisions in the sardine stock assessment model. However, another limitation of our study is that Stock Synthesis 3 could only use data produced by model C, precluding the direct comparison of the outcomes of the three models within the assessment modeling framework. Likewise, it would be valuable to develop sardine assessment models in SS versions that can allow the inclusion of ageing errors from any of the three assumed models.

Although we assumed that one reader was unbiased in model scenario A, this assumption was based on reader experience and age corroboration but not on age validation (Kalish et al. 1995; Kimura et al. 2006; Campana 2011). To date, there are no fish of known age to determine age-reading accuracy of Pacific sardines. CDFW has established a training set of otoliths that has been used for age corroboration among readers and to train and certify new age readers. Because this training set does not include any fish whose ages were validated, it cannot be used to address issues concerning ageing bias. While the periodicity of sardine growth increments have been corroborated in juvenile fish (Butler 1987) and in age 1–2 fish (Barnes and Foreman 1994), to our knowledge age corroboration from annuli in older mature fish has never been conducted. As patterns of increments in young fish cannot be applied to older fish, the lack of verification of increment formation in each and every age group can lead to systematic bias in age determination (Campana 2011). Most concerns regarding bias remain with ageing fish older than four years old (i.e., the age 5⁺ group). This age group was more frequent in the Pacific Northwest and British Columbia fisheries. Interpreting increments at the edge of older fish otoliths was challenging for all readers because it is usually difficult to differentiate check marks from annuli. For example, in the first year of life, a wide opaque increment near the focus followed by a fine translucent ring can be interpreted as a check mark; whereas the same mark present in a more distal area of the otolith may be interpreted as an annulus (Yaremko 1996). Thus, although we focused on quantifying ageing imprecision in this study, we recognize there may be systematic bias in the overall ageing process. Statistical models cannot account for such potential bias, which may ultimately affect the estimation of mortality and growth rates of the northern sardine population. Current stock assessments assumed a fixed natural mortality rate ($M = 0.4 \text{ yr}^{-1}$) across all ages, which may mask the impact of ageing bias in the oldest age classes on estimating demographic parameters. Further, the natural mortality rate was estimated for fish collected in the historical fishery (Murphy 1966), prior to the full recovery of the sardine stock (McCall 1979), and thus warrants new evaluation. As sardines have reoccupied the U.S. Pacific northwest and British Columbia since the 2000s, the catches taken by these fisheries (Hill et al. 2011) and their impact on growth, mortality and recruitment processes have substantially increased. Hence, it is imperative to develop and apply new methods to accurately age fish collected in these fisheries. Therefore, a second recommendation from this study would be to conduct in the short term a coast-wide evaluation of the polished method (McFarlane et al. 2010) for age 3⁺ fish collected from Mexico to Canada.

Further, this evaluation could be coupled with a long term tagging experiment to address and quantify ageing bias in Pacific sardine.

Because the model that assumed similar *SDs* among readers consistently had the lowest *AICc*, we suggest that it can be used to estimate and include standard deviation-at-age in future SS models of Pacific sardines, particularly when the assessments are developed on age-reading data that can be partitioned by fishery and by laboratory. However, a more systematic study based on a common set of otoliths needs to be conducted to test the validity of this assumed model for computing ageing imprecisions from data that involved readers from different laboratories.

ACKNOWLEDGMENTS

We especially thank the 13 anonymous readers from CICIMAR-IPN (Baja California Sur, Mexico); CDFW (CA, USA); SWFSC (CA, USA); WDFW (WA, USA), and PBS (DFO, BC, Canada) that participated in the age-reading process for this study. We are also grateful to Dr. Andre Punt for providing the Agemat model software and for various suggestions during the modeling process. Finally, we thank B. Wells, B. Javor, and K. Hill and three anonymous reviewers for providing useful comments on this manuscript.

REFERENCES

- Barnes, J. T. and T. J. Foreman. 1994. Recent evidence for the formation of annual growth increments in the otoliths of young Pacific sardines (*Sardinops sagax*). *Calif. Fish and Game*. 80: 29–35.
- Beamish, R. J. and D. A. Fournier. 1981. A method for comparing the precision of a set of age determinations. *Can. J. Fish. and Aquat. Sci.* 38: 982–983.
- Butler, J. L. 1987. Comparison of the larval and juvenile growth and larval mortality rates of Pacific sardine and northern anchovy and implications for species interaction. Ph.D. dissertation. University of California San Diego. 242 pp.
- Butler, J. L., M. L. Granados, J. T. Barnes, M. Yaremko, and B. J. Macewicz. 1996. Age composition, growth and maturation of the Pacific sardine (*Sardinops sagax*) during 1994. *Calif. Coop. Oceanic Fish. Invest. Rep.* 37: 152–159.
- Campana, S. E., M. C. Annand, and J. I. McMillan. 1995. Graphical and statistical methods for determining the consistency of age determinations. *Trans. Amer. Fish. Soc.* 124: 131–138.
- Campana, S. E. 2011. Accuracy, precision, and quality control in age determination, including a review and abuse of age validation methods. *J. Fish Biol.* 59: 197–242.
- Chang, W. Y. B. 1982. A statistical method for evaluating the reproducibility of age determination. *Can. J. Fish. and Aquat. Sci.* 39: 1208–1210.
- Conser, R., K. Hill, P. Crone, N. Lo, and R. Félix-Uraga. 2004. Assessment of the Pacific sardine stock for U.S. management in 2005. Pacific Fishery Management Council, November 2004. 125 pp.
- Demer, D. A., J. P. Zwolinski, K. A. Byers, G. R. Cutter, J. S. Renfree, T. S. Sessions, and B. J. Macewicz. 2012. Prediction and confirmation of seasonal migration of Pacific sardine (*Sardinops sagax*) in the California Current Ecosystem. *Fish. Bull.* 110: 52–70.
- Deriso, R. T., J. T. Barnes, L. D. Jacobson, and P. J. Arenas. 1996. Catch-age analysis for Pacific sardine (*Sardinops sagax*), 1983–95. *Calif. Coop. Oceanic Fish. Invest. Rep.* 37: 175–187.
- Fitch, J. E. 1951. Age composition of the southern California catch of Pacific mackerel 1939–40 through 1950–51. *Calif. Dept. of Fish and Game, Fish Bull.* 83: 1–73.
- Fournier, D. and C. P. Archibald. 1982. A general theory for analyzing catch at age data. *Can. J. Fish. and Aquat. Sci.* 39: 1195–1207.
- Hill, K. T., P. R. Crone, N. C. H. Lo, B. J. Macewicz, E. Dorval, J. D. McDaniel, and Y. Gu. 2011. Assessment of the Pacific sardine resource in 2011 for U.S. management in 2012. NOAA Technical Memorandum, NOAA-NMFS, 265 pp.
- Hill, K. T., N. C. H. Lo, B. J. Macewicz, P. R. Crone, and R. Félix-Uraga. 2009. Assessment of the Pacific sardine resource in 2009 for U.S. management in 2010. NOAA Technical Memorandum, NMFS-SWFSC-487, 241 pp.
- Hill, K. T., E. Dorval, N. C. H. Lo, B. J. Macewicz, C. Show, and R. Félix-Uraga. 2007. Assessment of the Pacific sardine resource in 2007 for U.S. Management in 2008. NOAA Technical Memorandum, NMFS-SWFSC-413, 157 pp.
- Kalish, J. M., R. J. Beamish, E. B. Brothers, J. M. Casselman, R. I. C. C. Francis, H. Mosegaard, J. Panfili, E. D. Prince, R. E. Thresher, C. A. Wilson, and P. J. Wright. 1995. Glossary for otolith studies. In D. S. Secor, J. M. Dean, and S. E. Campana (eds), Recent development of fish otolith. University of South Carolina Press, Columbia, South Carolina. pp. 723–729.
- Kimura, D. K., C. R. Kestelle, B. J. Goetz, C. M. Gburski, and A. V. Buslov. 2006. Corroborating the ages of walleye Pollock (*Theragra chalcogramma*). *Mar. Fresh. Res.* 57: 323–332.
- Lo, N. C. H., B. J. Macewicz, and D. A. Griffith. 2011. Migration of Pacific sardine (*Sardinops sagax*) off the West coast of United States in 2003–2005. *Bull. Mar. Sci.* 87: 395–412.
- Lo, N. C. H., B. J. Macewicz, and D. A. Griffith. 2005. Spawning biomass of Pacific sardine (*Sardinops sagax*) from 1994 to 2004. *Calif. Coop. Oceanic Fish. Invest. Rep.* 46: 93–112.
- Marr, J. C. 1960. The causes of the major variations of the catch of Pacific sardines, *Sardinops caerulea* (Girard). In Proceeding of the World Scientific Meetings on the biology of Pacific sardine and related species. Rome September 14–29, 1959. Stock and Area Paper 3, Vol. III, 667–691 pp.
- McCall, A. D. 1979. Population estimates for the waning years of the Pacific sardine fishery. *Calif. Coop. Oceanic Fish. Invest. Rep.* 20: 72–82.
- McFarlane, G., J. Schweigert, V. Hodes, and J. Detering. 2010. Preliminary study on the use of polished otoliths in the age determination of Pacific sardine (*Sardinops sagax*) in British Columbia waters. *Calif. Coop. Oceanic Fish. Invest. Rep.* 51: 162–168.
- Method, R. 2000. Technical description of the Stock Synthesis Assessment Program. NOAA Tech. Memo. NMFS-NWFSC-43, 46p.
- Morison, A. K., S. G. Robertson, and D. C. Smith. 1998. An integrated system for production fish ageing: image analysis and quality assurance. *N. Am. J. Fish. Manag.* 18: 587–598.
- Mosher, K. H. and H. H. Eckles. 1954. Age determination of Pacific sardines from otoliths. *U.S. Fish. Wildl. Ser. Res. Rep.* 37, 40 pp.
- Murphy, G. I. 1967. Vital statistics of the Pacific sardine (*Sardinops caerulea*) and the population consequences. *Ecology*. 48: 731–736.
- Murphy, G. I. 1966. Population biology of the Pacific sardine (*Sardinops caerulea*). *Proc. Calif. Acad. Sci.* 34: 1–84.
- Phillips, J. B. 1948. Growth of the sardine, *Sardinops caerulea* 1941–42 through 1946–47. *Calif. Div. Fish and Game. Fish Bull. No. 71*, 19 pp.
- Punt, A. E., D. C. Smith, K. Krusisc Golub, and S. Robertson. 2008. Quantifying age-reading error for use in fisheries stock assessments, with application to species in Australia's southern and eastern scalefish and shark fishery. *Can. J. Fish. and Aquat. Sci.* 65: 1991–2005.
- Reeves, S. A. 2003. A simulation study of the implication of age reading errors for stock assessment and management advice. *ICES J. Mar. Sci.* 60: 314–328.
- Richards, L. J., J. T. Schnute, A. R. Kronlund, and R. J. Beamish. 1992. Statistical models for the analysis of ageing error. *Can. J. Fish. and Aquat. Sci.* 49: 1801–1815.
- Rogers, P. J. and T. M., Ward. 2007. Application of a 'case building approach' to investigate the growth distributions and growth dynamics of Australian sardine (*Sardinops sagax*) off South Australia. *Mar. Fresh. Res.* 58: 461–474.
- Walford, L. A., and K. H. Mosher. 1950. Determination of the age of juveniles by scales and otoliths. In Studies on the Pacific sardine or pilchards (*Sardinops caerulea*). United States Department of Interior, Fish and Wildl. Ser. Special Scientific Report–Fisheries, No 15, pp. 31–95.
- Wolf, R. S. 1961. Age composition of the Pacific sardine 1932–60. United States Fish and Wildl. Serv. Research Report 53, 36 pp.
- Yaremko, M. L. 1996. Age determination in Pacific sardine, *Sardinops sagax*. NOAA Technical Memorandum, NMFS SWFSC-223. 33 pp.