Review of Age Precision Metrics with Recommendations for Future Use

##### Derek H. Ogle

###### Department of Mathematical Sciences and Department of Natural Resources, Northland College, 1411 Ellis Ave, Ashland, WI 54806

##### Joshua XX. Lyons

###### Department of Natural Resources, Northland College, 1411 Ellis Ave, Ashland, WI 54806

##### Gordon XX. Scott

###### Department of Natural Resources, Northland College, 1411 Ellis Ave, Ashland, WI 54806

Running head: Age Precision Metrics Review

# Abstract

WILL BE <300 WORDS HERE

# Introduction

Age, along with length, is one of the most important characterisitics of individual fish recorded by fisheries scientists. Age data is used to estimate growth, mortality, and recruitment (Spurgeon et al. 2015; Kerns and Lombardi-Carlson 2017; Quist and Isermann 2017), which are primary rates of interest for fisheries scientists (Ricker 1975; Hilborn and Walters 1992; Haddon 2011; Paukert and Spurgeon 2017), and in age-structured population models to estimate yield and set harvest regulations (Beverton and Holt 1957; Hilborn and Walters 1992; Haddon 2011; Methot and Wetzel 2013). Age, unlike length, is not directly observed but is most often interpreted from periods of fast and slow growth recorded on calcified structures such as scales, otoliths, fin rays or spines, or vertebrae (Campana 2001; Spurgeon et al. 2015; Quist and Isermann 2017). Age data can be highly variable and prone to errors as age may not be reliably recorded on or interpreted from calcified structures, especially for fish that are more than only a few years old (Campana 2001; Quist et al. 2012; Phelps et al. 2017). Errors in age data have resulted in incorrect estimates of growth, mortality, and recruitment which have led to misunderstanding the population dynamics of some fish populations (Lin Lai and Gunderson 1987; Reeves 2003; Bertignac and Pontual 2007; Yule et al. 2008; Koenigs et al. 2013; Hamel et al. 2016; Tyszko and Pritt 2017; Chang et al. 2019).

Sources of error in age estimation may be inherent to the calcified structure itself or part of the process of interpreting age from the calcified structure (Campana 2001; Morison et al. 2005; Buckmeier et al. 2017). Errors inherent to the structure are measurable, but cannot be controlled (Morison et al. 2005). Assessing this source of error is part of validating ages estimated from calcified structures, methods for which have been thoroughly reviewed (Maceina et al. 2007; Spurgeon et al. 2015; Buckmeier et al. 2017) and are not addressed further here. Errors related to the process of interpreting ages can be both measured and controlled (Morison et al. 2005) and consist of differences between the estimated and true age (i.e., accuracy) and differences among estimated ages from multiple interpretations of the same structure (either by multiple interpreters or the same interpreter multiple times). Differences among interpretations of the same structure may be systematic (e.g., one set of ages is always one year less than the other set of ages), patterned (e.g., one set of ages largely match the other set of ages up to a certain age and then become progressively more different after that age), or random. Identifying or measuring systematic or patterned differences among interpretations is referred to as identifying *bias* (or not) in the age estimates (Campana et al. 1995). Methods for assessing bias are described by Campana et al. (1995) and are also not discussed further here. In the absence of bias, *precision* is a measure of random differences among interpretations, or the repeatability of age estimates among interpreters (Campana 2001; Maceina et al. 2007) that do not display a bias (Campana et al. 1995). Precision and metrics for assessing precision are the focus this paper.

Precision metrics measure the repeatability of age estimates made on each of individual fish. Percent agreement, the simplest precision metric, is the percentage of fish for which the age estimates are the same (i.e., *exact percent agreement*) or differ by no more than a certain amount (e.g., differ by no more than one year). When , *percent partial agreement* is the percentage of fish for which the ages agree for some specific number of interpreters (e.g., at least two interpreters agree).

Other measures of precision result from computing various measures of variability among the age estimates within each of the fish and then averaging these values across the fish. The first two measures of variability are the familiar absolute deviation (AD) and standard deviation (SD) of estimated ages for each fish; i.e.,

where is the th age estimate for the th fish and is the mean of the age estimates for the th fish. The and may be scaled by to produce the percent error (PE) and coefficient of variation (CV) for the th fish, respectively; i.e.,

When and are averaged across all fish they produce the overall measures of precision called *average percent error* [; Beamish and Fournier (1981)] and *average coefficient of variation* [; Chang (1982)]. These measures are usually multiplied by 100 and expressed as a percentage. For example, the formulae for and are

When , (Chang 1982; Kimura and Anderl 2005) and, thus, and only differ by a constant value and provide functionally the same information about precision among age estimates. For , is directly proportional to (Chang 1982). However, this proportionality depends on and, thus, there is no constant proportionality between and , though they will likely be strongly correlated (Kimura and Anderl 2005).

Percent agreement, , and are the most commonly used measures of precision (Campana 2001; Morison et al. 2005). Percent agreement has been criticized as a useful composite measure of precision, largely because it varies widely among species and among ages within a species (Beamish and Fournier 1981; Kimura and Lyons 1991; Campana et al. 1995; Campana 2001; Maceina et al. 2007). Despite these critiques and likely due to its simplicity, Morison et al. (2005) found that 49% of 53 laboratories from 23 countries used percent agreement as a measure of precision among estimated ages. Beamish and Fournier (1981) proposed the to address the shortcomings of the percent agreement. Chang (1982) introduced , though he called it , as a better alternative to the because the standard deviation is a better estimator (unbiased and consistent) of variability than the absolute deviation, though Kimura and Lyons (1991) noted that this was only true if the age estimates were normally distributed.

In a review of 131 papers that reported precision values for fish age estimates, Campana (2001) found that most papers used the (57%) to measure precision, though and were used almost equally for studies of annual age and was used predominantly (84%) when estimating daily age from otoliths. The median and modal across 117 studies deemed to have valid precision estimates were 7.6% and 5%, respectively. No significant differences in precision were found among calcified structures used to estimate annual age. Campana (2001) reiterated that there is a direct conversion between and when and that both measures are “equally sensitive to precision differences among agers.” From this, Campana (2001) stated “it is not self-evident that one measure [APE or ACV] is to be preferred over the other.” From his literature review and informal discussions with scientists from fish ageing laboratories, Campana (2001) concluded that an of 5% would represent reasonable precision for many fish of moderate longevity and complexity of structure interpretation.

Other metrics of precision have been proposed or could be derived using similar logic. Chang (1982) also introduced the index of precision for the th fish as . Kimura and Anderl (2005) suggested, based on distributional theory, that the median age estimate for the th fish () should be used instead of for . Presumably the same argument can be applied to . This change results in a modified percent error and coefficient of variation for the th fish; i.e.,

Bauerlien et al. (2018)] suggested that if the variability among age estimates is not related to the age of the fish then may be an appropriate measure of variability, as compared to . Similar arguments suggest that may be an appropriate measure of variability, as compared to . Averaging these measures across all fish gives five alternative measures of precision among age estimates – *average index of precision* [; Chang (1982)], *average modified percent error* (), *average modified coefficient of variation* (), *average absolute deviation* [; Bauerlien et al. (2018)], and *average standard deviation* (). All but the and are typically multiplied by 100 and reported as a percentage.

Not suprisingly there are predictable relationships between several of these measures. When , , and because , and (Campana et al. 1995). Thus, when , , , , , and provide functionally the same information about precision among age estimates. Similarly, and provide functionally the same information about precision among age estimates. For , and are proportional and, thus, do not provide functionally different information about precision among age estimates.

It has been more than two decades since Campana (2001) reviewed the use of precision metrics in the published literature and made recommendation for their use. It has also been 15 years since Maceina et al. (2007) reported on a survey of fisheries scientists with respect to their use of precision metrics. In relatively recent years, Bauerlien et al. (2018) suggested that using a single summary precision metric may not be good practice and Buckmeier et al. (2017) described how the ASD may be more appropriate than the in some instances. Given the length of time since the last synthetic review of precision metrics and these newly raised questions, we report here on our comprehensive review of papers published since 2002 that reported age precision metrics and our extended analyses of data from multiple age interpretations reported in a variety of published papers. Our overall goal is to summarize the use and value of precision metrics since 2001, show results for precision metrics that are not typically used in the literature, and to provide specific recommendations on the use of precision metrics with respect to fish ageing studies. Specifically, we will answer the following questions from our review of the literature:

1. How often are each of the seven precision metrics introduced above used?
2. What is the distribution and summary statistics of values reported in the literature?
3. Do values reported in the literature differ by type of comparison (between or within interpreter), number of repeated interpretations (i.e., ), calcified structure, range of ages, sample size, or other characteristics of the fish (e.g., taxonomic category) or study (e.g., preparation method for the structure).
4. How often did the authors check for no bias before computing the precision metric?
5. How often did the authors check for a relationship between the precision metric and age?
6. How often was the precision metric related to age (when this relationship was examined)?

Under the *a priori* assumptions that only percent agreement, , and will be used extensively, few papers will have tested for bias prior to computing the precision metric, and few papers will have tested for a relationship between the precision metric and age, we will also answer the following questions from an extended analysis of some published data sets:

1. How are the various precision metrics related (beyond the known relationships identified above)?
2. How often is a bias detected (and, thus, precision should not be calculated)?
3. How often is the precision metric related to age (and, thus, a single overall measure of precision should not be calculated)?
4. What kind of relationship (linear, quadratic, or more complicated) between precision metric and age is most common?

# [A]Methods

[C]*Literature Review.*– XXX

[C]*Extended Analyses.*– XXX

WILL BE FLESHED OUT

# [A]Results

## [B]Literature Review

The final literature review data consisted of 1254 individual precision assessments from 383 papers published between 1996 and 2020. The number of papers with precision assessments generally increased over this 25 year period (Figure 1). Otoliths were the most commonly assessed calcified structure (51% of all assessments for Actinopteri), though vertebrae were most commonly assessed among elasmobranchs (82% of all assessments for Elasmobranchii; Figure 2). The number of precision assessments for Actinopteri increased in the late 2000s, primarily due to increased assessments of scales, spines, and fin rays, and then again in the late 2010s due to more otolith assessments (Figure 2). The number of precision assessments for Elasomabranchii was fairly constant over the 25 years, though the assessment of spines increased beginning in the late 2000s (Figure 2). Most assessments were between-readers rather than within-readers, with the exception that 53% of assessments with vertebrae were within-readers (Figure 3). The majority of assessments were between two readings (69%; Figure 4). The percentage of assessments using two readings increased to a peak in the late 2000s and decreased into the 2010s, primarily replaced with assessments among three readings, though assessments with four or more readings were more prevalent in the late 2010s (Figure 4).

The was the most used precision metric, though the and were used nearly as much (Figure 5). The was most often used simultaneously with either the , , or and, thus, was used rarely by itself (17% of studies that used ). The and were commonly used together with 46% of studies that used also using and 58% of studies that used also using . The percent of studies that used the generally increased from 1996 to 2020, whereas the percent of studies that used the was fairly steady before increasing in the late 2010s (Figure 6). However, the percent of studies that used the without also using the declined two approximately 20% from the early to late 2000s and has stayed steady since then (Figure 6). The was used rarely and always with either the or (Figure 5). The and were each used in only one study, and and were not used in any studies.

[C]*Extended Analyses.*– XXX

WILL BE FLESHED OUT

# [A]Discussion

## [B]Subsection

WILL BE FLESHED OUT

# [A]Acknowledgments

WILL BE FLESHED OUT

# Figures

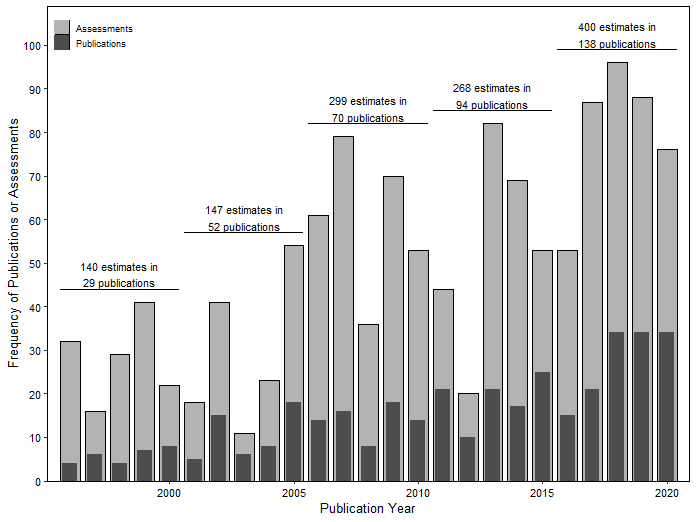


Figure 1

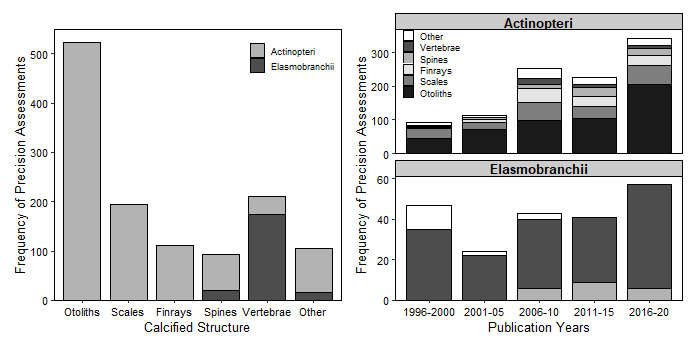


Figure 2

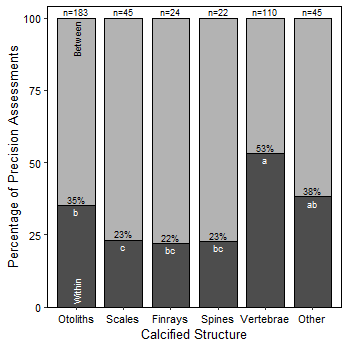


Figure 3

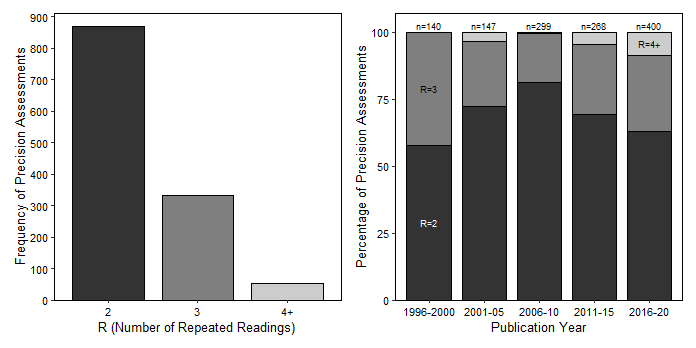


Figure 4

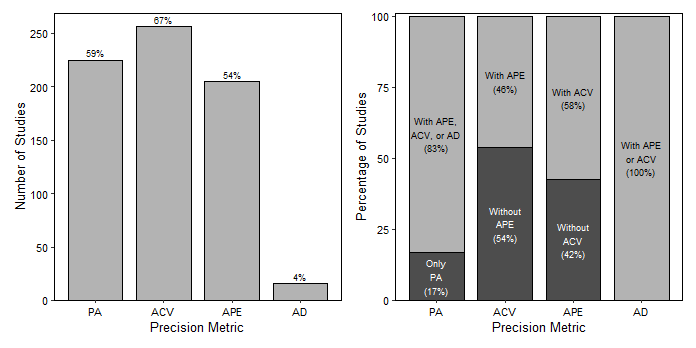


Figure 5

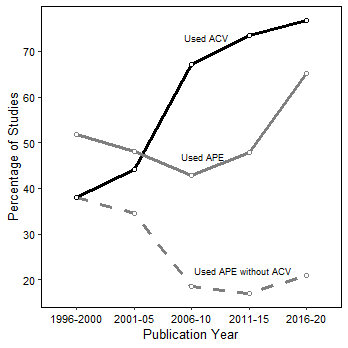


Figure 6

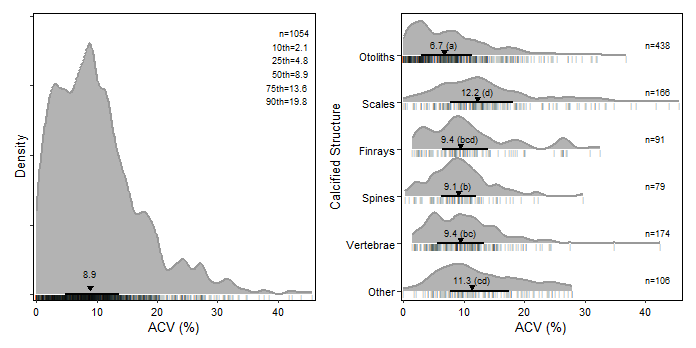


Figure 7

# Figure Captions

Figure 1. Number of individual assessments and number of publications with precision assessments by year from 1996-2020 used in the literature review. Total number of assessments and publications are also shown for five year periods.

Figure 2. Number of individual precision assessments by primary calcified structure types separated by major taxonomic class (Left) and further separated by five year periods (Right). Six and ten assessments for Holocephali and Petromyzonti are not included.

Figure 3. Percentage of individual precision assessments within calcified structure types that were within-reader assessments. Structures with the same letters have statistically equal percent of within-reader assessments. Overall sample size for each structure is shown above the bars. Eight observations that combined within- and between-reader assessments are not included.

Figure 4. Number of individual precision assessments by the number of repeated readings (Left) and further separated by five year period (Right) . The maximum number of repeated readings was R=9.

Figure 5. Number of studies that used each of the main precision metrics (Left) and the percentage of studies using each metric separated by how the metric was used relative to the other main metrics (Right). Note that =Percent Agreement, =Average Coefficient of Variation, =Average Percent Error, =Average Index of Precisions.

Figure 6. Percentage of studies that used the Average Coefficient of Variation (), Average Percent Error (), or the alone without also using the by five year period.

Figure 7. Density distribution and individual values (vertical marks) of Average Coefficient of Variation () values for all assessment (Left) and for assessments by main calcified structure types (Right). The median value is shown and marked with a triangle. The inter-quartile range (first to third quartile) is shown by the black horizontal bar. Calcified structures with different letters have significantly different medians. Note that this includes all Average Percent Error () values for two readings (=2) converted to values.

# [A]References

Need to fix species capitalization

Bauerlien, C. J., M. R. Cornett, E. A. Zielonka, D. P. Crane, and J. S. Bulak. 2018. Precision of calcified structures used for estimating age of Chain Pickerel <span class="nocase">*esox*</span> *niger*. North American Journal of Fisheries Management 38.

Beamish, R. J., and D. A. Fournier. 1981. A method for comparing the precision of a set of age determinations. Canadian Journal of Fisheries and Aquatic Sciences 38:982–983.

Bertignac, M., and H. de Pontual. 2007. Consequences of bias in age estimation on assessment of the northern stock of European Hake (<span class="nocase">*merluccius*</span> *merluccius*) and on management advice. ICES Journal of Marine Science 64(5):981–988.

Beverton, R. J. H., and S. J. Holt. 1957. On the dynamics of exploited fish populations. United Kingdom Ministry of Agriculture, Fisheries, London.

Buckmeier, D. L., P. C. Sakaris, and D. J. Schill. 2017. Validation of annual and daily increments in calcified structures and verification of age estimates. Pages 33–80 *in* M. C. Quist and D. A. Isermann, editors. Age and growth of fishes: Principles and techniques. American Fisheries Society, Bethesda, Maryland.

Campana, S. 2001. Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. Journal of Fish Biology 59:197–242.

Campana, S. E., M. C. Annand, and J. I. McMillan. 1995. Graphical and statistical methods for determining the consistency of age determinations. Transactions of the American Fisheries Society 124:131–138.

Chang, W. Y. B. 1982. A statistical method for evaluating the reproducibility of age determination. Canadian Journal of Fisheries and Aquatic Sciences 39:1208–1210.

Chang, Y.-J., J. Hsu, J.-C. Shiao, and S.-K. Chang. 2019. Evaluation of the effects of otolith sampling strategies and ageing error on estimation of the age composition and growth curve for Pacific Bluefin Tuna <span class="nocase">*thunnus*</span> *orientalis*. Marine and Freshwater Research.

Haddon, M. 2011. Modelling and quantitative methods in fisheries, 2nd edition. CRC Press, Boca Raton, FL.

Hamel, M. J., J. J. Spurgeon, C. J. Chizinski, K. D. Steffensen, and M. A. Pegg. 2016. Variability in age estimation results in ambiguity and false understanding of population persistence. North American Journal of Fisheries Management 36:514–522.

Hilborn, R., and C. J. Walters. 1992. Quantitative fisheries stock assessment: Choice, dynamics and uncertainty. Chapman; Hall, New York.

Kerns, J. A., and L. Lombardi-Carlson. 2017. History and importance of age and growth information. Pages 1–8 *in* M. C. Quist and D. A. Isermann, editors. Age and growth of fishes: Principles and techniques. American Fisheries Society, Bethesda, Maryland.

Kimura, D. K., and D. M. Anderl. 2005. Quality control of age data at the Alaska Fisheries Science Center. Marine and Freshwater Research 56:783–789.

Kimura, D. K., and J. J. Lyons. 1991. Between-reader bias and variability in the age-determination process. Fisheries Bulletin, U.S. 89:53–60.

Koenigs, R. P., R. M. Bruch, and K. K. Kamke. 2013. Impacts of aging error on Walleye management in the Winnebago System. North American Journal of Fisheries Management 33:900–908.

Lin Lai, H., and D. R. Gunderson. 1987. Effects of ageing errors on estimates of growth, mortality and yield per recruit for Walleye Pollock (<span class="nocase">*theragra*</span> *chalcogramma*). Fisheries Research 5:287–302.

Maceina, M. J., J. Boxrucker, D. L. Buckmeier, R. S. Gangl, D. O. Lucchesi, D. A. Isermann, J. R. Jackson, and P. J. Martinez. 2007. Current status and review of freshwater fish aging procedures used by state and provincial fisheries agencies with recommendations for future directions. Fisheries 32(7):329–340.

Methot, R. D., and C. R. Wetzel. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries Research 142:86–99.

Morison, A. K., J. Burnett, W. J. McCurdy, and E. Moksness. 2005. Quality issues in the use of otoliths for fish age estimation. Marine and Freshwater Research 56:773–782.

Paukert, C. P., and J. J. Spurgeon. 2017. Age structure. Pages 221–232 *in* M. C. Quist and D. A. Isermann, editors. Age and growth of fishes: Principles and techniques. American Fisheries Society, Bethesda, Maryland.

Phelps, Q. E., S. J. Tripp, M. J. Hamel, R. P. Koenigs, and Z. J. Jackson. 2017. Choice of structure for estimating fish age and growth. Pages 81–106 *in* M. C. Quist and D. A. Isermann, editors. Age and growth of fishes: Principles and techniques. American Fisheries Society, Bethesda, Maryland.

Quist, M. C., and D. A. Isermann. 2017. Age and growth of fishes: Principles and techniques. American Fisheries Society, Bethesda, Maryland.

Quist, M. C., M. A. Pegg, and D. R. DeVries. 2012. Age and growth. Pages 677–732 *in* A. V. Zale, D. L. Parrish, and T. M. Sutton, editors. Fisheries techniques, 3rd edition. American Fisheries Society, Bethesda, Maryland.

Reeves, S. 2003. A simulation study of the implications of age-reading errors for stock assessment and management advice. ICES Journal of Marine Science 60:314–328.

Ricker, W. E. 1975. Computation and interpretation of biological statistics of fish populations. Bulletin of the Fisheries Research Board of Canada 191:1–382.

Spurgeon, J. J., M. J. Hamel, K. L. Pope, and M. A. Pegg. 2015. The global status of freshwater fish age validation studies and a prioritization framework for further research. Reviews in Fisheries Science & Aquaculture 23:329–345.

Tyszko, S. M., and J. J. Pritt. 2017. Comparing otoliths and scales as structures used to estimate ages of Largemouth Bass: Consequences of biased age estimates. North American Journal of Fisheries Management 37:1075–1082.

Yule, D. L., J. D. Stockwell, J. A. Black, K. I. Cullis, G. A. Cholwek, and J. T. Myers. 2008. How systematic age underestimation can impede understanding of fish population dynamics: Lessons learned from a Lake Superior Cisco stock. Transactions of the American Fisheries Society 137:481–495.