

# Proposal

## Background

Age estimation of fish is necessary for the assessment of fish population dynamics (e.g., recruitment, growth, and mortality) and stock structure. Ages are most often estimated from calcified structures. Those ages may vary among readers, among calcified structures, or among readings by the same reader. It is important to understand both the accuracy and precision of age estimates to better understand how biases and variability may affect estimates of population dynamic metrics.

Accuracy is related to estimating the true age of the fish. Accuracy is related to validation which was extensively discussed by Campana (2001) and won't be discussed further here.

Precision is related to the consistency or repeatability of age estimates among readings (i.e., among readers or among readings by the same reader at different times). Precision was also discussed extensively in Campana (2001) but will be discussed further here given comments and questions in Bauerlien et al. (2018).

Suppose that  $R$  age estimates are made on  $n$  fish in a sample and, for simplicity, that  $R$  is constant (e.g., always  $R$  readers of the structure). Further suppose that  $x_{ij}$  is the  $i$ th age estimate for the  $j$ th fish and that  $\bar{x}_j$  is the mean of the  $R$  age estimates for the  $j$ th fish. For example, the data may look like the following where **fishID** is a unique fish identification number, **tl** is the fish's total length (mm), and **otolith1** and **otolith2** are ages estimates made from otoliths by two readers.

fishID	tl	otolith1	otolith2
1	345	3	3
10	173	1	1
20	478	7	7
30	346	2	3
52	350	4	4
68	568	17	18
91	518	10	11
93	550	12	11

There are at least four common measures of precision for an individual fish (say fish  $j$ ) – average absolute deviation (AAD),<sup>1</sup> standard deviation (SD), average percentage error (APE),<sup>2</sup> and coefficient of variation (CV).<sup>3</sup> These four measures are defined below.

$$AAD_j = \frac{\sum_{i=1}^R |x_{ij} - \bar{x}_j|}{R}$$

$$SD_j = \sqrt{\frac{\sum_{i=1}^R (x_{ij} - \bar{x}_j)^2}{R - 1}}$$

<sup>1</sup>Some authors refer to this as the *mean absolute deviation* (MAD). We use AAD here instead for congruency with abbreviations for other methods.

<sup>2</sup>Traditionally, APE has been used for the average of what is called the  $APE_j$  here. In other words, the traditional APE is equivalent to our AAPE.

<sup>3</sup>Traditionally, CV has been used for the average of what is called the  $CV_j$  here. In other words, the traditional CV is equivalent to our ACV. This follows Ogle (2016).

$$APE_j = \frac{AAD_j}{\bar{x}_j}$$

$$CV_j = \frac{SD_j}{\bar{x}_j}$$

From these formulae, it is seen that  $APE_j$  is simply  $AAD_j$  rescaled by dividing by the mean age estimate. Similarly,  $CV_j$  is simply  $SD_j$  rescaled in the same way. The thought process behind these rescalings is that if the error in age estimates increases with age then the measure of precision should be adjusted by age. In other words, a difference of one year on a 10-year-old fish will be treated the same as a difference of two years on a 20-year-old fish. Furthermore, Kimura and Anderl (2005) demonstrated that  $CV_j = \sqrt{2}APE_j$  when  $R = 2$ . Thus,  $CV_j$  and  $APE_j$  are simply constant multiples of each other when  $R = 2$ . It follows then that  $AAD_j$  and  $SD_j$  are also constant multiples of each other when  $R = 2$ . Kimura and Anderl (2005) noted that the constant factor of  $\sqrt{2}$  is approximately correct when  $R > 2$ .

Traditionally, the measures of precision for individual fish have been averaged across all  $n$  fish in the sample to develop a single metric of precision. These summaries are shown below with the first “A” in each acronym meaning “average” (e.g., “average average absolute deviation” or “average coefficient of variation”).

$$AAAD = \frac{\sum_{j=1}^n AAD_j}{n} = \frac{\sum_{j=1}^n \sum_{i=1}^R |x_{ij} - \bar{x}_j|}{nR}$$

$$ASD = \frac{\sum_{j=1}^n SD_j}{n} = \frac{\sum_{j=1}^n s_j}{n}$$

$$AAPE = \frac{\sum_{j=1}^n APE_j}{n} = \frac{\sum_{j=1}^n \sum_{i=1}^R \frac{|x_{ij} - \bar{x}_j|}{\bar{x}_j}}{nR}$$

$$ACV = \frac{\sum_{j=1}^n CV_j}{n} = \frac{\sum_{j=1}^n \frac{SD_j}{\bar{x}_j}}{n}$$

Note that the  $APE_j$  and  $CV_j$  and thus the  $AAPE$  and  $ACV$  are usually multiplied by 100 and treated as a percentage.

Kimura and Anderl (2005) suggested, based on distributional theory, that the divisor for  $APE_j$  should be the median age estimate ( $\tilde{x}_j$ ) rather than the mean age estimate. If this same argument is extended to the  $CV_j$  then this gives rise to two new precision metrics

$$APE2_j = \frac{AAD_j}{\tilde{x}_j}$$

$$CV2_j = \frac{SD_j}{\tilde{x}_j}$$

These two metrics can be averaged across all fish to compute the overall measures of precision  $AAPE2$  and  $ACV2$ . For  $R = 2$ ,  $\bar{x}_j = \tilde{x}_j$  and there will be no difference between  $AAPE$  and  $AAPE2$  or  $ACV$  and  $ACV2$ .

However, when  $R > 2$ , then using the median may remove the influence of an age estimate from a reader that is an outlier in comparison to the other readers.

Finally, Chang (1982) suggested the “Index of Precision”  $D$  as

$$D_j = \frac{CV_j}{\sqrt{R}}$$

However, note that when  $R = 2$  then  $D_j = APE_j$  because  $CV_j = \sqrt{2}APE_j$  when  $R = 2$ . This metric will be labelled as  $AD$  when averaged across all  $n$  fish.

For example, the metrics for individual fish may look like the following.

otolith1	otolith2	mean	median	mode	SD	CV	CV2	AAD	APE	APE2	AD
3	3	3.0	3.0	3	0.000	0.000	0.000	0.0	0.000	0.000	0.000
1	1	1.0	1.0	1	0.000	0.000	0.000	0.0	0.000	0.000	0.000
7	7	7.0	7.0	7	0.000	0.000	0.000	0.0	0.000	0.000	0.000
2	3	2.5	2.5	NA	0.707	28.284	28.284	0.5	20.000	20.000	20.000
4	4	4.0	4.0	4	0.000	0.000	0.000	0.0	0.000	0.000	0.000
17	18	17.5	17.5	NA	0.707	4.041	4.041	0.5	2.857	2.857	2.857
10	11	10.5	10.5	NA	0.707	6.734	6.734	0.5	4.762	4.762	4.762
12	11	11.5	11.5	NA	0.707	6.149	6.149	0.5	4.348	4.348	4.348

and a summary across all  $n$  fish may look like the following (note that  $ACV2$ ,  $APE2$ , and  $AD$  are not shown because  $R = 2$ ).

n	validn	R	PercAgree	ASD	ACV	AAAD	AAPE
151	151	2	62.252	0.309	4.719	0.219	3.337

## Problem Statement

Precision is (anecdotally ... see below) most often measured by the  $ACV$ . The  $ACV$  is designed to address the issue that differences among age estimates of say one year for young fish is more important than the same one year difference for old fish. In other words, the measure of precision should be adjusted for the (estimated) age of the fish. However, two possible issues arise with the use of  $ACV$ . First, interpretation is made more difficult. For example, an  $ACV$  of 10% is more difficult to interpret than say an  $ASD$  of 1.2. In the latter, 1.2 can be interpreted as ages differing by 1.2 years on average. However, the analyst likely will convert the  $ACV$  to actual age differences for interpretation (e.g., “an  $ACV$  of 10% means differences of 1 year for age 10 fish but 2 years for age 20 fish”). Second, the  $ACV$  is meant to remove the effect of age on the measure of precision. However, it is possible that  $ACV$  is still related to the age of the fish. If  $ACV$  is related to the age of the fish then the practice of reporting a single  $ACV$  value (averaged across all fish and, thus, all ages) is misleading. As suggested by Bauerlien et al. (2018), the analyst should assess the relationship between any of the precision metrics and age before summarizing with a single value.

Questions that we wish to consider with this research are:

1. Which precision metric is most commonly used in the literature?
2. What is the (statistical) distribution of precision metrics found in the literature?
3. How often have authors tested if their chosen precision metric was related to age or not?
4. How often are precision metrics (especially  $ACV$ ) related to age?
5. What kind of relationship (linear, quadratic, or more complicated) between precision metrics and age is

most common?

6. What are the relationships between the various precision metrics?
7. Do other metrics of precision exist and are they useful to fisheries biologists?
8. What advice should be given to fisheries biologists when choosing a precision metric?

## Research Direction

### Literature Review

We will attempt to review as many studies as possible that examined the precision among readers or among multiple readings by the same reader of fish calcified structures. We will record (at least) the following information from each study will be recorded in a flat database (e.g., spreadsheet).

- Study ID (Author name(s) appended with publication year ... e.g., Lepaketal\_2017)
- Species name
- Study site
- Publication year
- Structure (e.g., scales, otoliths, finrays, vetebrae, cleithra)
- Structure2 (e.g., dorsal, pectoral, pelvic)
- Process (e.g., whole, sectioned, crackburn)
- Number of readers ( $R$ )
- Number of samples ( $n$ )
- Age range (minimum and maximum age among all readers)
- Percent agreement (exact) ... NA if not reported
- AAPE ... NA if not reported
- ACV ... NA if not reported
- ASD ... NA if not reported
- AAAD ... NA if not reported
- ACV2 ... NA if not reported
- AAPE2 ... NA if not reported
- Whether (or not) a relationship between a precision metric and age was examined (Y/N)
- If a relationship with age was examined, what was that relationship (none, linear, quadratic, other)

Citations will be stored in the FishAgePrecision Group Zotero database. The flat database will be stored in a Github repository. Results from the flat database extracted from the publications will be used to answer the first three questions in the Problem Statement.

### Re-Analysis of Past Studies

We will attempt to obtain raw data on multiple age estimates of the same fish from as many sources as possible. We will the (re)analyze these data to record the following items.

- Study ID (Author name(s) appended with publication year ... e.g., Lepaketal\_2017)
- Species name
- Study site
- Structure (e.g., scales, otoliths, finrays, vetebrae, cleithra)
- Structure2 (e.g., dorsal, pectoral, pelvic)
- Process (e.g., whole, sectioned, crackburn)
- Number of readers ( $R$ )
- Number of samples ( $n$ )
- Age range (minimum, 10th percentile, 90th percentile, and maximum mean age among all readers)
- Percent agreement (exact)
- AAPE

- ACV
- ASD
- AAAD
- ACV2
- AAPE2
- AD
- Type of relationship between precision metrics and age
  - Fit linear and quadratic models with and without modeling heteroscedasticity (i.e., generalized least-squares models)
    - \* Determine if heteroscedasticity needed to be modeled
    - \* Determine if the quadratic term was needed
    - \* Determine if a linear relationship existed

Raw data files will be stored as CSV files with a corresponding meta-data text file (an example has been created for Lepak\_etal\_2017). The data will be initially processed with `agePrecision()` from the `FSA` package. Information above will be extracted from this result using `precisionData()` written for this project. The resulting object will be saved as an RDS file for post-processing to produce a `data.frame` (and ultimately a flat database) of the items above.

## Simulation Study

XXX NEED TO THINK ABOUT THIS MORE XXX

## References

- 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 *Esox Niger*. North American Journal of Fisheries Management 38.
- 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.
- 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.
- 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.