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 readings by the same reader, or among calcified structures. 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 (Yule et al. 2008; Hamel et al. 2016; Tyszko and Pritt 2017).

Accuracy, or the ability to estimate the true age of a fish, is related to validation which was extensively discussed by Campana (2001) and won't be discussed further here. Precision is 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 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 ith age estimate for the jth 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 four common measures of precision for an individual fish (say fish j) – average absolute deviation (AAD),[1] standard deviation (SD), average percentage error (APE),[2] and coefficient of variation (CV)[3] – as defined below:

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

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

---

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

[2]Traditionally, APE has been used for the average of the APE_j here. In other words, the traditional APE is equivalent to our AAPE.

[3]Traditionally, CV has been used for the average of the CV_j here. In other words, the traditional CV is equivalent to our ACV. This follows Ogle (2016).

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

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

where $\bar{x}_j$ is the mean of the R age estimates for the jth fish. 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 absolution deviation" or "average coefficient of variation").

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

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

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

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

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) suggestd, 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

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

Note, however, if $R = 2$ then $D_j = APE_j$.[4] This metric will be labelled as AD when averaged across all n fish.

For example, the metrics for individual fish may look like those in the following table:

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	–	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	–	0.707	4.041	4.041	0.5	2.857	2.857	2.857
10	11	10.5	10.5	–	0.707	6.734	6.734	0.5	4.762	4.762	4.762
12	11	11.5	11.5	–	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 table:[5]

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

---

[4]Because, when $R = 2$, $CV_j = \sqrt{2}APE_j$ and the $\sqrt{2}$ in the numerator and denominator will cancel each other.
[5]Note that $AACV2$, $AAPE2$, and AD are not shown because they are equal to ACV, APE, and APE when $R = 2$.

Problem Statement

Prior to Campana (2001), ACV was used slightly more than APE as a precision metric. Since then, it appears that precision is most often measured by the ACV, largely based on the recommendation of Campana (2001). 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 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. The analyst will likely convert the ACV to a 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"). However, the ASD of 1.2 can be directly interpreted as ages differing by 1.2 years on average. 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. This implies that ACV may not be the universally best precision metric.

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. What are the relationships between the various precision metrics?
5. How often are precision metrics (especially ACV) related to age?
6. What kind of relationship (linear, quadratic, or more complicated) between precision metrics and age is most common?
7. Are any of the above affected by structure type, fish species, longevity, sample size, number of readings, etc.
8. Do other metrics of precision exist and are they useful to fisheries biologists?
9. 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 of ages estimated from calcified structures within or between readers. The following general information about each study will be recorded in a flat database (e.g., spreadsheet):

Field	Description
studyID	Unique identifier for each study/publication (created by Zotero)
pubyear	Publication year
country	Country where study was conducted
studysite	More specific description of the study site
marine	Whether (TRUE) or not (FALSE) the fish are a marine species
minedrefs	Whether (TRUE) or not (FALSE) we examined the references to find other precision metric papers
note	Other notes

In addition, the following specific information will be recorded in a second flat database:

Field	Description
studyID	Unique identifier for each study/publication (created by Zotero)
species	Species of fish
structure	Broad structure description (e.g., cleithra, finrays, otoliths, scales, vertebrae)
structure2	Specific structure description (e.g., sagittae, dorsal, pectoral, pelvice)
process	Description of how the structure was processed (e.g., whole, sectioned, crack-and-burn)
type	Type of calculation (between for between-readers, within for within-reader)
R	Number of readings
n	Number of fish read
agemin	Minimum observed age (or mean age)
agemax	Maximum observed age (or mean age)
PA0	Percent agreement (perfect) metric
PA1	Percent agreement (within 1 year) metric
PAother	Whether (TRUE) or not (FALSE) other percent agreement metrics were reported
AAPE	Average average percent error metric
ACV	Average coefficient of variation metric
ASD	Average standard deviation metric
AAAD	Average average absolute deviation metric
ACV2	Average coefficient of variation (using the median) metric
AAPE2	Average average percent error (using the median) metric
AD	Average index of precision metric
checkbias	Whether (TRUE) or not (FALSE) checked for systematic bias among readers
biasmethod	Method used to test for systematic bias among readers (e.g., agebiasplot, chisquare, pairedt)
checkrelage	Whether (TRUE) or not (FALSE) checked for a relationship between the precision metric and age.
typerelage	Type of relationship between the precision metric and age (e.g., none, linear, quadratic, other)
notes	Other notes.

This database will have a separate row for each unique combination of species, general structure, specific structure, process, and type of calculation. For example, if one paper examined between- and within-reader precision for whole otoliths, sectioned otoliths, pectoral spines, and cleithra for three species of fish, then this study would populate 24 (two types of calculation × four specific structures × three species) rows in this flat database (each with the same study ID).

Citations will be stored in the FishAgePrecision Group Zotero database. The flat database will be stored in a Github repository. This information will be used to answer the first three and the seventh 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. General information about each study will be stored in the same flat database as for the literature review. In addition, we will analyze each data set to record two types of information. First, we will record the following information about mean ages and precision metrics for each data set.

Field	Description
studyID	Unique identifier for each study/publication (created by Zotero)
species	Species of fish
structure	Broad structure description (e.g., cleithra, finrays, otoliths, scales, vertebrae)
structure2	Specific structure description (e.g., sagittae, dorsal, pectoral, pelvice)
process	Description of how the structure was processed (e.g., whole, sectioned, crack-and-burn)
type	Type of calculation (between for between-readers, within for within-reader)
R	Number of readings
n	Number of fish read
validn	Number of fish in the analysis
agemin	Minimum observed mean age
age10	10th percentile of observed mean ages
age90	90th percentile of observed mean ages
agemax	Maximum observed mean age
PA0	Percent agreement (perfect) metric
AAPE	Average average percent error metric
ACV	Average coefficient of variation metric
ASD	Average standard deviation metric
AAAD	Average average absolute deviation metric
ACV2	Average coefficient of variation (using the median) metric
AAPE2	Average average percent error (using the median) metric
AD	Average index of precision metric

Additionally, we will examine the relationship between precision metrics and mean age for reach data set. We will fit the following four generalized least-squares (GLS) models for the relationship of a precision metric on mean age:

1. Linear model assuming homoscedasticity for the precision metric relative to mean age.
2. Linear model assuming heteroscedasticity for the precision metric relative to mean age follows a power function (i.e., either increasing or decreasing variance in the precision metric with increasing mean age).
3. Quadratic model assuming homoscedasticity for the precision metric relative to mean age.
4. Quadratic model assuming heteroscedasticity for the precision metric relative to mean age follows a power function.

In situations with small sample sizes ($n < $ XXX) the two models with heteroscedasticity may not be able to be fit. Following fitting these models, we will determine if heteroscedasticity is present by comparing models 1 and 2 and models 3 and 4. We will then determine if the quadratic term is necessary by either comparing models 1 and 3 or models 2 and 4. Finally, if a quadratic term is not necessary then we will test for a significant linear relationship between the precision metric and mean age by testing that the slope in the linear model is zero or not. These set of tests will only be conducted with the SD_j and CV_j metrics

given the tight and linear relationships between AAD_j and SD_j and between APE_j and CV_j. The following information will be recorded from these tests.

Field	Description
studyID	Unique identifier for each study/publication (created by Zotero)
species	Species of fish
structure	Broad structure description (e.g., cleithra, finrays, otoliths, scales, vertebrae)
structure2	Specific structure description (e.g., sagittae, dorsal, pectoral, pelvice)
process	Description of how the structure was processed (e.g., whole, sectioned, crack-and-burn)
type	Type of calculation (between for between-readers, within for within-reader)
var	Precision metric variable being examined (SD or CV)
HetInLM	p-value testing for heteroscedasticity in the linear model
HetInQM	p-value testing for heteroscedasticity in the quadratic model
QMNoHet	p-value testing for quadratic term in homoscedastic models
QMHet	p-value testing for quadratic term in heteroscedastic models
SlopeLMNoHet	p-value testing significance of slope in homoscedastic linear model
SlopeLMHet	p-value testing significance of slope in heteroscedastic linear model
Hetero	Summary of presence of heteroscedasticity (neither, linear, quadratic, both)
Quad	Summary of quadratic term importance (neither, no hetero, hetero, both)
Slope	Summary of slope significance (neither, no hetero, hetero, both)

Raw data files will be stored as CSV files with a corresponding meta-data text file, which will be named after the `studyID` variable (an example has been created for lepak_age_2017.CSV and lepak_age_2017_META.txt). Each data file will be initially processed with `agePrecision()` from the `FSA` package. Specific information (see below) will be extracted from this result using `precisionData()` written for this project and the resulting object will be saved as an RDS file for post-processing. All of the resulting RDS files will then be read and compiled to produce two data.frames that contain the precision metric summary information and the results of the tests for the relationship between the precision metrics and mean age. Results in these two data.frames will be used to answer the fourth through seventh questions in the Problem Statement.

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.

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.

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.

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.