**Probable impacts of ageing error precision and bias on a stock assessment of gray triggerfish (*Balistes capriscus*)**

Derek Chamberlin1, Nicholas Fisch2, Robert Ahrens3, Kelli F. Johnson4, and William F. Patterson5

1National Marine Fisheries Service, Alaska Fisheries Science Center, 7600 Sand Point Way N.E., Building 4, Seattle, WA 98115, USA

2Pacific Biological Station, Fisheries and Oceans Canada, 3190 Hammond Bay Road,Nanaimo, British Columbia V9T 6N7, Canada

3National Marine Fisheries Service, Pacific Islands Fisheries Science Center, 1845 Wasp Blvd., Building 176, Honolulu, HI 96818, USA

4National Marine Fisheries Service, Office of Science and Technology, 1315 East-West Highway, Silver Spring, MD 20910, USA

5University of Florida, School of Forest, Fisheries, and Geomatics Sciences, 7922 NW 71st Street, Gainesville, FL 32653, USA

Abstract

Introduction

Integrated fishery stock assessment models are critical tools in modern fisheries science which use a statistical framework to estimate stock status and biological reference points for management advice. In the United States, the majority of these assessments rely on age-structured models, which use age composition data to estimate key parameters such as recruitment and natural mortality (CITATIONS). Accurate and precise age estimates are therefore essential; errors in ageing can propagate through models, introducing bias or increasing uncertainty in management benchmarks derived from assessment outputs.

Statistical catch-at-age (SCAA) models use time series of age composition data to track cohorts across years, while integrating additional information such as abundance indices, fleet-specific catch and discards, and length composition data (Deriso et al., 1985; Fournier and Archibald, 1982), all of which are fit statistically in a population dynamics model with process and observation error (Maunder and Punt, 2013). The use of age simplifies the transition of fish from one time step (e.g., year class) to another, which can lead to a reduction in parameter confounding relative to catch-at-size models (Fisch et al., 2019). However, when age data are biased due to, for instance, misinterpretation of ageing structures, assessment outputs such as estimates of productivity, abundance, and reference points can be systematically biased (Lai and Gunderson, 1987; Reeves, 2003).

Gray triggerfish, *Balistes capriscus*, is one such species where ageing error is a potential stock assessment concern (Chamberlin et al., 2024; Potts et al., 2023). Gray triggerfish age has historically been estimated by counting translucent zones in dorsal spine sections (Allman et al., 2016; Hood and Johnson, 1997; Kolmos et al., 2013; Potts, 2014; SEDAR 43, 2015). Recently researchers demonstrated that the historical dorsal spine protocol of Kolmos et al. (2013) underestimates age, while otolith opaque zone counts from whole sagittal otoliths, as well as an updated dorsal spine ageing protocol produce accurate age estimates (Chamberlin et al., 2024; Potts et al., 2023).

The northern Gulf of Mexico, also known as the Gulf of America and hereafter referred to as “the Gulf”, gray triggerfish stock is estimated to have historically experienced overfishing and was assessed to be overfished in the early 1990s (SEDAR 9, 2006; SEDAR 43, 2015). Assessments as late as 2006 indicated fishing mortality far exceeded the fishing mortality that produces the maximum sustainable yield (FMSY) (Porch, 2001; SEDAR 9, 2006; Valle et al., 2001). As a result, a rebuilding plan was developed in 2008 that reduced commercial and recreational catch limits and instituted a closed season in both sectors to reduce fishing mortality and allow the Gulf gray triggerfish stock to recover (amendment 30A citation). Despite these management measures, the stock did not recover at expected rates and was not rebuilt by the end of the rebuilding plan timeline (Crabtree, 2015; Gregory, 2015; SEDAR 43, 2015) A second rebuilding plan was subsequently implemented by the Gulf Council in 2017 (Amendment 46 citation). Given the historical error in age estimates produced from dorsal spine translucent zone counts (Chamberlin et al., 2024; Potts et al., 2023), it is plausible past stock assessments overestimated stock productivity, leading to overly optimistic stock recovery projections (Lai and Gunderson, 1987; Reeves, 2003; Tyler et al., 1989).

Simulation analysis enables the examination of the effects of ageing error on stock assessment benchmarks (Lai and Gunderson, 1987; Reeves, 2003), and is a useful approach to explore the impacts of parameter bias in stock assessment models (Monnahan et al., 2016; Stawitz et al., 2019). This approach uses an operating model to simulate a hypothetical or real-world fishery, a sampling model to simulate data collection from the fishery, and these data are then used in an estimation model to estimate performance measures and evaluate the effects of uncertainty in data inputs on stock assessment benchmarks. In simulation analysis, the true population parameters are known, thus one can evaluate the performance of the stock assessment model relative to known values. This allows one to define and control for assumptions inherent in the assessment framework. This approach has been used to evaluate the effects of sampling protocols, alternative methods of mortality estimation, and the relative utility of age- versus size-structured assessment models (Fisch et al., 2019; Gwinn et al., 2010; Lee et al., 2011; Punt et al., 2017), thus is well-suited to examine the effects of ageing error on an integrated statistical catch-at-age stock assessment.

The objectives of this study were two-fold; 1) examine the effects of different forms of ageing bias on a simplified age-structured assessment framework reflecting the life history of Gulf gray triggerfish, and 2) examine the effects of empirical ageing error on the most recent Gulf gray triggerfish production stock assessment (SEDAR 43, 2015). The dual approach was chosen as the most recent production assessment is a complex stock assessment with catch and age composition data from five fleets, nine indices of abundance (six fishery-dependent and three fishery-independent), and 37 estimated parameters. This increases the potential for confounding variables and interactions that are difficult to parse within a complex simulation framework. Thus, pairing the more complex assessment with a simpler framework allows us to reduced those confounding variables and interactions and examine trends that can be generalized beyond the Gulf gray triggerfish stock. Results are discussed in the context of the effects of different types of ageing bias on assessment estimates and the effects of historical bias on the gray triggerfish assessment and management, including the contribution of ageing error to the lack of recovery in the Gulf gray triggerfish stock.

Methods

Population simulation operating model

The population simulation operating model was based on the most recent assessment of Gulf of Mexico Gray Triggerfish (SEDAR 2015). The parameter values and equations governing the model are given in Tables 1 and 2, respectively. The population model runs for 94 years, with a fishing time series beginning in year 26 (years 26-94 experience fishing). The fishing time series aggregated fishing intensity values across fishing fleets from the assessment, with the shrimp fishery removed. For fishery selectivity, we searched for parameter values that roughly matched the dome-shaped selectivity from each fleet in the assessment (Figure 1, this MS; Figure 3.2.32 – SEDAR 2015).

One hundred simulation iterations were run for the population model specification, with stochasticity included using log-normal draws for recruitment deviations about a Beverton-Holt stock recruitment relationship in each year. A sampling model then sampled from that population and fishery exploitation model to simulate the data collection process.

Sampling model

Annual landings were simulated using a log-normal distribution with a standard deviation (SD) of 0.05. A fishery index was simulated using a lognormal distribution with an SD of 0.25 (Table 3). The catch age-composition available for select years throughout the time series was simulated using a multinomial distribution. Age composition data was simulated from years 26, 36, 46, 51, 56, 61, 66, 71, and consistently from year 76 to 94. The number of samples began at *n*=30 in year 26 of the model (first year with fishing), increasing by 10 each decade until year 51, and then increasing by 10 every 5 years until year 76 where it was fixed at *n*=100 for the rest of the time series. Once the true observed age composition was simulated, ageing error was included by drawing from a multinomial distribution with probabilities defined using an ageing error matrix **X** (the probabilities any specimen with true age *a* will be read as age 0, 1, …, *A*).

Ageing Error Matrix

Ageing error was simulated from an ageing error matrix derived from a normal distribution with different scenarios for bias and precision, or standard deviations of at age, (). The probability that an individual age  will be coded age  was calculated using the cumulative distribution function of the standard normal ()

Ageing error was then simulated for the observed catch at age data by drawing from a multinomial distribution with the sample size equal to the number of fish true age in the sample and the probability vector as .

Ageing Error Scenarios

We implemented four different scenarios for bias and imprecision for age reading (Figure 2).

1. No ageing error.
2. Constant bias with age.
3. Linear increase in bias with age.
4. Curvilinear bias relationship.

The constant bias scenario (Scenario 2) mirrored the systematic under-ageing scenario reported in (Reeves, 2003). Constant under-ageing is a common phenomenon when the first annulus is mis-identified (CITATION). The linear scenario (Scenario 3) bias slope parameter was set at 0.25 to mirror the maximum bias observed in Scenario 4, allowing for a more direct comparison between bias shapes. The curvilinear scenario (Scenario 4) bias was estimated using the nwfscAgeingError R package, with a curvilinear relationship between observed and predicted age and a linear relationship between standard deviation and predicted age, (Punt et al., 2008; Thorson et al., 2012) comparing paired gray triggerfish otolith and historical dorsal spine-derived age estimates(Chamberlin et al., 2024). The standard deviation-at-age in Scenarios 3 and 4 were based on the standard deviation-at-age estimates from the aforementioned nwfscAgeingError model.

Why you chose what you chose for AE definition

Estimation model

The estimation/assessment models were identical in structure to the population operating model. They began in year 26 of the time series (first year with data). The fixed-effect parameters estimated within the model included the unfished recruitment, the recruitment SD, a natural mortality parameter, four selectivity parameters, catchability of the fishery index, and annual fully selected fishing mortality levels in each year (Table 1). Recruitment deviations in each year and those that make up the initial unfished abundance were treated as random effects. All other parameters were fixed at their true values from the operating model. The assessment models also fit the data using the same likelihoods used to generate data in the sampling model.

Experimental Design

The experimental design was a full factorial where each sampling model scenario was fit by estimation models with each of the four ageing error scenarios specified. This resulted in 16 different model fits (with 100 replicates), four of which were correctly specified.

Model Fitting

Models were fit using Template Model Builder (TMB, Kristensen et al., 2016) with recruitment deviations specified as random effects and the remainder of the fixed effect parameters identified in Table 1. TMB calculates the marginal negative log-likelihood given the fixed effect parameters using the Laplace approximation to integrate over random effects. Fixed effect parameters are estimated via minimizing the marginal negative log-likelihood within the program R using the nlminb function. Random effects and derived quantities are then predicted using empirical Bayes (Kristensen et al., [2016](javascript:;)). Starting parameter values for each simulation iteration were chosen by sampling from a uniform distribution with the bounds specified as 50% below and 50% above the true parameter value. If a model failed to converge according to a non-positive definite hessian matrix, new starting parameters were drawn and the model fit again. This process was repeated a maximum of 10 times. Standard errors of parameter estimates were extracted in addition to those of derived quantities using the generalized delta method built into TMB.

Performance Metrics

We evaluated model performance by calculating three different performance metrics: relative error in annual spawning stock biomass, error in the terminal F-ratio (F/Fmsy), and error in the terminal year of B-ratio (B/Bmsy). Relative error was calculated as

And the error for the ratios was simply the estimated value – true. F-ratio was defined as the fishing mortality estimated in the terminal year of the model divided by the estimated fishing mortality at maximum sustainable yield. Similarly, B-ratio was defined as the terminal year estimated biomass divided by the biomass at maximum sustainable yield estimated by the model.

We also, as a diagnostic, present the relative error in natural mortality, unfished recruitment, fishing mortality, Fmsy, and Bmsy in the supplemental file.

Results

Convergence

Model convergence rates varied between the scenarios, with all correctly specified scenarios resulting in convergence rates greater than 90% (Table 3). The lowest convergence rates occurred with sampling models fit with estimation models that assumed ageing error had a linear bias with age. Outside of the correctly specified scenario for linear bias, the convergence rates were 41, 14, and 19% for EMs fit to sampling models without ageing error, with a constant bias, and with a curvilinear bias. All other scenarios had convergence rates above 70%.

Performance Metrics

The correctly specified scenarios were all approximately unbiased in each performance metric (diagonal panels, Figures 3-4). The one slight exception to this is the linear ageing bias estimation model, which had a slight positive bias in SSB and F-ratio (although this model had very low convergence).

*SSB*

When there was no ageing error in the data, assuming ageing error had a constant bias or a linear bias led to positive biases in annual SSB, although the former became more unbiased through time and the latter less biased (Figure 3). The model that assumed there was a curvilinear bias resulted in very little bias in annual SSB.

When the data had a constant bias in age reading, assuming there was no ageing error or curvilinear error led to negative biases in SSB on the order of 25%. The model that assumed a linear bias in age reading was approximately unbiased although had very low convergence.

When there was a linear bias in age reading, assuming no ageing error or curvilinear error led to a negative bias in SSB at the end on the time series, on the order of ~10%. Estimation models that assumed a constant bias in age reading at age resulted in positive bias in SSB on the order of 10-15%.

When there was a curvilinear bias in age reading, assuming no ageing error in the estimation model led to approximately unbiased SSB at the end of the time series. The estimation models that assumed a constant or a linear bias in age reading led to positive biases in annual SSB on the order of 20-30%.

*F- and B-ratio*

When there was no ageing error in the data, assuming ageing error had a constant bias led to a negatively biased F-ratio and a positively biased B-ratio (~0.25, Figure 4). The opposite was true for the scenario where the assessment model assumed a linear bias in ageing error, with positively biased F-ratio and negatively biased B-ratio (where the F-ratio was substantially biased (>0.50). The model that assumed there was a curvilinear bias resulted in minimal bias in each of the ratios.

When the data had a constant bias in age reading, assuming no ageing error or a curvilinear ageing bias led to a positive bias in F-ratio and a negative bias in B-ratio, although both were minimal (median < 0.1). The estimation models that assumed a linear bias in age reading had very low convergence although of those that did converge, the median F- and B-ratios were in the opposite direction as the other scenarios although with more variability.

When there was a linear bias in age reading, assuming no ageing error, constant bias, or curvilinear bias each led to positive biases in F-ratios, on the orders of approximately 0.25, 0.10, and 0.20, respectively.

When there was a curvilinear bias in age reading, assuming no ageing error or linear bias led to positive biases in F-ratio and negative biases in B-ratio (with substantially more bias for the linear bias EM for the models that converged). When the estimation model assumed a constant bias in age reading, the F- and B-ratios were approximately unbiased.

Discussion

Curvilinear seems to be the least worst outcome.