Title

The effect of length bin width on growth estimation in integrated age-structured stock assessments

Authors

Cole C. Monnahan*[1], Kotaro Ono[2], Sean C. Anderson[2,3], Merrill B. Rudd[2], Allan C. Hicks[4], Felipe Hurtado-Ferro[2], Kelli F. Johnson[2], Peter T. Kuriyama[2], Roberto R. Licandeo[5], Christine C. Stawitz[1], Ian G. Taylor[4], Juan L. Valero[6]

* Corresponding author: monnahc@uw.edu

[1]Quantitative Ecology and Resource Management, Box 352182, University of Washington, Seattle, WA 98195, USA

[2]School of Aquatic and Fishery Sciences, Box 355020, Seattle, WA 98195-5020, USA

[3]School of Resource and Environmental Management, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada

[4]Fishery Resource Analysis and Monitoring Division, Northwest Fisheries Science Center, NOAA, 2725 Montlake Blvd. East, Seattle, WA 98112-2097, USA

[5]Fisheries Centre, Aquatic Ecosystems Research Laboratory (AERL), University of British Columbia, Vancouver, BC, Canada V6T 1Z4

[6]Center for the Advancement of Population Assessment Methodology, Scripps Institution of Oceanography, La Jolla, CA 92037, USA

Abstract

Analysts conducting stock assessments using integrated, age-structured models must discretize length data into a limited number of bins (data bins). Furthermore, some modeling frameworks also allow users to specify a distinct structure for how lengths of fish are represented in the model (model bins). The effect of choices regarding the number and width of these bins on model output is unclear, and these choices are made inconsistently in assessments across regions and species. Here, we used the Stock Synthesis modeling framework, and the ss3sim stock assessment simulation package, to explore the effects of choices about length discretization on stock assessment performance for three fish life-history types and four data cases. We found that, with all other aspects of a model fixed, increasing the model bin width tended to increase estimates of spawning biomass, but this effect depended on the shape of length-based processes (e.g., growth, maturity, and selectivity). Thus, we suggest analysts using model bins wider than 1 cm explore the effect of this decision on derived management quantities. In the context of estimation, there generally was a predictable tradeoff between estimation accuracy and model run time, with finer model and data bins always improving estimation accuracy and model convergence, but increasing run time. In some cases, wider data bins reduced run time (by up to 50%) with little sacrifice in model estimation performance, particularly those using conditional age-at-length data. This study identifies key aspects to consider when binning length, and provides pertinent information for stock assessment best practice guidelines.

Keywords

fisheries stock assessment; simulation testing; somatic growth; Stock Synthesis; ss3sim

1 Introduction

Integrated, age-structured fisheries stock assessment models are complex, powerful, and flexible tools for analyzing the status of a fish stock (Hilborn and Walters, 1992). However, this complexity often requires an analyst to make a variety of subjective biological, statistical, and modeling decisions, the effects of which are often poorly understood (Maunder and Piner, 2015). One such decision is how to discretize fish length measurements into 'bins' for analysis. In reality, growth is a continuous process, yet in assessment models, length data, and processes which depend on length, must be broken into discrete bins. Length bin specification is of central importance when constructing size-structured models (Drouineau et al., 2008), but it also is important in age-structured models because many important biological and fishery processes are a function of length (e.g., growth, maturity, and selectivity). In addition, lengths, which are easier to measure than ages, are a common source of data used to inform estimates for key processes like growth.

Within some types of age-structured stock assessment models, analysts must specify two distinct types of length bins. First are 'data bins,' which specify the resolution of the observed length data (e.g., length compositions). For example, length measurements from a fishery may be recorded to the nearest 1 cm, and bins must span the observed length range (i.e. 10 to 50 cm). Second are 'model bins,' which define the length dynamics within the model. For the same example, the model bins may need to range from 5 to 100 cm to appropriately capture fish in the population that are too small to be selected by the gear, and larger fish which were previously available to fishermen. The choice of data bins is limited by the properties of the observed data, whereas the choice of model bins may be based on prior observations or may be a subjective decision. In many age-structured stock assessment modeling frameworks, the data and model bins match. The common bin width is decided upon based on the bins in which the length measurements are collected, or some aggregation subjectively chosen by the analyst. For cases where the data and model bins do not match, model bins need to be mapped to the data bins (typically via aggregation) to calculate the likelihood of the expected proportions at length, conditional upon the observed data. Distinct bin types are possible for any custom-built model, as well as the widely-used age-structured population modeling framework Stock Synthesis (SS; Methot and Wetzel, 2013). Therefore, depending on the modeling framework, an analyst must decide the minimum length, maximum length, and bin width (together the 'bin structure'), and whether to have distinct model and data bins.

The choice of bins represents a tradeoff between model performance and accuracy. Increasingly fine model bins characterize length-based processes at a finer scale, but also increase computational requirements. Finer bins are therefore expected to increase accuracy, but may increase model run time (i.e., slower estimation). Conversely, increased bin width may reduce the accuracy of model estimates, but reduce model run time. Reducing model run time may free up time for analysts to conduct sensitivity tests or perform Bayesian analyses (e.g., Stewart et al., 2013). However, guidelines on best practices for binning strategies (i.e., setting the width and thus the number of bins) to balance this tradeoff are not readily available. Consequently, decisions are typically *ad hoc* and likely based on factors such as preferences from personal or colleague experience. A non-exhaustive survey of stock assessments from the U.S. West Coast, Gulf of Mexico, South Atlantic, mid-Atlantic, and Australia found a wide variety of bin widths were used in assessments, with little relation to maximum length or other life-history characteristics (Fig. 1).

Szuwalski (2015) used simulation to explore the effects of increasing bin width on the precision and run time of a size-structured stock assessment model. He found biases in mature biomass and tradeoffs between precision, model stability, and run time, and recommended setting the bin width based on the goal of the analysis. Simulation testing has also been used to study age-structured stock assessment models for a wide range of topics, such as selectivity (Crone and Valero, 2014), steepness of the stock-recruit relationship (Conn et al., 2010), the value of data (Ono et al., 2014), retrospective patterns (Hurtado-Ferro et al., 2014), and time-varying natural mortality (Johnson et al., 2014). Here, we explore tradeoffs between run time and accuracy of growth and management quantities with increasing length bin widths for three life histories and two types of data (age vs. conditional-age-at length) in an age-structured stock assessment model.

2 Materials and Methods

2.1 Overview

We generated true population and fishery dynamics from an operating model (OM), and then ran stock assessments via an estimation model (EM). The OMs and EMs were parameterized from actual assessments and modified to generate and assess simplified, but realistic, dynamics. Process and sampling error were added to the OM values to simulate variable dynamics and data collection. During model development, we verified that under base conditions (i.e., the same model structure between OM and EM and unbiased data sampling) the EM parameter estimates were unbiased. This ensured any observed bias was caused by the hypothesis under investigation. We then varied the data and model length bin structures in the EMs and investigated how these differences affected the precision and bias of estimated growth and management quantities.

We conducted our analysis in R (version 3.2.2; R Core Team, 2015) using the stock assessment simulation framework ss3sim (Anderson et al., 2014a; Anderson et al., 2014b), which uses SS (version 3.24o; Methot and Wetzel, 2013) to both generate data and run the assessment. Cross-testing different modeling frameworks for simulation and estimation can provide useful insights (Deroba et al., 2014); however, we chose to use the same model framework (self-testing) because it allowed us to isolate the effects of the specific properties being varied (e.g., bin width). Following the reproducible and transparent philosophy of ss3sim, this simulation, consisting of the model configurations, results, and code to run and process the simulations, is available online and the analysis is reproducible using freely available tools (see Appendix A for details).

2.2 Model Configurations

2.2.1 Biological assumptions

We chose three archetype stocks to represent a broad range of life histories (Table 1). The 'cod' model setup represents a life history with a medium life span, moderate growth rate, and low recruitment variability (derived from North Sea cod; *Gadus morhua*; R. Methot, NMFS, NOAA, pers. comm.). The 'rockfish' model setup represents a life history with a long lifespan, low growth rate, and low recruitment variability (derived from yelloweye rockfish; *Sebastes ruberrimus*; Taylor and Wetzel, 2011). Finally, the 'flatfish' model setup represents a life history with a medium lifespan, high growth rate, and moderate recruitment variability (derived from yellowtail flounder; *Limanda ferruginea*; R. Methot, pers. comm.). The OMs were parameterized with biological parameters estimated in their respective stock assessments (Table 2), whereas the dynamics of the simulated fisheries were simplified from their respective assessments (e.g., single sex and area) and standardized between species (see section 2.2.2).

We modeled somatic growth for both the OM and EM using the specialized von Bertalanffy growth function (Schnute, 1981), as parameterized in SS:

$$L(a) = L_{\infty} + \left(L_{\min} - L_{\infty}\right)e^{-k\left(a-a_1\right)},$$

where *L(a)* is the mean length of fish at age *a,* and a_1 is a young age that is well-represented in the data. SS also linearly interpolates the length of fish younger than a_1, adjusts mean length within the plus group, and normally distributes lengths at each age around the mean length-at-age (see equations A.1.3-A.1.14 in Methot and Wetzel, 2013). In this study, all five growth parameters were estimated in each EM: mean length at minimum age (L_{min}), mean length at maximum age (parameterized to be L_{∞}), Brody growth coefficient (k), and coefficients of variation for young and old fish (CV_{young} and CV_{old}). We considered only constant growth parameters, assuming no time variation, growth morphs, or platoons (i.e., growth classes; Goodyear, 1984).

We assumed a Beverton-Holt stock-recruitment relationship, with steepness fixed at the true value in the EM, but estimated unfished recruits (R_0). We also assumed independent and identically distributed annual recruitment deviations, with their magnitude based on the recruitment error estimated in the original models (Table 1), as the only source of process error. Natural mortality, the length-weight relationship, and maturity curves were all fixed in the EM at their true values (Table 2).

We applied a procedure to correct for bias in estimated recruitment deviations that can arise in a penalized likelihood framework (Methot and Taylor, 2011). This iterative procedure is impractical to conduct for every model replicate in a simulation study, because it requires estimating and inverting a Hessian matrix, and then rerunning the model. Therefore, we used the same parameters for all replicates of a given scenario (unique combination of life history, data case, and bin-width case; see below). For each scenario we estimated bias adjustment parameters using 10 and 20 replicates in the data-rich and data-limited scenarios, respectively, and then used the average of those parameters for all replicates of that scenario.

2.2.2 Fishery assumptions

Model configurations included one fishery and one survey. We ran the OMs without fishing for 25 years as a 'burn-in' period for the cod and flatfish models, and used a 100 year burn-in period for the rockfish model due to its long lifespan. After the burn-in period, fishing was simulated for 75 years. We used an increasing and then decreasing exploitation pattern (i.e. 'two-way trip'; Magnusson and Hilborn, 2007), specified in terms of instantaneous F, rather than catch. The exploitation pattern was calculated as a function of the maximum sustainable yield (MSY) value for a given life history (Fig. 2d). Specifically, F increased linearly for 40 years to F_{high}, the value which led to catch at equilibrium of 0.9MSY (such that $F_{high} > F_{MSY}$), and then decreased linearly for 35 years to F_{low}, the value which leads to equilibrium catch of 0.9MSY (and $F_{low} < F_{MSY}$). Previous studies using the ss3sim framework found little benefit in evaluating additional fishing patterns (Hurtado-Ferro et al., 2014; Johnson et al., 2014; Ono et al., 2014).

The fishery selectivity curve matched the maturity curve (logistic), and was length based, time invariant, and estimated in the EM. Survey selectivity was shifted to the left of the fishery curve (i.e. smaller fish were selected in the survey), such that the length at which half of fish were selected in the survey was 80% of what it was in the fishery. The survey catchability parameter (q) was fixed at 1 in the OM, but estimated in the EM. More details on the model configurations can be found in the model configuration files (Appendix A) and Table 2.

2.2.3 Data quantity and quality

We used four types of data: (1) an index of abundance from the survey, (2) length compositions from the survey and fishery, and either (3) age compositions or (4) conditional age-at-length (CAAL) compositions from the survey and fishery. CAAL composition data are created from paired age and length observations, and represent the age structure within a given length data bin (e.g., He et al., 2015). Using CAAL data, instead of both age and length compositions of the same sampled fish, is preferable because it avoids including the same data twice. CAAL data are expected to be more informative about growth than marginal age composition data (He et al., 2015; Methot, 2015), although to date few studies have examined this data type.

Abundance indices were generated using a lognormal distribution. Age and length compositions were assumed to be independent and generated from a multinomial distribution. CAAL data were generated for a given fleet and year using the following procedure. First, we sampled from the expected length distribution with sample size N_{length} to get the observed number of fish in each length bin $N_{length,bin}$ (the length compositions). Second, we assumed all fish were aged, and for each length bin took a multinomial sample of size $N_{length,bin}$ with probabilities set to the true distribution of ages, given the length bin. We repeated this procedure across all length bins, fleets, and years to construct an observed CAAL matrix. The CAAL data were inherently tied to the length compositions (e.g. as if a trip measured lengths and ages for all fish), in contrast to the age compositions which were generated independently of the length data (e.g. as if one trip measured only ages and a second only lengths). Index, length, age, and CAAL data were all unbiased, and the level of observation error was controlled by the sample sizes for compositions, or the standard deviation (in log space) for the index of abundance. We did not include ageing error in the models or data sampling, and EM samples sizes were fixed at the true effective sample size.

We varied the quantity (number of years, number of samples) and type of age data (age composition or CAAL) to test the impact of binning across four hypothetical data cases (Fig. 2e). Since the length of burn-in periods differed between life histories, we report years since the start of fishing. Our ‘rich’ case included fishery compositions (length, and age or CAAL) of 125 fish sampled in years 10, 20 to 45 every 5 years, and then annually from 46 to 75. The survey operated every other year from year 50 to 75, with compositions samples of 500 fish and standard deviation of the log of the index of 0.2. Our 'limited' data case had fishery compositions of 20 fish sampled in year 60, and then annually from 65 to 75. The survey sampled every other year from 68 to 75, with compositions of 20 fish and standard deviation of the log of the index of 0.2. Both of these data cases included an index of abundance from the survey and length compositions, but we also varied whether there were age compositions (assumed independent of the length data) or CAAL data (dependent on the length data) to create a total of four data cases.

2.3 Binning methods and strategies

The OM and EMs for each life history were set up with identical model and data bin ranges (minimum and maximum lengths; Table 1). In all cases, the OM used 1 cm model and data bin widths, so that the models' internal calculations and expected values mimicked a data collection procedure where lengths were grouped in 1 cm bins. We set the minimum bin well below the smallest observed fish, and we set the maximum bin large enough to contain more than 99.5% of fish if the population were in an unfished state, effectively eliminating the plus group.

2.3.1 Structural impact of model bin widths

Some properties and derived quantities of a model inherently depend on the model length bin structure used. In contrast, the data bin structure has no impact on these quantities, as it is only used in the calculation of the likelihood. Therefore, before testing the impact of data bin width in the estimation context, we investigated how the true *MSY*, annual spawning stock biomass (*SSB*), and recruitment changed with increasing model length bin widths. To explore this, we ran the OMs for model bin widths of 1-20 cm without recruitment (process) variation, and estimation turned off (i.e., growth and other parameters fixed). Thus any differences in *MSY*, *SSB*, or recruitment were caused only by differences in the model bin structure. These results were used to help interpret the estimation accuracy in subsequent sections. We used the quantities from the 1 cm case (the highest resolution) as a baseline against which to compare wider bins used in this section.

2.3.2 Tradeoffs of model speed and accuracy with increasing bin widths

In this section, we explored the impact of data bin width on estimation. We used the same set of commonly used data bin widths (1, 2, 5, 10, and 20 cm) across all three life histories. Although stock assessments rarely use widths greater than 10 cm (Fig. 1), we included a 20 cm width to explore model performance beyond the typical range. The total number of bins for life-history types, for a given bin width, varied due to differences in maximum lengths (Table 3). For example, a 5 cm bin width would lead to many more bins for a large elasmobranch than it would for a small forage fish. An alternative experimental design would be to specify a set of bin width ratios (the ratio of bin width to asymptotic length) to standardize the resolution tested across life history types. However, this led to widths that were not integers or were unlikely to be used in practice (e.g., 13 cm). These bin widths also introduced technical difficulties, because CAAL data bins must align with the model bins, and we sought identical bin ranges to facilitate comparisons of model run time. Therefore, we chose to specify the widths above and note the bin width ratio for each life history (Table 3).

In addition to varying the data bin widths, we also tested two cases for EM model bin widths (the ranges remained fixed). In the first case, we matched the model and data bin width, as if the option of a distinct model bin structure was not available. In the second case, we left the model bin width set at 1 cm and only varied the data bin width. The difference in performance of these two cases allowed us to differentiate the impact of the two types of bins, as well as quantify the advantage of separate model and data bin structures.

2.4 Model convergence and performance

Verifying model convergence, as is typically done with an actual assessment (e.g., by trying multiple starting values, or checking for an invertible Hessian), is impractical in the context of a simulation study with tens of thousands of model runs. During development of the simulation we found models with a maximum gradient less than 0.1 and no parameters stuck on their bounds consistently provided reliable results. Therefore, if a replicate of an EM failed to meet these conditions (or failed to converge at all) we considered it non-converged and excluded it from our results. We ran 200 replicates of each data-rich scenario and 400 for each data-limited scenario to account for the higher uncertainty and non-convergence of the latter, and reported convergence rates by scenario.

EM parameters were initialized at the true (OM) parameter values, with the exception of the recruitment deviations (intialized at zero), and R_0 which was initialized at a higher value to help stabilize estimation, particularly for life histories with higher recruitment variability. Bounds were set wider than typically used (Table 2) and no priors were included on any parameters.

We used relative error ($RE = \left(\hat{\theta} - \theta\right)/\theta$) and median absolute relative error ($MARE = median\left(\left|\hat{\theta} - \theta\right|/\theta\right)$) between OM (θ) and EM ($\hat{\theta}$) parameters and across replicates to quantify estimation performance (accuracy and precision). We focused on growth parameters, but also tracked two quantities of interest to management: spawning stock biomass at maximum sustainable yield (SSB_{MSY}) and biomass in the last year relative to unfished biomass ('depletion'). We also tracked the EM run time and number of iterations to convergence.

3 Results

3.1 Structural impact of model bins

The magnitude and pattern of RE for quantities of interest from increasing model bin width in the OM (using the 1 cm model as baseline) varied between life-history type, with cod exhibiting the smallest effects and rockfish the largest (Fig. 3). Relative error in both SSB_{MSY} and depletion oscillated with increasing model bin width, and was generally positive (Fig. 3a,c,e). We observed two patterns in annual *SSB* values across life histories (Fig. 3b,d,f). First, *SSB* always increased (positive RE) with increasing model bin width, with the increase being most pronounced between 10 and 20 cm model bin widths. Second, the annual changes in RE for *SSB* were highly variable among life histories. For example, with a 20 cm model bin width, the cod life history showed little change in RE, flatfish RE increased and then decreased, and rockfish RE only increased.

3.2 Tradeoffs of model performance and speed

Patterns were generally similar among the three life histories, so we only present results for the flatfish models in the main text. Figures for the rockfish and cod models can be found in the supplementary material (Appendix A; Figs. S1-S6).

For the data-rich, age-composition scenario, estimates of management quantities and growth parameters were generally unbiased (Fig. 4a-g). The three exceptions were CV_{young}, L_{min}, and k with matching data and model bins of 5 cm and 10 cm. For example, median RE values were 6.8% for CV_{young}, -20.5% for L_{min}, and 6.5% for k for 10 cm bin widths. Despite the biases in growth parameters, the management parameters were relatively unbiased, with median RE value of -1.7% for both SSB_{MSY} and depletion.

For the data-rich, CAAL scenario, estimates of management quantities and growth parameters were also generally unbiased with 1 cm model bins. However, the CAAL data scenarios tended to be more sensitive to increasing model bin widths, compared to the scenario with age compositions (Fig. 4h–n). For example, with 10 cm data and model bin widths, growth parameters were biased, and the median RE for SSB_{MSY}

was 20.3% and -16.1% for depletion, substantially higher than the -1.7% for both with age composition data. The most notable improvement seen by using CAAL data over age compositions, in the data-rich scenario, was in the reduced uncertainty of the growth parameters for 1 cm model and data bin widths, particularly the *CV* parameters and *k*.

Compared to the data-rich scenarios, the data-limited scenarios generally had similar, but exaggerated, patterns with substantial bias and variability for all bin-width cases (Appendix A). Interpretation of the data-limited scenarios was further complicated by convergence issues (see below).

The rate of convergence, defined here as a maximum gradient of less than 0.1 and no parameters on their bounds, declined as data bin widths increased (Fig. 5a-d). We found that nearly 100% of all data-rich scenarios with 1 cm model bins converged. Convergence rates for data-limited scenarios followed the same pattern, but were much lower than those from data-rich scenarios. For example, with data and model bins of 1 cm, the data-limited scenarios had convergence rates of 63.5% and 90.5% for age compositions and CAAL data, respectively. Models that failed to converge typically were stuck at the lower bounds of CV_{young} or L_{min}, or had a high gradient. We were unable to get reliable bias adjustment parameters for the scenario of 20 cm data and model bins with data rich CAAL for the flatfish life history, which prevented us from implementing this scenario.

As expected, the run time per iteration consistently decreased as the data bin width increased (Fig. 5m-p). Scenarios with matching model and data bin widths ran faster compared to those with 1 cm model bins (Fig. 5e-h). One notable exception, unique to the flatfish model in the data-rich CAAL scenario, was the substantial increase in iterations (and thus run time) as model bin width increased (Fig. 5j). Run times were on average 1.6 times longer for CAAL scenarios compared to age scenarios (range 0.9 to 2.8) for the data-rich scenarios (Appendix A). The pattern in run time for data-limited scenarios was not consistent across life histories, likely due to convergence issues.

We found a consistent improvement in estimation with increased run time but only for data rich CAAL scenarios (Fig. 6). For example, by switching from 1 cm to 5 cm data bins in the data-rich, CAAL scenario, run time was less than half and with minimal change in estimated management quantities

4 Discussion

We tested the impact of length bin specification on model performance and accuracy of age-structured stock-assessment models, and conclude higher bin resolution improves performance, but increases model run time. The width of model bins was found to be particularly important, as both operating model behavior and estimation accuracy improved with finer model bins. Wider data bins introduced bias into growth parameter estimates, but did not substantially affect management quantity estimates. As a result, we recommend analysts use fine model bin widths, and determine the optimal tradeoff between better parameter estimation and increased run time when specifying data bin width. For models without the option of distinct model bins, we expect this tradeoff to occur faster, such that wider bins are less likely to be a good option.

We caution analysts that increasing the model bin widths can inherently change model outputs, including important management quantities, even before data are included. This effect likely occurs because aggregating lengths into equally spaced bins is a linear transformation applied to inherently non-linear processes (e.g., selectivity, maturity, and growth). For example, the model assigns a weight to all fish in a length bin by using the length at the mid-point of the bin, even though weights within the bin are not uniform. Growth, maturity, and selectivity interact, and it is therefore difficult to predict how a model will change with increasing model bin widths. In this study, we found spawning biomass tended to increase with increasing model bin widths. This effect was more pronounced for life histories that are fully selected when they are well below their maximum size (e.g. cod) than slower-growing species (e.g., rockfish). However, we encourage analysts to explore the sensitivity of their models to increasing model bin widths, because the effect is difficult to predict, but straightforward to check.

We found the best estimation performance (i.e., increased precision and accuracy of REs) resulted from the use of 1 cm model and data bin widths. This was not surprising as the discretization error is minimized with finer bins, and those estimation model configurations matched our generated truth from the operating models. However, there was a minimal loss of accuracy when model bin width was held at 1 cm and data bin width was increased to 5 or even 10 cm in some cases. This result suggests that reliable estimates may be possible, even when lengths are measured coarsely (e.g., visual surveys as in SEDAR, 2005). This also validates the use of distinct model and data bin widths in SS, and suggests analysts building custom models strongly consider adding this feature. We recommend analysts use a fine resolution for model bins, regardless of the resolution of the data and assuming run time is not an obstacle. For example, if an analyst is given length-composition data in 5 cm bins, we advise they still use a 1 cm model bin width to maximize estimation accuracy.

In some situations, decreasing run time may be particularly appealing. For instance, simulation testing (Deroba et al., 2014) and management strategy evaluations (Punt et al., 2014) can have tens of thousands model replicates, and Bayesian inference with the Markov chain Monte Carlo algorithm can have numerous iterations within a model (e.g. Stewart et al., 2013). Another situation where decreasing run time could be appealing may be with models that use CAAL data, because the data matrices are much larger than for marginal age composition data: $N_{age} \times N_{length} \times N_{year}$ versus $N_{age} \times N_{year}$, where N_{length} and N_{age} are the number of age and length bins and N_{year} is the number of years of age data for a fleet. In these situations where run time is an obstacle, the tradeoff with performance may be more important to understand. For cases where run time decreased with increasing bin widths, the estimation of growth typically degraded, but management quantities often remained stable. We therefore encourage analysts to consider the goal of their model (management advice, simulation testing, etc.), and explore the effect of aggregating data into wider bins, or increasing model bin widths, for the purposes of reducing run time.

As with any simulation study, we made a set of assumptions that should be considered when interpreting our results. For example, our model setups included a single fishery, sex, area, and source of process error. We also fixed parameters, such as steepness of the stock-recruit relationship and natural mortality, at their true values, which would be impossible in a real assessment. One particularly important assumption to highlight is that both age and length data were generated without measurement error (in the sense that fish were always assigned to the correct age or length bin). It is unclear how these two assumptions affected our results, and may be worth investigation in future studies. The simulated data were also unbiased with known effective sample sizes and had no outliers or other properties that conflict with the multinomial likelihood used (e.g., over-dispersed or dependent; Francis, 2014; Maunder, 2011). Our approach of using simplified models and generating idealized data contrasts somewhat with other studies that used specific empirical assessments and the bootstrapping feature of SS (e.g., Crone and Valero, 2014). Studies which use this approach use more realistic models, but the results may be difficult to generalize to other stocks and systems. We also expect the results from our simplified models will apply to many custom-built age-structured models, and thus our conclusions are applicable beyond SS models.

We see simulation testing of simplified biological and fishery systems as a vital first step to understanding how stock assessment models perform under more realistic conditions. Here, we outlined tradeoffs which are important for analysts to consider when binning length in age-structured stock assessment models. We show that specification of model and data length bins can affect estimates of management quantities and demographic parameters, suggesting analysts should carefully consider how length bins are specified when fitting stock assessment models.

5 Appendix A

Reproducible code, model configurations, results, and additional plots and tables are available at https://github.com/ss3sim/binning.

6 Acknowledgements

This growth research was supported through the Center for the Advancement of Population Assessment Methodology (CAPAM) in La Jolla CA, USA, as part of the good practices in stock assessment modeling program. This publication is partially funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement No. NA10OAR4320148, Contribution No. [TBD]. CCM, PTK, and CCS were partially supported for this work by Washington Sea Grant. MBR and CCS were supported by the NSF Integrative Graduate Education Research Traineeship (IGERT) Program on Ocean Change. Partial support for this research came from a Eunice Kennedy Schriver National Institute of Child Health and Human Development research infrastructure grant, R24HD042828, to the Center for Studies in Demography and Ecology at the University of Washington. RL was funded by Conicyt. We thank Ian Stewart, Mark Maunder, and André Punt for feedback on the initial study design. We also thank Rick Methot and Chantel Wetzel for technical discussions on SS and interpretation of results.

7 Tables

Table 1. The three life-history operating model configurations, including minimum and maximum length bin (measured in cm), growth parameters, and the parameters natural mortality (M), steepness (h), and recruitment variability (σ_R).

Life history	Min Bin	Max Bin	L_{min}	L_{∞}	CV_{young}	CV_{old}	k	M	h	σ_R
Cod	10	190	20.0	132.0	0.10	0.10	0.20	0.20	0.65	0.40
Flatfish	2	102	12.7	47.4	0.20	0.20	0.35	0.20	0.76	0.80
Rockfish	10	110	18.0	62.0	0.13	0.13	0.05	0.08	0.44	0.50

Table 2. Biological, fishery, and modelling parameters used for each life-history type. Lower and upper bounds are given in percentages of the true value from the OM, except for the *CV* and catchability parameters, which are absolute.

Variable Name	Symbol	Flatfish	Cod	Rockfish	Estimated	Lower	Upper
Natural mortality ($year^{-1}$)	M	0.20	0.20	0.08	No	-	-
Minimum age (year)	a_1	0.5	1.0	1.0	-	-	-
Maximum age (year)	A_{max}	25	25	70	-	-	-
Length at a_1 (cm)	L_{min}	12.7	20.0	18.0	Yes	5	500
Length at A_{max} (cm)	L_∞	47.4	132.0	62.0	Yes	5	500
Growth rate ($year^{-1}$)	k	0.347	0.200	0.047	Yes	5	500
CV at L_{min} *(-)*	CV_{young}	0.20	0.10	0.13	Yes	0.01	0.50
CV at L_∞ *(-)*	CV_{old}	0.20	0.10	0.13	Yes	0.01	0.50
Length-weight scaling (kg cm)	α	1.00E-05	6.80E-06	9.77E-06	No	-	-
Allometric factor (-)	β	3.00	3.10	3.17	No	-	-
Maturity slope (cm^{-1})	Ω_1	-0.400	-0.276	-0.400	No	-	-
Length at 50% maturity (cm)	Ω_2	28.90	38.18	38.78	No	-	-
Log mean virgin recruits (-)	$ln\ R_0$	10.5	18.7	5.6	Yes	4	20
Steepness (-)	h	0.76	0.65	0.44	No	-	-

Recruitment variability (-)	σ_R	0.7	0.4	0.5	No	-	-
Fishery length-at-50% selectivity (cm)	S_1	36.4	50.8	46.4	Yes	10	200
Fishery length selectivity slope (cm)	S_2	4.3	5.1	4.2	Yes	0	500
Survey length-at-50% selectivity (cm)	S_3	30.6	41.8	38.7	Yes	10	200
Survey length selectivity slope (cm)	S_4	4.3	5.2	4.2	Yes	0	500
Survey log-catchability	$ln\ q$	0	0	0	Yes	-20	20

Table 3. Data bin-width cases for each life history in the estimating models. All operating models had 1 cm model and data bins. Bin width ratio is the length bin width divided by mean asymptotic length (L_∞).

		Number of bins			**Bin width ratio**		
	Bin width (cm)	Cod	Flatfish	Rockfish	Cod	Flatfish	Rockfish
Base case	1	180	100	100	0.008	0.021	0.016
Small	2	90	50	50	0.015	0.042	0.032
Medium	5	36	20	20	0.038	0.105	0.081
High	10	18	10	10	0.076	0.211	0.161
Extreme	20	9	5	5	0.152	0.422	0.323

8 Figures

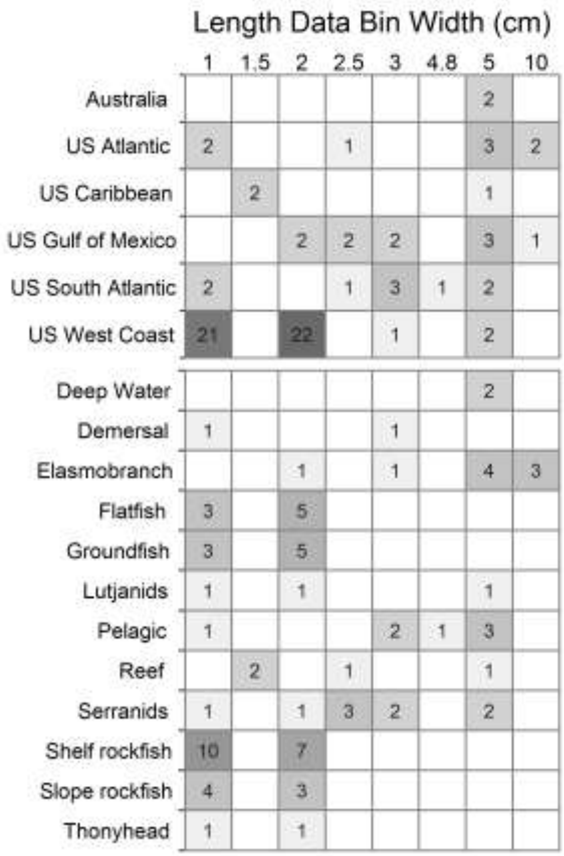

Fig. 1. Frequency of data length bin widths used by region (top panel) and by species grouping (bottom panel). Columns represent bin widths in cm, and cells contain counts with darker shading indicating higher counts. Results are from a non-exhaustive survey of Stock Synthesis models in the U.S. and Australia.

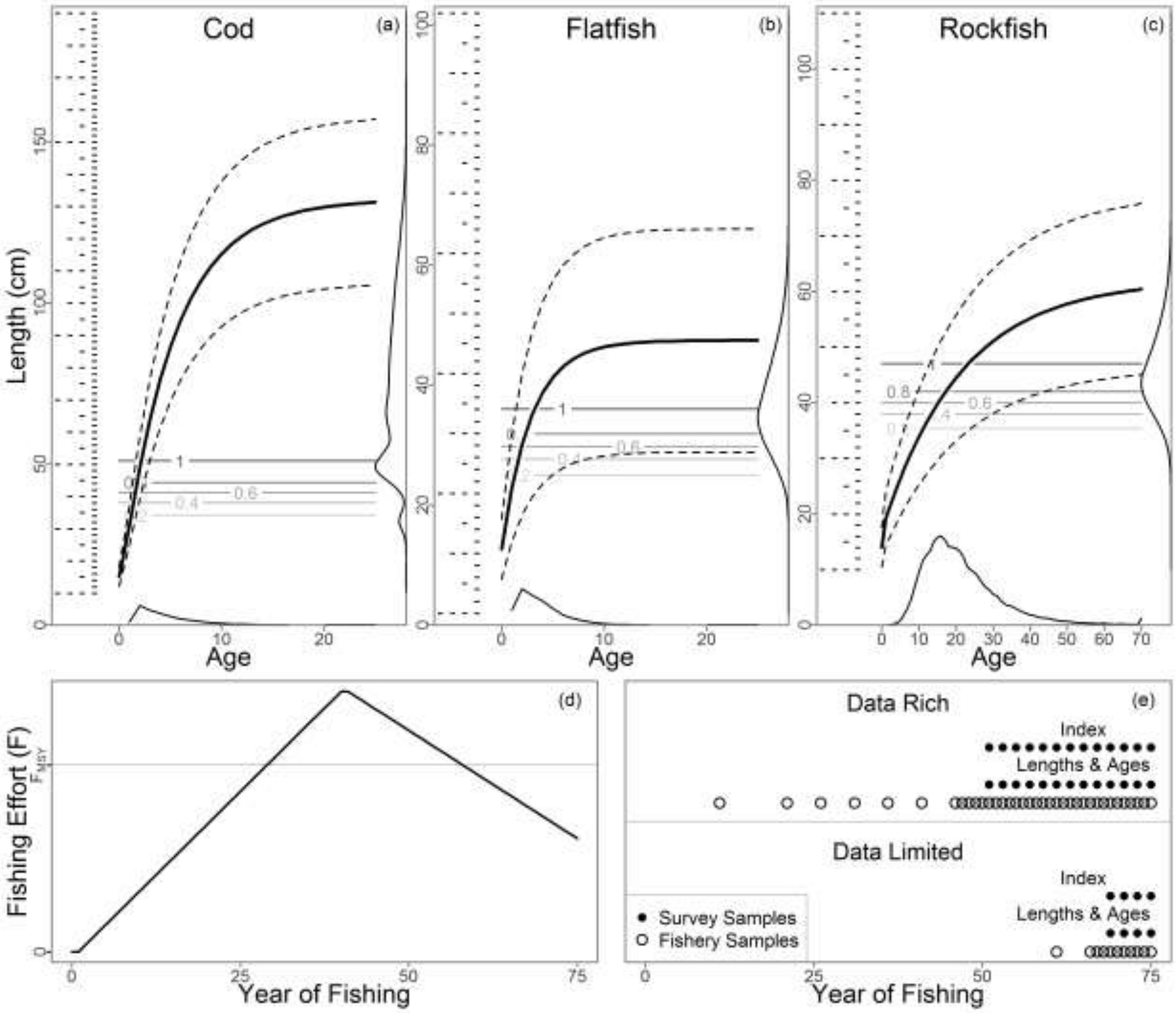

Fig. 2. Experimental design showing the life histories, fishing pattern, and data cases. (a–c) Growth functions are shown with their 95% confidence intervals (solid and dashed curved lines), while selectivity contours (length-based and matching maturity) are shown as shaded horizontal lines. Four bin-width cases (2, 5, 10, and 20 cm) are shown as columns of dashed lines on the left of each panel. The average fishery age and length distributions from a single replicate of the data-rich case are also shown as density plots on the respective axes. (d) The pattern of instantaneous fishing effort, relative to F_{MSY}. (e) The data type and years for two data cases. Ages can either be compositions or conditional age-at-length. Sample sizes are given in the text.

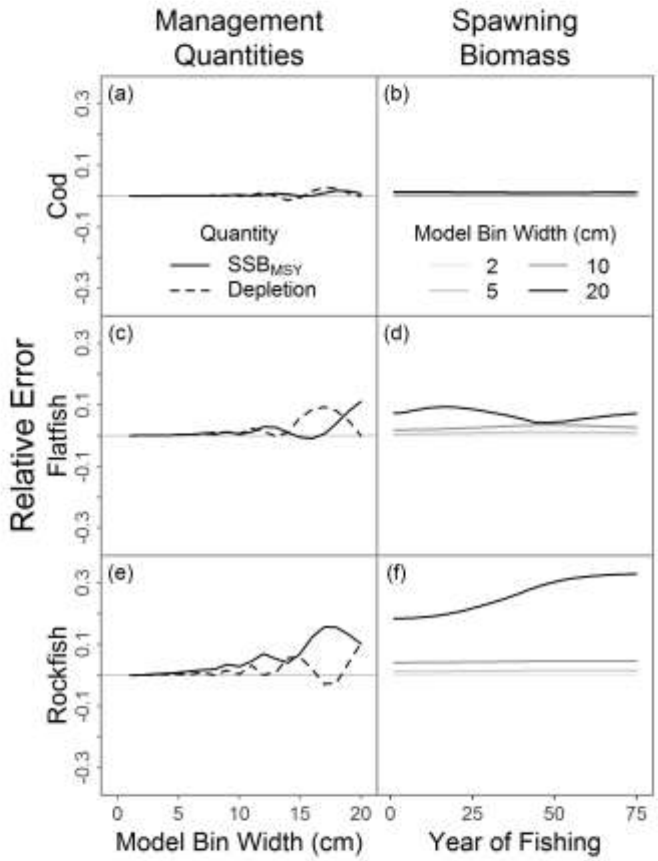

Fig. 3. The effect of internal model structure for variable model length bin widths in the operating model, while leaving parameters fixed and setting recruitment deviations to zero. Derived values such as management quantities, annual spawning biomass, and recruitment change because they depend on the internal fish length resolution. Relative error is calculated using the values from a 1 cm model as the base comparison. *SSB_{MSY}* is spawning stock biomass at maximum sustainable yield and depletion is biomass in year 75 relative to unfished biomass.

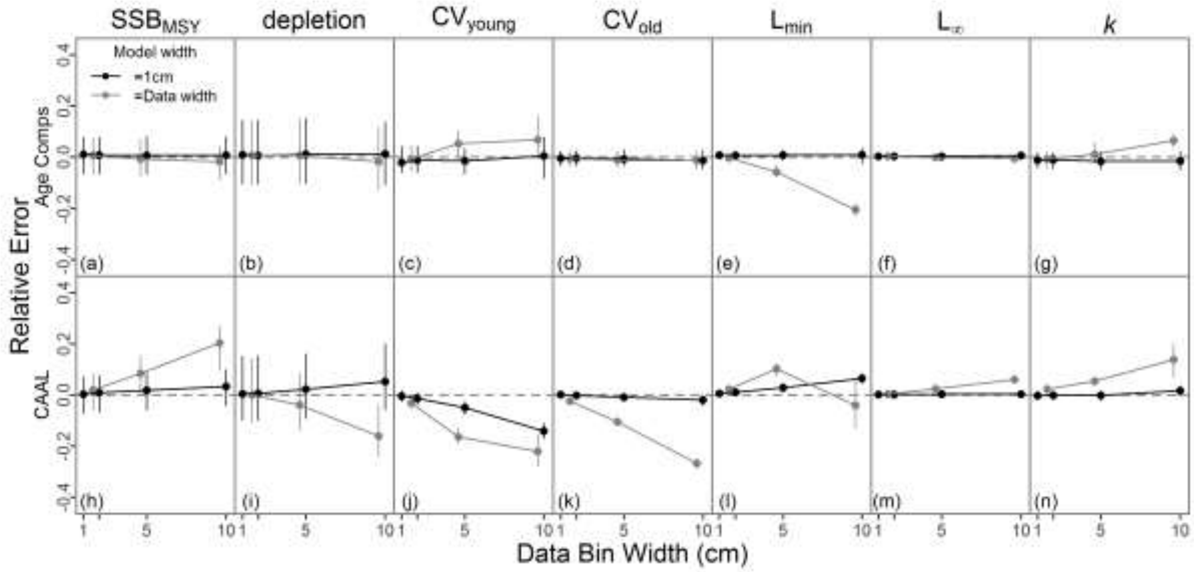

Fig. 4. Relative error of estimates of management quantities and growth parameters for data-rich cases across data bin widths for the flatfish life history. Points and lines show the median and interquartile range. Shading is used to indicate whether the model bin width was held at 1 cm (black) or whether it matches the data bin width (gray). Scenarios with less than 50% convergence are not shown. Cases of 20 cm bins are not shown for clarity.

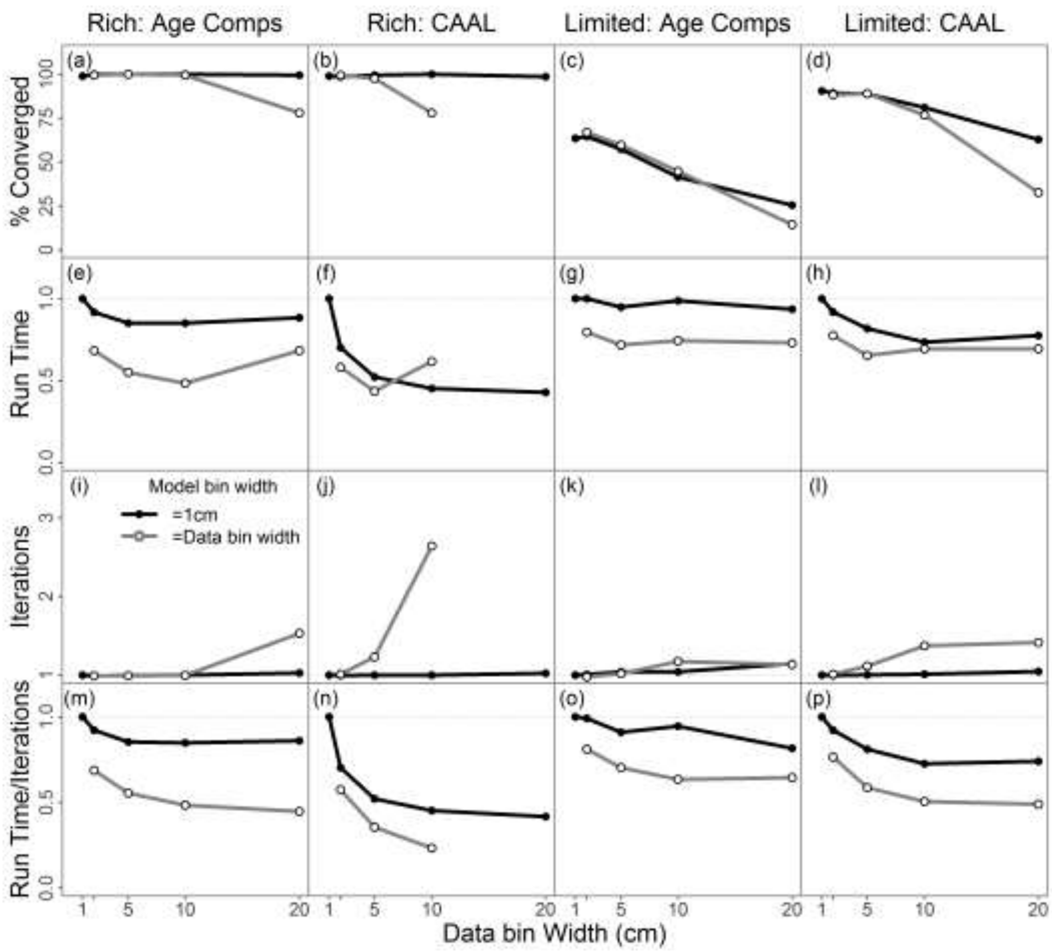

Fig. 5. Performance metrics (rows) for the flatfish model for four data cases (columns) by data bin width. Convergence is defined as the maximum gradient less than 0.1 and no parameters stuck on their bounds. The run time, number of iterations, and run time per iteration (last three rows) are normalized by data case and are relative to a model with 1 cm model and data bins.

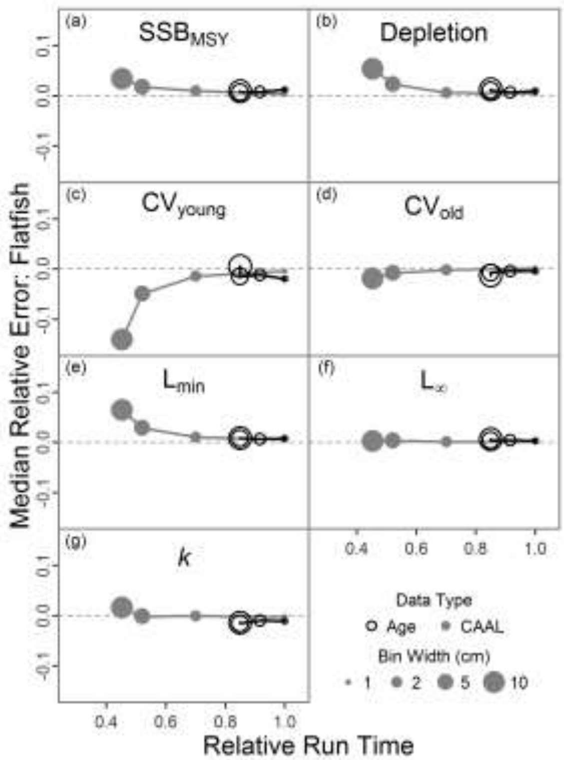

Fig. 6. Tradeoff between run time and relative error for the data-rich case for the flatfish model. Median run time and median relative error are shown for each data bin width for the case where the model bin was fixed at 1 cm. Relative run time is calculated relative to the slowest run time for each data type. Cases of 20 cm data bins are not shown for clarity.

9 References

Anderson, S.C., Monnahan, C.C., Johnson, K.F., Ono, K., Valero, J.L., 2014a. ss3sim: An R package for fisheries stock assessment simulation with Stock Synthesis. PLoS One. 9:e92725.

Anderson, S.C., Monnahan, C.C., Johnson, K.F., Ono, K., Valero, J.L., Cunningham, C.J., Hurtado-Ferro, F., Licandeo, R., McGilliard, C.R., Szuwalski, C.S., Vert-Pre, K.A., Whitten, A.R., 2014b. ss3sim: Fisheries stock assessment simulation testing with Stock Synthesis. R package version 0.8.2.99.

Conn, P.B., Williams, E.H., Shertzer, K.W., 2010. When can we reliably estimate the productivity of fish stocks? Can. J. Fish. Aquat. Sci. 67:511-523.

Crone, P.R., Valero, J.L., 2014. Evaluation of length- vs. age-composition data and associated selectivity assumptions used in stock assessments based on robustness of derived management quantities. Fish. Res. 158:165-171.

Deroba, J.J., Butterworth, D.S., Methot, R.D., De Oliveira, J.A.A., Fernandez, C., Nielsen, A., Cadrin, S.X., Dickey-Collas, M., Legault, C.M., Ianelli, J., Valero, J.L., Needle, C.L., O'Malley, J.M., Chang, Y.J., Thompson, G.G., Canales, C., Swain, D.P., Miller, D.C.M., Hintzen, N.T., Bertignac, M., Ibaibarriaga, L., Silva, A., Murta, A., Kell, L.T., de Moor, C.L., Parma, A.M., Dichmont, C.M., Restrepo, V.R., Ye, Y., Jardim, E., Spencer, P.D., Hanselman, D.H., Blaylock, J., Mood, M., Hulson, P.J.F., 2014. Simulation testing the robustness of stock assessment models to error: some results from the ICES strategic initiative on stock assessment methods. ICES J. Mar. Sci. 72:19-30.

Drouineau, H., Mahévas, S., Bertignac, M., Fertin, A., 2008. Assessing the impact of discretisation assumptions in a length-structured population growth model. Fish. Res. 91:160-167.

Francis, R.I.C.C., 2014. Replacing the multinomial in stock assessment models: A first step. Fish. Res. 151:70-84.

Goodyear, C.P., 1984. Analysis of Potential Yield per Recruit for Striped Bass Produced in Chesapeake Bay. N. Am. J. Fish. Manage. 4:488-496.

He, X., Field, J.C., Pearson, D.E., Lefebvre, L.S., 2015. Age sample sizes and their effects on growth estimation and stock assessment outputs: Three case studies from U.S. West Coast fisheries. Fish. Res. 10.1016/j.fishres.2015.08.018.

Hilborn, R., Walters, C.J., 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty. Springer Science & Business Media.

Hurtado-Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., Licandeo, R., McGilliard, C.R., Monnahan, C.C., Muradian, M.L., Ono, K., Vert-Pre, K.A., Whitten, A.R., Punt, A.E., 2014. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES J. Mar. Sci. 72:99-110.

Johnson, K.F., Monnahan, C.C., McGilliard, C.R., Vert-pre, K.A., Anderson, S.C., Cunningham, C.J., Hurtado-Ferro, F., Licandeo, R.R., Muradian, M.L., Ono, K., Szuwalski, C.S., Valero, J.L., Whitten, A.R., Punt, A.E., 2014. Time-varying natural mortality in fisheries stock assessment models: identifying a default approach. ICES J. Mar. Sci. 72:137-150.

Magnusson, A., Hilborn, R., 2007. What makes fisheries data informative? Fish. Fish. 8:337-358.

Maunder, M.N., 2011. Review and evaluation of likelihood functions for composition data in stock-assessment models: Estimating the effective sample size. Fish. Res. 109:311-319.

Maunder, M.N., Piner, K.R., 2015. Contemporary fisheries stock assessment: many issues still remain. ICES J. Mar. Sci. 72:7-18.

Methot, R.D, 2015. User manual for Stock Synthesis. Version 3.24s. Available at: http://www.st.nmfs.noaa.gov/Assets/science_program/SS_User_Manual_3.24s.pdf.

Methot, R.D., Taylor, I.G., 2011. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. Can. J. Fish. Aquat. Sci. 68:1744-1760.

Methot, R.D., Wetzel, C.R., 2013. Stock Synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fish. Res. 142:86-99.

Ono, K., Licandeo, R., Muradian, M.L., Cunningham, C.J., Anderson, S.C., Hurtado-Ferro, F., Johnson, K.F., McGilliard, C.R., Monnahan, C.C., Szuwalski, C.S., Valero, J.L., Vert-Pre, K.A., Whitten, A.R., Punt, A.E., 2014. The importance of length and age composition data in statistical age-structured models for marine species. ICES J. Mar. Sci. 72:31-43.

Punt, A.E., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A., Haddon, M., 2014. Management strategy evaluation: best practices. Fish. Fish. 10.1111/faf.12104.

R Core Team, 2015. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Schnute, J., 1981. A Versatile Growth-Model with Statistically Stable Parameters. Can. J. Fish. Aquat. Sci. 38:1128-1140.

SEDAR, 2005. SEDAR 8 - Caribbean Yellowtail snapper stock assessment report. SEDAR, North Charleston, SC. Available online at: http://sedarweb.org/docs/sar/S8SAR1_CaribYTSfinal.pdf.

Stewart, I.J., Hicks, A.C., Taylor, I.G., Thorson, J.T., Wetzel, C., Kupschus, S., 2013. A comparison of stock assessment uncertainty estimates using maximum likelihood and Bayesian methods implemented with the same model framework. Fish. Res. 142:37-46.

Szuwalski, C.S., 2015. Biases in biomass estimates: The effect of bin width in size-structured stock assessment methods. Fish. Res. 10.1016/j.fishres.2015.06.023.

Taylor, I., Wetzel, C., 2011. Status of the US yelloweye rockfish resource in 2011 (Update of 2009 assessment model). Pacific Fishery Management Council, Portland, OR. Available online at: http://www.pcouncil.org/wp-content/uploads/Yelloweye_2011_Rebuilding.pdf