University of Hawaii Manoa
Department of Oceanography
1000 Pope Rd
Honolulu, HI 96822

6 July 2020

Dear Editors,

We would like to extend our thanks again to yourself and our reviewers for your time and effort in reviewing our manuscript. Your constructive feedback has vastly improved the clarity of our work. We express our thanks both here and in our acknowledgements section.

In the following document, we have presented the major comments of each reviewer and the changes that have been made to the manuscript to address them. Our responses are italicized. Additionally, a spreadsheet detailing the minor concerns of the reviewers is also included with our resubmission. This document outlines the line-specific comments of reviewers as well as how the comment has been addressed.

We hope the editor and reviewers find the improved manuscript satisfactory for publication.

Thank you again for your work, as it has improved our own,

Stephen R. Scherrer Ph.D.

On behalf of all authors.

Reviewer 1

Fishery Selectivity

The first concern is with respect to fishery selectivity. The tagging program was fishery-dependent and used hook-and-line gear, which likely selectively samples various size classes of fish; I didn't see this accounted for in your description of your likelihood function. An example of this penalty is pasted below from Kapur et al. (2019) – there is TMB code for estimation available online which accompanies this paper.

$$L(D|\theta) = \prod_i S_{L_i} \frac{1}{\sqrt{2\pi}\,\sigma a_i} e^{-(L_i - \hat{L}_i)/(2[\sigma a_i]^2)} / \int_{-\infty}^{\infty} S_l \frac{1}{\sqrt{2\pi}\,\sigma a_i} e^{-(\hat{L}_i - l)/(2[\sigma a_i]^2)} dl$$

(6)

where L_i is the observed length at a given age a_i, $\hat{L}_i$ is the corresponding estimate based on VBGF parameters θ, S is a logistic selectivity function with parameter L_{50}, the length at which 50 % of individuals (male or

Kapur, M., Haltuch, M., Connors, B., Rogers, L., Berger, A., Koontz, E., Cope, J., Echave, K., Fenske, K., Hanselman, D., Punt, A.E., 2020. Oceanographic features delineate growth zonation in Northeast Pacific sablefish. Fish. Res. 222. https://doi.org/10.1016/j.fishres.2019.105414

Unfortunately, this dataset did not come with information on what gear was used to catch these fish, nor can we assume commercial fishers recapturing fish used standard gear. This makes accounting for gear selectivity non-trivial. We have instead implemented a sensitivity analysis per the suggestion of reviewer 3 that implicitly accounts for differences in catchability by synthesizing data such that the distribution of fish of all sizes is uniform. We can then compare the parameter estimates previously obtained to the estimates from fitting the same models to this synthetic data and determining how different the resulting parameters are. This provides an indication of the bias that sample size (which gear selectivity contributes to) imparts on the estimation process. The details of this analysis are found in Methods section 2.10 and Results Section 3.4 and summarized in table 6.

Lack of Priors

The second concern regards the implementation and presentation of the Bayesian model. Considering the previous work on growth for this species, it is strange that uninformative priors were used for all input parameters (line 120). Even a moderately broad prior (ie to account for sexual dimorphism) would be acceptable, but we do have a sense for the general range for this species. It is unsurprising that the MLE & Bayes methods performed similarly based on this setup, considering they used the same data source.

All four bayesian models have been refit using Linf and K parameters from Andrews et al. 2012 as priors.

It would simplify the reader's understanding of Model 1 vs Model 2 if you would use the language for hierarchical Bayesian models, whereby individual growth parameters are sampled

from a distribution. Please remove the line 125 where k is referred to as fixed, and simply state outright that k is estimated once for the entire population, same story for L129 regarding L infinity. This pops up several times (L171, "k is treated as fixed unknown parameter") – just say it was estimated for the entire population.

We have updated the verbiage in the methods section to reflect reviewer 1's suggestions and ensured that consistency is maintained through the rest of the document.

Finally, it isn't clear to me if/how the MLE and Bayesian methods were compared aside from the bootstrapped intervals overlapping. You mention "cross validation iterations to determine model structure" L349, with no prior description of what this is or what it does – nor a definition of "mssr". Once the corrections above are made, it would be useful to directly compare the best-fit model estimates from each approach.

MSSR (Mean sum of squared residuals) has been updated to a more conventional metric (RMSE - Root mean squared error). The best MLE model was determined through cross validation (see the methods section) but current computational constraints make cross validation with Bayesian models unfeasible. Instead we've used quantitative metrics to determine which Bayesian models are credible (CV, DIC), results of the sensitivity analysis, and similarity to MLE estimates.

Reviewer 3
Sensitivity Analysis
When comparing model 11 with models 1-5, the authors show that the inclusion of additional growth data improved the predictive capacity of growth models compared to tagging data alone (Lines 356 - 359), likely due to the "omission of the largest individuals from models 1-5" (line 410). Along those lines, my main suggestion for the authors is to conduct a sensitivity analysis for the influence of the distribution of samples in their data (i.e. varied representation of samples by age class).

Since this approach assumes individual variability, it is important to have sampled the entire distribution of length at age for results to be accurate. The authors' integrated approach addresses this, but not completely. Assuming that sampling was sufficient to describe the true mean and standard deviation of length at age, simulating observations from a distribution such that each age class is equally represented and then re-fitting the growth model could provide a metric to compare the results of the model fit to the raw data with for detecting this influence. Samples in the oldest age classes are unlikely to represent the full distribution of lengths at age, as large and old fish are naturally scarce. This, and general underrepresentation of older age classes compared to younger ones can bias the estimation of L∞ in some cases. This was suggested by the authors as a possible reason for differences in their results when compared to those of O'Malley et al. (2015) (line 380) and in discussing differences between model 11 and models 1-5 (lines 408 - 411). Indeed, this has also been suggested in publications providing recommendations for age and growth studies and describing their fundamentals (e.g. Cailliet and Tanaka 1990), but is not frequently examined quantitatively (e.g. Bolser et al. 2018). It is possible that model 11 included enough data such that this effect is negligible, but we do not know for sure unless a sensitivity analysis is conducted. Examining differences between models fit to the raw data and models fit

to a dataset with equal representation of samples at age would allow the authors to identify if the distribution of samples has influenced their results.

Differences in sample distribution can also confound comparisons between studies and sexes, such as the ones made in the discussion of the present study (paragraphs beginning on lines 381 & 425). In my opinion, it would be worth reviewing the sample distribution of the datasets associated with the studies being compared, if they are available, to identify differences in sample distribution that could confound comparisons (e.g. a disparity in samples at older ages).

The authors did an excellent job of acknowledging the influence of sample distribution - particularly where estimation of L∞ is concerned - in the portions of text mentioned thus far, but an explicit analysis would greatly enhance this paper and strengthen their conclusions.

A sensitivity analysis has been implemented using Bolser et al. 2018 as a framework. Methods and results of this analysis can be found under sections 2.10 and 3.4 respectively and summarized in table 6.