To the Review Board:

Thank you for reviewing our manuscript. Below in boldface we have responded to the reviewer's comments, which we found fruitful and constructive. Where applicable, we have pasted in the changed text in blue both here and in the manuscript for the reviewer's convenience.

Reviewer #2's chief critiques regarded the parameter values used in the IBM simulation model, which they felt should be more sablefish-like. We have repeated this analysis using more sablefish-like parameters to demonstrate that the results (method performance) are indeed scale invariant. We also included a second analysis wherein the performance criteria were relaxed per R1's suggestion, which increased the method's accuracy to 100% for several scenarios.

Per R2's request regarding gear selectivity, we have modified the analysis of sablefish survey data to include a length-based selectivity penalty on the lognormal objective function for data collected from British Columbia, which uses externally-estimated length selectivity parameters in their age-based model. This change resulted in slight shifts to the estimated VBGF parameters and resultant length estimates, but not to the qualitative trends in sablefish growth throughout the region. Other changes included improving figure clarity and a rewrite of the discussion.

On behalf of my coauthors, I thank the reviewers for their careful consideration of our manuscript and useful feedback, which has improved the study.

Sincerely,

Maia Sosa Kapur

University of Washington

October 2019

Reviewers' comments:

Editor: I agree with the 2 reviews that the manuscript needs major revision to be publishable. Please address all comments by both reviewers in the revision.

Reviewer #1:

Review of Fish9315. Data-driven approach reveals…sablefish

This is a nice study that combines a simulation study of a new method, with a comparison to an existing method (STARS), and with an application to a real fishery data set (sablefish). Rather ambitious. The approach looks relevant and thoroughly done. My criticisms are mostly about presentation, recognizing that these can be addressed in a revision.

We appreciate your comments and have done our best to address your suggestions for improving the presentation of the approach and results.

A have a few of things to say about the science:
1. The exact match for identifying a true breakpoint (l. 220) was acceptable but the presentation of the results and discussion about this could be simplified. See minor comment about l 313.

Understood – see response below.

2. I did not understand why you accepted some change points in Figures 5 and 6 but not others. For example, in 5f, the smoother peaks twice, with confidence limits not overlapping with the zero line, but you only picked the larger one to place as a dashed line in 5g. Same thing with 5b,f, and 6b,d. Why only pick one and if so, which one? Please explain.

Our discussion has been updated to explain this L405: "Since the purpose of this analysis was diagnostic (the detection of where the spline is changing the most), we were able to avoid undue influence from this parameter by a) selecting only the value corresponding to the maximum first derivative and b) that had confidence intervals not containing zero, which are common in highly curved splines. We also chose to use only the maximum absolute value of the derivative to avoid splitting the spatio-temporal surface into many small zones, which may have led to problems of small sample size, or ultimately be unrealistic to implement in a population dynamics model of the fishery and stock."

3. I was wondering if you would point out that sablefish ages have some error associated with them. This started to come up in the discussion, l. 450-453, but more as an orphan sentence. Perhaps some statement that these ages across agencies should be comparable, based on age workshop results, is appropriate at the least.

A good point also raised by R2. We have added to the discussion L 422: "In addition, we did not simulate nor consider error or bias in the aging (i.e., otolith reading) process (Cope and Punt, 2007), which would potentially introduce uncertainty in breakpoint detection. Based on aging workshops conducted for sablefish, we consider aging results used in the case study to be roughly comparable between regions (Fenske et al., 2019)."

4. The introduction lays out the topic well but it overlook an application of GAMS with clustering analysis for defining spatial stock structure of a marine fish (Winton et al. 2014). I assume this is an oversight, and that it would be fruitful to comment on this method as something available, and if the authors have an opinion, then what do they think of this as an alternative.

Thank you for this reference, which strengthens the argument for a data-driven approach (it was indeed an oversight). We have included commentary on this approach in our discussion rewrite, L386: "Alternate GAM-based methods, such as the clustering approach applied in Winton et al. (2014), have also demonstrated that detecting spatial structure through a spatially explicit process can reveal distinct sub-areas in fish traits (e.g. mortality). That study also found that models did not necessarily require explicit ecosystem data (like temperature) to perform as well as models with only spatial information."

Presentation

1. The term 'region' becomes difficult to follow. It is used in both a general sense in some places (i.e., l. 26, 79) but also in more specific ways, such as 3 'regions' (AK, BC, CC) or 5 regions (Figure 7, or line 482-3 that identifies a specific 'region 3' in lowercase). By the discussion, I was confused enough that I could not follow some parts.

We have changed references to AK/BC/CC "regions" to "management area(s)" or "within political boundaries", and retained the term "region" for the growth zones detected via the GAM analyses.

2. In a related sense, the repeated use and disuse of acronyms for places was annoying (see bottom half of page 13, in particular).

We have replaced all acronyms with the full names for clarity.

3. The discussion seems bloated and came in and out of focus. For example, the paragraph beginning l. 447 did not seem to make a coherent point. Paragraph beginning l. 454 unnecessarily invokes ecosystem-based management, when the results of this paper clearly have relevance to single-species management (or I just did not understand what the authors intended here). By the time I got to Figure 7, I was pretty confused, likely resulting from a couple of factors: 1) no obvious outline structure to the discussion, 2) no background information on the ecosystem or management context (I am from the east coast), and 3) the confusing depiction of Figure 7 (see minor comments). I can see the point of mentioning counter-gradient growth variation, but the comparisons of sablefish to silversides is a stretch, considering that the latter is an annual species that spawns in the intertidal zone. I was less clear by discard rates were coming up on l 567. My recommendation for the discussion is to develop a clear outline that support the main thesis of this paper (see, for example. L 38) and revise to cut the discussion in half. Recommendation for major revision relates mostly to the discussion.

Helpful and fair. We have re-drafted the entire discussion beginning L362 to match the following outline, and the length has been reduced, principally by 1) removing repetition of study results 2) moving discussion of the model performance vs STARS to supplementary material.

Short outline of discussion:

1) **Implications of simulation results**
    a. **Performance of the method**
    b. **Caveats of the method**
    c. **Intended uses and future research of the GAM based method**
2) **Implication of sablefish results**
    a. **Contextualizing our findings with other work**
    b. **Conextualizing our findings within the ecosystem**
    c. **Future directions for sablefish research**

4. The conclusion section seems unnecessary.

We have deleted this section.

Minor **We have made the spelling-related corrections mentioned below.**

l. 1, it should be 'Data-driven' as this compound modifies approach

Another reviewer felt this nomenclature was redundant so it has been removed from the title, and replaced most occurrences with "model-based".

l. 151, please be more specific about how you rounded. For example was a value between 22.5 and 23.4 assigned as 23? Or was it 23.0-23.9 = 23? You have a strict threshold for accepting a simulated sample, so it seems worth being specific here.

Yes, we updated this sentence L137: "For each parameter, we identify at which predictor value (e.g., latitude) the maximum absolute value of the first derivative is obtained; this is rounded to the nearest integer (e.g. a value between 22.5 and 23.4 would be rounded to 23) and defined as the "breakpoint" if its 95% confidence interval (generated using the standard error estimates for the derivative) does not include zero."

l. 155, you define a degree in (standard, not nautical, if I understand that correctly) miles here but in km on l. 404; check journal format and pick one

L309 now defines a standard degree in km.

l. 205 start a new paragraph at 'Under each scenario'?

L195 A new paragraph now starts here.

l. 279 Waite and Mueter 2013 not in literature cited

This has been added.

l. 275, this begins a rather long paragraph that addresses more than one topic. Break up in to 2-3 paragraphs, emphasizing why you are estimating an asymptotic value for predicted length. I was a bit unsure of this, after the paper seems to say it would use a size at age data approach not estimated from models.

A good point; we have softened some of the introductory language to indicate that we are using a combination of a "data-driven" and information theoretic approach. This section starting L272 is now broken into several paragraphs and now clarifies how asymptotic length is treated with the following: "We employed a stepwise exploration of whether estimates of L_∞ were significantly different between detected regions using the method and generated from this ecosystem break using the entire, non-sub-sampled dataset. Asymptotic length was used to ease comparison between estimated values and those used in the current assessments."

l. 307-311, I consider it poor style to write sentences that do nothing but point to a figure. Generally editors want you to make a point in a sentence that ends with the corresponding figure or table in parentheses at the end of the sentence, if only to keep things short and concise.

A fair comment –the sentence pointing to Figure 4 L306 now states: "…displays the coverage probabilities for the 95% confidence intervals and proportion of simulations wherein the correct breakpoint was detected perfectly or with a "relaxed" criteria (within 2 degrees, roughly 220 km, or 2 years), demonstrating the success rate of the method across a variety of simulations."

l. 313+ some of this is rather tedious. It appears that section 3.1 is making two points: 1) the success of the method using 'exact match' and 2) the success if you loosen up the match criterion. I would rewrite strong topic sentences for these two paragraphs and revise accordingly. In association with that, why don't you add 3 panels to figure 4 that show the success rate with +/- 1 or 2 degrees latitude (rather than the exact match) which should simplify the text in this section 3.1.

This is a good suggestion; we have updated Figure 4 to have 3 additional panels which show the success rate when the criterion is relaxed to +/- 2 degrees. The text starting L312 has been shortened and now reads:

"For all scenarios, the method achieved the highest coverage probabilities for the length-at-age 0 (L_1) [48%-97% coverage for three scenarios and 27% in the scenario with overlap]. Coverage probabilities for length-at-age 15 (L_2) were slightly lower [43% - 74% for three scenarios and 16% in the scenario with overlap]. In terms of spatial breakpoint detection, there was not a qualitatively strong difference in the method's ability to correctly detect latitudinal vs. longitudinal breakpoints across scenarios. Our GAM-based method correctly detected the lack of a breakpoint in 86% of simulations without breaks; there was no discernable pattern to the spurious spatial breakpoints identified in the remaining

simulations. The method did less well at detecting the accurate breakpoints for scenario 4 (a "true" spatial break at 48°), assigning the break between 45° and 50° longitude in 100% of simulations; similarly, for the scenario with a single breakpoint at 25°, the GAM-based method was 100% accurate when the criteria were relaxed to include breaks from 24° to 26°. Relaxing the criteria in this manner increased the method's accuracy to over 90% for all scenarios except one…We computed the mean absolute error in both L_1 and L_2 estimates across scenarios and found the maximum error to be 1.84 cm for L_1 and 6.98 cm L_2, both obtained in scenario 1. Finally, we did not find the method's accuracy sensitive to either halving or reducing the sample size by 25%; see Supplementary Table A2."

l. 357 I was not sure what the antecedent of 'initial stratification' was so I had trouble following this.

This sentence has been clarified L348: "Parameter estimation at this temporal stratification generated 95% confidence intervals for L_∞ which overlapped for males within all regions and for females in region 5 (Supplementary Figure A12)."

l. 362 I was not sure what the antecedent of 'this set' was so I had trouble following this.

This sentence has been clarified L352: "Once re-aggregated and re-estimated, we did not find overlapping confidence intervals for L_∞ for any adjacent regions (Supplementary Figure A14), so this set of specifications (five spatial regions for both sexes, and a temporal break for females in regions 1 through 4) was retained as our final spatiotemporal stratification."

Figure 4. These colors did not work in my b/w hardcopy. Yellow, in particular, did not print well. Also, why is the order of scenarios different here than in Table 2? That seems like an unnecessary way to confuse the reader.

Thank you for noting this; we have changed the colors in all figures to be B/W friendly, and ensured the scenario order is consistent throughout tables/figures.

Figure 7. There is too much on this one figure and the legend explains too little. In the text, you talk about three regions in some places (AK, BC, CC. l. 111-113) but there are 5 regions here. I could figure out that dotted lines mark 10 degree latitudes or longitude lines (but maybe that should be in the legend), but I was not sure what demarked the 5 regions. There was some mention of a 4th and 5th region (l. 366-368) but I eventually realized I was not given enough information to understand the point of this figure or to follow much of the discussion.

Understood; the 5 regions are in fact those detected by the GAM analysis and are not strictly at 10-degree intervals. Line 355+ now states: "The stratification consists of three regions bounded on their western border by a break at 130˚W; from south to north, these regions (labeled 1, 2 and 3 on Figure 7) are defined by latitudes 36˚N and 50˚N. They correspond generally to Monterey, CA and the northern tip of Vancouver Island, BC. Region 4 is the area between 130˚W and the ecosystem break at 145˚W (roughly Cordova, AK). Datapoints collected to the west of the ecosystem break are assigned to region 5."

Table 1. reference to 1996-current in the foot note seems incomplete; what is the terminal year? **Good catch, 'current' has been replaced with '2018'**

Literature citations. Many are incompletely formatted.

We have double checked these and will work with the copy editor to ensure they meet journal specifications.

Cited.
Winton, M. V., Wuenschel, M. J., & McBride, R. S. (2014). Investigating spatial variation and temperature effects on maturity of female winter flounder (Pseudopleuronectes americanus) using generalized additive models. Canadian Journal of Fisheries and Aquatic Sciences, 71(9), 1279-1290. doi:10.1139/cjfas-2013-0617

Reviewer #2:

Data driven approach reveals oceanographic features delineate growth zones in northeast pacific sablefish.

General
This paper proposes to detect spatial and/or temporal breakpoints in fish size-at-age using estimated derivatives of spline-based smoothing functions of latitude, longitude, and time. The authors develop a individual-based model simulation to test the efficacy of the proposed method given hypothetical scenarios for regional differences (or lack of) in growth parameters. The method is then used to estimate spatio-temporal breakpoints in growth patterns of sablefish in the northeast Pacific.

Strengths: Overall, the paper presents a solid quantitative approach to a problem that is fairly common in fisheries oceanography. The simulation study is valuable in providing a way to "ground-truth" the method's reliability in absolute terms, as well as against other methods (although see below).

Weaknesses: The paper has a few weaknesses. First, the Introduction could be more concise and to the point about the actual method and its applicability. The stated justification of the method against "typical" approaches is not warranted and should be revised or removed.

Agreed; the introduction has been shortened.

Second, although the simulation study is warranted and presented reasonably well (but see below re IBM), I found the actual parameter and data scenarios unrealistic. For instance, the range of fish sizes in the simulations (from 6-8 cm at age-0 to 258 cm at age-15?) probably rules out any extant fish species. I don't know how the results would change, but I suspect that such as range provides an advantage to precision of growth parameter estimates (especially for a CV~10%), which are key to detecting regional differences in the simulations. Therefore, the simulations are impossible to judge and would need to be redone for a more realistic scenario.

Noted and revised; we repeated the IBM simulation using more sablefish-like parameters (see below).

Recommendation: Reconsider after major revision and review.

Introduction

L43-66. The first paragraph of the Introduction could be deleted without affecting the quality of the paper. In fact, it would probably help to clarify what the paper is actually about. The second paragraph (L67+) is more direct and clearly indicates the topic (which is not management boundaries as implied on L43).

This is a good suggestion. We deleted the first paragraph, and moved the ~3 descriptive sentences for the paper objective to the end of the Introduction. The paper now begins with L41 "There is no consensus on how to model region-specific growth patterns in assessment or population dynamics models. Fish somatic growth rates are …"

L52-66. This is not a particularly convincing argument for two reasons. First, the method presented here is just a variation of the "typical" approach described on L52 for linking biological observations to oceangraphic properties. Second, the "data-driven" approach was historically described as a "shotgun" search for correlations. One can always fit models to spatial data and then find oceanographic features to "explain" various observations. In the quest to separate correlation from causation, specifying a priori hypotheses that generate specific, falsifiable predictions is ALWAYS prefered over shotgun approaches. So, in this sense, the proposed data-driven approach sounds nice, but is the weaker form of scientific inference. One would have to ignore a lot of the philisophy of science to accept that a data-driven approach is more scientific than a priori hypotheses - exactly the opposite of the argument presented here.

This is a fair comment, and perhaps our explanation of our approach excessively denounced the *a priori* hypothesis approach, as in practice, we used a combination of the data-driven and information-theoretic methods by testing the ecosystem break at 145*. We have modified the language of this sentence L64+ to read: "An alternative tool is a model-based method that identifies break points in fish size-at-age, which can then be used to aggregate data and estimate parameters related to somatic growth. The significance of these breaks can be falsified by comparing overlap in growth parameter estimates and tested against or among pre-specified breaks of interest (i.e. an area with a known ecosystem regime). "

L102. I suggest that "our method", "the method", "the proposed method" be given a name.

We've updated these references to specify "the/our/the proposed GAM-based method".

L106. This should start a new paragraph. In any case, what is presented here could also be deleted since it is not really about detecting spatial patterns in fish growth parameters.

We have deleted this introductory paragraph as suggested above.

Why would spatial trends in size-at-age imply stock structure when (i) sablefish are highly mobile and (ii) there is already no genetic evidence of differentiation?

Agreed – we have replaced the mention of stock structure with 'spatial variation in stock traits' on L60 – as it should be considered in operating model development, but as you correctly point out doesn't mean that these are independent sup-populations.

Methods

\

Why is the IBM so complicated for such a simple problem? Interactions between individual growth and selectivity seem more important than having a stock recruit relationship for a fish that is in unfished equilibrium.

The IBM was repurposed from more complex assessment studies by the first author for this work; the stock-recruit was simply an efficient way to generate new individuals in an accepted manner. We attempted to keep it simple by having the population remain unfished and without sampling bias. Below we describe how we've modified the analysis to account for selectivity in BC data during growth estimation.

L134. I would agree if I knew that that selectivity is a constant function of age across regions. So, what happens if selectivity is length-dependent (which it probably is) and there are spatial/regional differences in survey gear (which there are).

We understand this issue. Currently, selectivity is an independent function of length for both Alaska and the West Coast, and the revised operating model for BC also does not account for size-selectivity in sampling the population (which was the source of our confusion). We agree that there may in fact be spatial differences in gear selectivities leading to different ages sampled in the dataset, and we attempted to address this using the sub-sampling regime (to ensure even sampling across regions and ages when fitting the GAM) and only fitting the GAM to a single age and sex.

However, to address the fact that the BC model is still in development, we have modified the length estimation procedure to represent the length-based selectivity inherent in BC sampling practices and updated the text accordingly, which included removing the now-irrelevant speculation on selectivity from the discussion.

The methods section L279 describing this change now reads:

"To account for length-based selectivity, which is implemented only for the British Columbia data, we applied a penalty to the likelihood function as follows:

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

Where L_i is the observed length at a given age a_i, $\hat{L}_i$ is the corresponding estimated based on VBGF parameters θ, S is a logistic selectivity function with parameter L_{50}, the length at which 50% of individuals (male or female) are fully selected, set to 52.976 cm (Samuel Johnson, SFU, pers. comm.)

Equation 2 $S_L = \frac{1}{1+\exp(L_{50}-L)}$

As length-based selectivity is assumed constant in both the California Current and Alaskan assessments, S_L is set to 1.0 when fitting data points from those regions."

Cited: Department of Fisheries and Oceans. (2016). A Revised Operating Model for Sablefish (Anoplopoma Fimbria) in British Columbia, Canada. ***Department of Fisheries and Oceans, Canada, 3190 Hammond Bay Road Nanaimo, BC V9T 6N7*****, (April). https://doi.org/http://www.dfo-mpo.gc.ca/csas-sccs/**

L142-143. The units of these variables are missing. **Thanks, we have updated L197.**

L146. "uncertainty" has many meanings. Specifically, you are computing the standard error of the estimated derivatives. **Thanks, clarified on L132: "The standard error of the derivative estimates are computed as…"**

L152: I don't understand the "95% confidence interval does not include zero". Aren't you estimating a latitudinal break-point? Maybe I am getting confused between estimation using actual data vs simulations. If so, then I suggest separating the two - i.e., don't even mention the simulation study until section 2.2. (note: reading the simulation section didn't clarify this question. L222 describes how breakpoints were detected?)

We see your confusion; the same approach for detecting breakpoints and estimating growth parameters (aside from the selectivity change described above) was used in both the simulation and sablefish application sections, which is why it's presented once here. Additionally, the simulation section uses the 95% CI of detected breakpoints as a performance metric (section 3.1), and we separately examine 95% CI of estimated Linf from the sablefish data to discard statistically insignificant breaks (see below).

To clarify, we updated the sentence on line 137 to read: "…defined as the "breakpoint" if its 95% confidence interval (generated using the standard error estimates for the derivative) does not include zero."

Later, when discussing the 95% CI for estimated growth parameters, we specify (L348): ". Parameter estimation at this temporal stratification generated 95% confidence intervals for L_∞ which overlapped for males within all regions and for females in region 5 (Supplementary Figure A12)."

L156. There is a lot to unpack in this one sentence. Within this sentence, is a general software package - TMB - really important to the point here? TMB mainly generates a gradient function.

We felt it useful to mention the software used, but have now simplified this sentence L146: "For each of these new aggregated data sets, the parameters of the VGBF; L_∞ - asymptotic length [cm], k - the rate at which asymptotic length is approached [cm/yr] and t_0 - the estimated age at length zero in years) are estimated using maximum likelihood assuming that

the error is normally distributed with zero mean and variance σ). This study performed estimation in Template Model Builder (Kristensen et al., 2016).”

L174: Eq 5 needs some editing of parentheses ()

Thanks, we added the missing parenthesis, L166.

L176. The "bias-corrected lognormal error" is unclear here.

Thanks, this was unnecessarily detailed; we have changed L168 to simply state “lognormal error”. The bias correction occurs when estimates are converted out of log space.

L180-182. Hopefully, there are typos in the L1 and L2 values here. Do you mean, e.g., 6.2 cm and 21.5 cm? There are no 258 cm (8.5 feet!) long sablefish. Also, do any of the actual datasets contain sablefish in the 6-8 cm size range? I don't see how bottom longline or trap surveys could ever catch individuals this size since sablefish are mostly pelagic during their first year and certainly wouldn't be able to mouth a large circle hook (or foolishly enter a trap full of adult sablefish). This range of L1 and L2 will be very optimistic about the estimability of the growth parameter k, since it is largely determine near the origin.

See next comment – these were not meant to be sablefish-like values, but we have since changed the analysis.

L208. Similar to above: L_inf = 150 cm is not realistic for sablefish.

We understand your confusion; the initial simulation study was not designed to imitate sablefish life history values specifically, and the study results are scale-invariant (i.e. the results would be identical with L1 and L2 were doubled). To demonstrate this, and per your comment below we decided to change the values used in the simulation study to more closely resemble sablefish (A1 @ 3 yrs, Linf ~ 70cm). The results (in terms of method performance) are unchanged.

L208. log(sigma) = 0.1 means sigma = 1.1 - is this on log(length) or length?

Sigma = 1.1 on length (this is only a start value for the estimation). For simplicity we changed the text to not state sigma in log space, L197.

L210. It is unrealistic to have age-0 fish in a length-at-age dataset, especially for sablefish. I would like to see how this method does with more realistic data, which would involve a1 ~ 3-5 yr and a L_inf ~ 70 cm. Lower L_inf would compress the growth pattern, while higher a1 would mask growth at young ages, making detecting differences in growth parameters more difficult and more sensitive to individual variation in growth - sigma - which is also high for sablefish.

This is fair – as noted above, we changed the simulation study to have values more similar to sablefish, and as expected the method performance is scale invariant.

L245. What "ecologists"? I expected a reference.

This sentence was removed with the discussion rewrite.

Results

I can't comment much on the simulation results because I don't think they are relevant.

Figure 1. Besides the values being unrealistic, it is hard to tell any differences in the bubble sizes in the figure. **We changed the values to be more sablefish-like (as described above) and have increased the scale of the contrast in bubble sizes in these figures to aid in interpretation. They are also now in greyscale.**

Figure 6. It is clear from Fig 6 that there could be multiple maxima/minima of spatial or temporal derivatives. Why chose the single largest one only?

This was also mentioned as a point of confusion by R1. Our discussion has been updated to explain this L405: "Since the purpose of this analysis was diagnostic (the detection of where the spline is changing the most), we were able to avoid undue influence from this parameter by a) selecting only the value corresponding to the maximum first derivative and b) that had confidence intervals not containing zero, which are common in highly curved splines. We also chose to use only the maximum absolute value of the derivative to avoid splitting the spatio-temporal surface into many small zones, which may have led to problems of small sample size, or ultimately be unrealistic to implement in a population dynamics model of the fishery and stock."

Discussion

Is growth zonation biologically significant? 95% intervals could be small bc of sample size. **That is a good point, and impacts on assessment results would have to be considered in the context of the fecundity relationships in given regions, which was outside the scope of this study. We've added a line L485: "We note, however, that the procedure used to eliminate 'overlapping' L_∞ estimates concerned only statistical differences in values (and are therefore sensitive to sample sizes). The biological significance of these values would need to be investigated in the context of fecundity and length-weight differences between regions."**

L443. What might one expect if ageing error were taken into account? How would the form of ageing error and growth parameters interact to affect the bias? For instance, if a fish reaches L_inf by age-25, then does an ageing error of +/- 5 years matter for fish length-at-age 35+? **This is a good point. We have included in our discussion a mention of aging error concerns for this region, L422: "In addition, we did not simulate nor consider error or bias in the aging (i.e., otolith reading) process (Cope and Punt, 2007), which would potentially introduce uncertainty in breakpoint detection. Based on aging workshops conducted for sablefish, we**

consider aging results used in the case study to be roughly comparable between regions (Fenske et al., 2019).”

L461. I remain sckeptical about these generalizations given that the simulation conditions favored highly precise growth parameter estimation. Try the simulations based on actual parameter estimates and size ranges representative of each region.

Per your comment above we decided to change the values used in the simulation study to more closely resemble sablefish (A1 @ 3 yrs, Linf ~ 70cm).

L480-481. Gear selectivity is not specific to fishery-dependent data. All sampling gear is size-/age-selective to some degree. **A good point, we have repeated the analysis of BC data to account for length-based selectivity (which is estimated externally for that region’s assessment), see above.**

L481-487. I am curious as to why this is curious to the authors. The BC sablefish assessment clearly uses length-based selectivity in the assessment, so how could be it "unknown to" and "not reflected in" the current assessment? The growth parameter estimation doesn't account for size selectivity of trap gear (I don't think any of the other regions do either).

We recognize the confusion here. Please see comment regarding L134 above for how we now account for size selectivity of trap gear in the growth estimation. Per this change, we removed the entire section of the discussion which wrongly stated the absence of size selectivity in BC and discussed accounting for it as a future direction.

L543. I appreciate the theoretical discussion here and how it relates to the observed patterns for sablefish. Perhaps this could be expanded a bit to give a scenario that would explain sablefish observations.

Thanks. Another reviewer suggested a complete re-write and shortening of the discussion. We have retained some theoretical material, and included the following discussion of potential sablefish scenarios L504: “A plausible scenario which would generate our observed results could be that changes in fisher behavior or climate in the last ~10 years caused female sablefish to move northward in greater numbers, or simply experience size-based truncations in regions to the east of 145 due to fishing pressure. Each of these phenomenae would have an inverse effect on resultant size-at-age, with fish entering the northern ecosystem tending to grow larger and high, persistent fishing pressure in any region leading to truncations in terminal size. Because we only detected slight declines size-at-age between time periods for female sablefish, it is possible that either fishery-related effects simply have not lasted long enough to be strongly evident, or such effects are being counteracted by more fish entering ecosystems favorable to higher terminal sizes. A closer examination of sex-related movement would be useful towards this understanding.“

L551. I don't see where this paragraph is going. The topic sentence doesn't seem related to the overall content.

The re-write of the discussion has updated this section (L482); this paragraph describes reasons why the temporal break could have been more pronounced for females than males.

L571. Are fishing mortality rates in the different regions actually large enough to substantially affect observed length-at-age? I doubt it, but you could test these hypotheses using the IBM.

To clarify, the IBM doesn't use *F*; it models an unfished population. You are correct that this could be explored with a re-configuration of the model to simulate different levels of fishing pressure, though the scope of this study was not to investigate if/how the method could detect changes in observed length-at-age due to F explicitly.

Table 1. There is not much discussion in this paper about how the survey methods could affect perception of regional differences in growth rates. For instance, it is likely that trawl surveys have dome-shaped selectivity for length, which would tend to generate smaller L_inf and higher k values (especially where trawls tend to be more selective for smaller fish compared to other gears).

As stated above, the trawl surveys used for Alaska and the West coast use asymptotic *age* based selectivity, with an independent selectivity function for length (all lengths = 1). See comments above regarding the updated analysis meant to address selectivity concerns for regions in which length-based selectivity is currently considered.