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.

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.
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.
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.
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.

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

Minor
l. 1, it should be 'Data-driven' as this compound modifies approach
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.
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
l. 205 start a new paragraph at 'Under each scenario' ?
l. 279 Waite and Mueter 2013 not in literature cited
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.
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
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.
l. 357 I was not sure what the antecedent of 'initial stratification' was so I had trouble following this.
l. 362 I was not sure what the antecedent of 'this set' was so I had trouble following this.
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.
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.
Table 1. reference to 1996-current in the foot note seems incomplete; what is the terminal year?

Literature citations. Many are incompletely formatted

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.

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.

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)

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.

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

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.

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?

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.

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).

L142-143. The units of these variables are missing.

L146. "uncertainty" has many meanings. Specifically, you are computing the standard error of the estimated derivatives.

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?)

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.

L174: Eq 5 needs some editing of parentheses ()
L176. The "bias-corrected lognormal error" is unclear here.

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.

L208. Similar to above: $L\_inf = 150$ cm is not realistic for sablefish.

L208. $\log(\text{sigma}) = 0.1$ means sigma = 1.1 - is this on log(length) or length?

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 $a1 \sim 3\text{-}5$ yr and a $L\_inf \sim 70$ 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.

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

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.

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?

Discussion

Is growth zonation biologically significant? 95% intervals could be small bc of sample size.

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+?

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.

L480-481. Gear selectivity is not specific to fishery-dependent data. All sampling gear is size-/age-selective to some degree.

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).

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.

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

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

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).