Comments on Progress Report

December 12, 2022

It is looking good. It demonstrates the comprehensive nature of the model and both the amount and diversity of the PIT-tag data. Some comments:

1. Somewhere be sure to clearly state that your model pools detections across years except in the detection efficiency component. Annual differences in migration behavior will be incorporated by including environmental and operational covariates in a future version of the model.
   1. I added this text in the second paragraph of section 3.3.2 (“Movement model”)
2. Rate vs probability: All probabilities can be considered rates, but not all rates can be considered probabilities. Be careful which term you use. I am thinking especially of the Introduction. “Survival rate” or “mortality rate” might refer to an instantaneous daily mortality rate, for example, or to survival per dam passed, whereas you appear to mean a population-level (or release-level) probability. Actually, we often omit either “rate” or “probability” when it is obvious, and just talk about “survival” or “mortality”. E.g., “… is implicated by low survival of Steelhead kelts, which decreases …” or “with mortality of 84-96% for kelts…”.
   1. I made these fixes in the introduction and executive summary, removing the word “rate” or replacing it with “probability” if I was referring to modeling results. In the summary statistics when talking about the frequency of overshooting and fallback, I also removed the word “rate.” For example, I changed “55% overshooting rate at McNary Dam” to “55% of tagged individuals overshot McNary Dam.”
3. Introduction: Be careful not to draw species-level conclusions from preceding work, both because those studies used only a sample of data and because you are later claiming that their analysis was insufficient. For example, instead of saying, “fallback rates have a positive relationship with river discharge (Boggs et al. 2004)”, it is more correct to say “Boggs et al. (2004) observed a positive relationship between fallback rates and river discharge.” Similar with the Richins results.
   1. Made these changes in the last paragraph of the introduction.
4. The word “data” is plural. So it should be “data are” or “data were”, not “data is” or “data was”.
   1. Changed all of the instances I could find.
5. Did we agree to use ESU or DPS?
   1. DPS. Changed all of the times that I referred to “ESU” to “DPS”.
6. Results: the text that refers to a table should come before the table, not after.
   1. Fixed!
7. Results: summary statistic tables (Tables 3-5): The proportions are not very useful without some measure of variability or the sample size. Also, because these tables are meant to simply give the reader a better understanding of the data you are working with, it is more appropriate to show counts here rather than proportions. The proportions look too much like probabilities, and as such they do not account for detection efficiency so they can be misleading. So, my recommendation is to restructure Table 3 to have the following columns: natal origin, sample size (i.e., number of tags from the natal origin that you are working with), number of tags exhibiting the behavior, and the percentage of tags exhibiting that behavior (=column 3 divided by column 2; okay to present as a %). You may also want to include a column that identifies the ESU/DPS. Tables 4 and 5 should have similar structure.
   1. For Table 3, I created a new table with columns natal origin, DPS, sample size, number of fish seen outside of DPS, and percent. For Tables 4 and 5, I replaced zeros with NAs where overshoot and fallback, respectively, are not informative. I reformatted Tables 4 and 5 slightly to show percent, number observed, and total sample size together in one cell. To format these tables similarly to Table 3, I would have to create a new table for every dam, which I think gets too long and unwieldy.
8. I like figures 4-19. Be sure to define the intervals in the caption, and also use a consistent x-axis range for all plots – or at least most plots. If there is a plot whose estimates are all very small or has some that are very large, you might consider a different range for the x-axis, but this should be the exception rather than the rule.
   1. I added text the caption indicating that the point estimates shown are the median values, and the intervals show the 95% credible interval. A consistent x-axis range of 0 – 0.85 has been applied to all plots.
9. When stating a point estimate in the text, it is appropriate to also provide the 95% credibility interval. It may not always be required, such as when you are presenting a range of point estimates, but it is appropriate in cases such as “… Methow River Steelhead had a 0.05 probability of straying…” (=Section 4.2.2). By providing the CI, you are indicating that (1) this is an estimate (which you may still want to emphasize, depending on the sentence structure) and (2) there is uncertainty in that estimate.
10. Table 6 (homing probabilities conditional on overshoot): I like this comparison – it helps really demonstrate the impact of overshoot and how it varies by population. As noted above, a more appropriate comparison would use the ratio of the overshoot to non-overshoot probabilities, with the 95% CI computed directly on that ratio. If that is doable in the amount of time you have, please consider adding a column that includes the ratio. You can keep the existing columns – the reader will want the individual estimates as they are, as well as the ratio.
    1. I added a column for the ratio of homing probability for overshooting vs. non-overshooting individuals to Table 6.
11. Section 4.4: Fallback probability by dam and origin – these are good results to show but let’s look for a more understandable presentation of them. It is concerning/confusing to see high probabilities of fallback presented for dam and origin combinations that were not observed in the data. For example, Fifteenmile Creek has high probabilities of fallback over all the upstream dams even though only 3 Fifteenmile steelhead were observed falling back over any dam upstream of McNary, and none were observed falling back over upper Columbia dams (and in fact none were observed overshooting those dams, either). Please confirm that the values presented in Table 7 are correct. If they are correct, you will need to address them or present them differently. If they are correct, do they arise because you equated migration behavior probabilities across origins for these non-homing regions (i.e., out of ESU and either upstream or not on direct pathway from ocean)? That would make sense, but it will be hard to convince a reader that they are informative when no fish from that population was assessed for the behavior (e.g., no Fifteenmile fish were available upstream of PRA for estimating fallback there). One option is to report NA for those dams where no fish were eligible for fallback assessment. Another option is to separate the table into three tables based on ESU, and discuss each ESU separately, similar to Section 4.2. If I were writing it, I would do both (separate by ESU and use NA’s in place of estimates based on no data for that natal origin). Also consider expanding the text that accompanies this section – that may happen naturally if you subdivide by ESU.
    1. Yes – these values are due to the fact that we calculated one movement probability for all fish in a DPS when they were outside of the DPS boundaries. I separated this table into three tables and put NAs for all of the fallback movement probabilities for which there weren’t any fish of that natal origin, and added more text to describe the values.
12. Can you also present the overshoot probability by dam and origin, condition on being the mainstem state where the natal tributary is? (Similar to Section 4.4, also considering comments on Section 4.4 above) - if there is time
    1. A probability of overshooting each dam, conditional on starting in the mainstem state that connects to the natal tributary, would require a series of simulations that I don’t have time for at this point. I’ve provided the same tables for each of the three DPSs as the fallback probability tables – the probability of overshooting a dam, conditional on being in the state directly downstream of the dam. Happy to take this out, if you think it’s misleading.
13. Discussion: Some potential additional topics are: (1) annual differences, (2) implications of assuming common out-of-ESU behavior, and (3) impact of spillway detections at LGR on perceived fallback at that dam (could it be estimated with available data, what other data would be required). Also comparing our results to existing literature.
    1. A discussion of future analyses that would address the impact of the spillway PIT antennas at Lower Granite has been added in the discussion under the heading “Detection probabilities in downstream passage routes.”
14. We should have a short conclusions section. See suggested structure in the text.
    1. Added this!
15. Appendix 2 tables: we have the similar problem here as in Table 7 with sizeable estimates for behaviors that were not observed in the data for individual populations. In this case, it is more reasonable to present these values because this is what the model says, but I would mark them with an asterisk to indicate that the parameter was equated across all natal origins in the ESU. In fact, you should probably do that for all such equated parameters, whether or not there were non-0 counts in the data for that population.
    1. I put an asterisk next to every probability that was shared across all origins within the DPS, and also added information in the table captions to describe this.
16. Significant digits: Instead of always using 2 significant digits (e.g., 0.000045), let’s always use 2 digits to the right of the decimal place. Values that are smaller than this (e.g., 0.000045) should be represented as <0.01. This should help standardize results and avoid implying a level of precision that we do not have. Also, be sure to include any trailing 0’s that are needed to get to 2 digits (e.g., 0.4 is 0.40) – R does not do that automatically, but you can use the format() function.
    1. Made this change in all tables and figures throughout.
17. The captions on some figures and tables could use more information. The reader should be able to understand what is being shown without having to read the text of the report. See particular comments in the report.
    1. Addressed the comments in the report and made sure that captions were informative.
18. Similar to previous comment: be sure to identify what the intervals represent in figures and/or tables – 95% credibility interval? range?
    1. Added this in all figure and table captions.
19. The formatting needs some work. The margins are too small, it should use at least 1.5 line spacing, and the page numbers are all in roman numerals. There may be an issue with how Markdown is using the CBR report format. Craig (at CBR) can help with the formatting of the final document, if it is not practical to make the formatting changes in MarkDown. He will need to work with the final version after you have made all your other changes.
    1. I made some formatting changes to the report in Word after knitting the R markdown file. If you’d like to see more formatting changes I’m happy to work with Craig.
20. Similar to the above comment: the figures use a lot of memory and do not always print (when I printed the report, the pages with Sections 3.1 and 3.2 did not print.) Is there a way to include them in a way that is more stable? Save them as pictures and reload them, for example? I don’t know if this is possible in Markdown or if it needs to be done later.