

Responses to the comments of Editor and Reviewers

CJFAS-2023-0139R2

Associate Editor comments to author:

I have received two thorough rereviews of cjfas-2023-0139.R1, Assessing the Impacts of Environmental and Ecological Variables on the Performance of Fraser Sockeye Salmon Forecast. Both reviewers indicated that the revised manuscript was substantially improved. Reviewer 1 had several important concerns that should be addressed by the authors, and reviewer 2 recommended accepting the manuscript. I agree with reviewer 1 that some revision is still needed. In particular,

1) the authors should double check that their equations are correct.

[1]Response: We double checked the equations and made corrections (details provided below: Point 2 of Reviewer 1).

2) A metric of bias would be a useful addition to the other performance metrics. If all the approaches have similar bias, then the average level can be reported in the text.

[2]Response: Thanks for your suggestion; accordingly, a metric of bias (Normalized Forecast Metric, NFM) was added to measure bias (Lines 224-227). For clarity and consistency, we have also slightly modified the performance metrics (Lines 214-216) with equations provided.

3) The affect of discretizing spawning performance at 0%, 50%, and 100% should be briefly discussed in the discussion.

[3]Response: We added a few sentences to discuss this issue and included an example to explain the assumption; please refer to our response to Point 1 of Reviewer 1 for details.

I commend the authors on their thorough revisions, and I recommend minor revisions before reconsidering this manuscript for publication. My minor comments by line number are below.

41. Please include the scientific name for sockeye salmon.

[4]Response: Thank you, the scientific name has been added.

149. Add "to be" after "found".

[5]Response: We made this change accordingly.

469. Change "forecast" to "forecasting".

[6]Response: We made this change accordingly.

Reviewer comments to author:

Reviewer: 1

Comments to the Author

In general, the authors have done a good job responding to the reviews and addressed most of my concerns. However, some of the added material causes me some concern:

Main concerns:

1. Categorizing spawning success as 0%, 50%, or 100% seems rather coarse and arbitrary. While this appears to be the precedent set by the cited Grant et al 2011 and Ogden et al. 2015 citations, I would like to see a little more justification of this approach, and discussion of any associated caveats, within this paper.

[7]Response: Thank you for your suggestion. We agree that 50% is an arbitrary value, however, we think that it would be more subjective for field technicians to quantify the percentage of spawn if we include more than current categories. As long as the partially spawned females are normally distributed and the samples are representative of the partially spawned females (1% to 99%), we believe it is a reasonable assumption for effective female spawner calculation. For clarification, we added additional text "Although 50% in spawning success is arbitrary and covers a wide range of partially spawned females, this category is reasonable for estimating effective female spawners, as long as the partially spawned females are normally distributed and the samples are representative of the partially spawned females (i.e., 50% spawned fish represents half of a fully spawned fish and half an un-spawned fish, Lingard et al., 2013; Stuart LePage, per. Comm.). For example, if 7000 females are observed fully spawned, 1000 females are 50% spawned and 2000 are dead un-spawned, the effective female spawners are 7500 ($7000 + 1000 \times 50\%$), the percent spawn is 75% ($\frac{7500}{7000+1000+2000}$), and the pre-spawn mortality is 25% ($1 - 75\%$)."(Line 138-145)

The new material on forecast performance metrics (lines 201-208) seem to have some problems as follows:

2. No citation for "mean absolute relative error (MARE)" is provided and this is not a metric I have encountered before (though I assume it is a straightforward combination of elements of MAE and MRE, but this is not reflected in the equation given) or turned up a reference for in a quick google search. There does not seem to be anything "relative" in the equation for MARE given on line 207 which is identical to the formula for MAE on line 205. The plots for MAE and MARE in Figure S3 look highly similar but not identical, so hopefully the correct equation (which is not stated) for MARE was actually used, but this is not clear from the manuscript.

[8]Response: Thank you for your careful review. Since MARE is not commonly used, we changed back to MRE(mean raw error) that was used in previous studies (Haeseker et al., 2008; Grant et al., 2010), but added the absolute calculation when doing the ranking. Doing so also helps facilitate comparisons with previous studies on the same salmon stocks. (Line 214-216)

3. Lines 207-208 state that MARE and MAPE represent "the longterm bias of forecasts" but since they are based on the absolute values of all differences between forecasts and "observations", they do not capture the direction of bias nor do they fully separate the effects of

bias and precision. Although MPE had issues I pointed out in my previous review, it is at least a signed metric of bias, as would be something like the median or mean log accuracy ratio ($\log y/x$) which would be symmetric in its sensitivity to over- versus under-forecasting (and the mean might better address the authors concerns about medians from small sample sizes, as well as considering the magnitude as well as rank of errors). At minimum, the misleading language around bias on line 208 needs to be revised. However, it now seems like the paper includes no metrics of bias -- would there not be management scenarios where bias is more concerning than overall accuracy? But there is no discussion of bias or its implications in the Discussion. At minimum, the Discussion should acknowledge that metrics of bias were not considered.

[9]Response: Thank you for your comments and suggestion; accordingly, a metric of bias (Normalized Forecast Metric, NFM) was added to measure bias (Lines 224-227). For clarity and consistency, we have also slightly modified the performance metrics (Lines 210-219) and removed the misleading language around bias. Additional results and discussion are included in the updated manuscript (Line 249-252 and Line 350-352)

Minor comments:

4. I appreciate the clearer description of the normalization of RMSE and standard deviation. However, these qualifiers are generally dropped later in the text, which can be confusing since the equation presented for RMSE is the non-normalized version. I suggest including the modifier "normalized" everywhere it would apply, throughout the manuscript.

[10]Response: Thank you for your suggestion. We added "normalized" before RMSE where it would apply (i.e., relating to Taylor diagram). We only applied normalized RMSE in Taylor diagram. That is primarily because we would like to use normalized standard deviation (SD). The normalized SD offers a way to compare models across stocks. Since RMSE and standard deviation are not independent, RMSE is also normalized. The raw RMSE was used for other ranking analysis, because standardization is a monotonically increasing function, the model ranking results remain the same.

5. The description of "false alarms" on lines 405-406 reads a little awkwardly, consider revising for clarity, although at least now I think I know what scenarios the authors are referring to.

[11]Response: Thank you for your suggestion. We re-wrote this sentence "Models with false alarms (bad models but appear to have good performance temporarily) will either demonstrate erratic behavior or deviate further from the observation." (Line 48-430)

6. On lines 281-286, while better performance of the forecast in 2022 and 2023 vs 2021 is worth mentioning, I think the wording needs to be more cautious given the limited ability to infer much about forecast performance based on $n=1$ or $n=2$. I'd also be cautious about attributing all of the observed improvement to the use of Taylor diagrams per se.

[12]Response: We agree with the reviewer and added another sentence after this paragraph: "Although the evaluation period is not long enough to make a conclusion, it would be interesting to continue monitoring the performance of this implementation in the near future." (Line 309-311)

7. Some of the added text (e.g. lines 392-396, 405-406, and 423-426) does not flow as well as the previous text, it may have been written hastily, and minor revisions for clarity for the recently added text might be considered.

[13]Response: Thanks for your careful review. We have revised the text (lines 413-419, 429-431, 444-455, 500-502) for clarity and better flow.

Reviewer: 2

Comments to the Author

The authors have done a great job responding to the first round of reviews. I appreciated the expanded technical discussion of some of the novel methods being used (Taylor diagrams, standard deviation as a forecast metric), and the broader treatment of different forecast approaches throughout the manuscript. It was also good to see some model diagnostics and output in the supplementary materials. The revised discussion is better suited to the results of the study and the authors have presented the appropriate caveats additional considerations to their work. I think the manuscript is acceptable for publication in its present revised form.

[14]Response: We appreciate Review 2 taking time to read this manuscript again and provide positive feedback. Thank you.