General feedback:

<u>SalishSeaCast</u>

The use of SalishSeaCast in your evaluations as opposed to solely relying on limited observations is a great choice. However, I feel that relying so heavily on a model without discussing its limitations and how they could relate to your interpretation of HOTSSea's effectiveness could make your evaluations somewhat misleading. Please include a discussion of SalishSeaCast limitations before presenting comparisons to it. Thank you, we now have added a note explaining that HOTSSea will share limitations of SalishSeaCast's (SSC) due to a lack of wetting and drying capabilities, climatologies used for river runoff, and apparent issues related to vertical mixing (Lines 235 – 240, Line 832). Other aspects of SSC model skill with respect to physical properties have been previously reported (Olson et al., 2020; Soontiens et al., 2016; Soontiens & Allen, 2017), though not necessarily using the same observational data or the subdomain definitions used here; thus our analysis possibly represents a first look at this model's performance in some respects. We feel it is not within the scope of this article to dedicate more space to targeted evaluation of SSC; Figure 4, for example quantifies some pertinent aspects of the SSC model performance.

<u>Python packages</u>

I wouldn't normally reference which python packages are used (ex. lines 113-119), I think that giving access to your code and referencing it at some point (as you have done) is sufficient. It is also difficult to imagine how a package is used without looking at the notebook/script, so I don't think that this information adds much to your paper. Removed the unnecessary specifics, as suggested.

Adding a README to your zenodo and GitHub repository, pointing to which folders one should look into for specific parts of the paper (such as temperature bias correction or SalishSeaCast evaluations), may be more useful to those interested in applying your methods. Readme was present in original submission; however, it could be clearer where the folders are for the analysis and bias correction – it was revised to clarify.

<u>Statistical equations</u>

The equation for widely known statistical metrics do not need to be detailed in the text (equations 1 and 2 for example), just mention which were used to keep the paper concise. It is much more important to describe what the metric reveals about a dataset. Where lesser-known

–

metrics were used, or modifications to traditional ones were made, then it makes sense to keep the equations.   Thank you, this has now been addressed.

<u>Puget Sound Evaluation</u>

It is necessary to draw the line somewhere, and you explain in the outline of future work that the collation of observations in Puget Sound is a necessary next step. However, since a large portion of your evaluations rely on SalishSeaCast, which covers Puget Sound, - it is not clear why Puget Sound was not included in that analysis. I think that the paper would benefit from either the inclusion of this work or a justification for why it was left out. Stressing that hindcasts of Puget Sound have already been done, as you mention on line 81, may do the trick.

Thank you. The original motivation was to synthesize the inputs for the Ecospace ecosystem modeling work focused on the Strait of Georgia and this is stated in the revised manuscript intro. Also, we mention how data was a limiter for evaluation of Puget Sound other models have hindcasted Puget Sound (though not as far back). (Lines 563, 597)

Line-by-line comments:

Line 56 – Feely et al., (2010) and Ianson et al., (2016) are also worthwhile regional acidification papers to reference here. Thank you – added these (lines 58-59)

Line 253 - I found this part confusing, explicitly state the difference between v0.14 and v0.16 to increase clarity. Thank you – we made clarifications.

Table 3 – I went back to this table many times while reading the text, here are a few suggestions to increase clarity:

- Are HRDPS1 and HRDPS different? If not, keep naming consistent. – Typo. Fixed.
- I think you are missing the "H" in "RDPS2 (10 km)" – we agree – this was confusing and there was a typo. We fixed it and added a table footnote to clarify.
- I liked the addition of approximate resolution next to HRDPS1 and RDPS in the second row. Include this for all the forcings. – as suggested, we have added approximate horizontal resolutions for all forcings. We feel it redundant to repeat all horizontal resolutions throughout Table 3, though, when it is the purpose of Table 1 is to do this.
- In your "evaluation purpose" for v0.16 I think that adding that this run went back to the "full" atmospheric forcing resolution would increase clarity on the difference between v0.14 and v0.16. – Made edits that help increase clarity, as suggested

Figure 2 – Add Nanoose station (the back star) to the legend. - Done

Table 4 – Normalize the CTD count in each subdomain by area in order to discuss heterogeneity more accurately in **line 274**. - Done

–

Line 287, 293, 297 – Define acronyms. - Done

Lines 302-309 – This belongs earlier, after line 282 perhaps, where you describe the model depth and time indexing. Tell us (the reader) about what you're conducting statistics on before telling us about the statistical metrics. – We made the suggested change

Line 315 – Accidental paragraph break? – Fixed.

Line 324 – I can't find this dataset in Table 2. If I am not mistaken, then why not include it? If you do include it, then I don't believe that the sentence "Canadian buoy data were downloaded…" is needed. The final row in Table 2 references the buoy data. We agree the sentence is redundant, so we removed it as suggested. We changed 'wave buoys' instrument label in Table 2 to 'buoys' for clarity (since these buoys collect SST and other info than just wave height).

Line 351 – "Described above" at the beginning of a new section is confusing, specify which section it was described in. – we made changes to make it less confusing, referencing the Table summarising experiments explicitly.

Line 367 – Remove "it" – Done.

Line 382 – I also don't see Nanoose station in table 2, I'm a bit confused about why some observations were not included. – Nanoose was included in second line of Table 2 but we agree it was buried in the description and was confusing especially because the dataset was prominent to the paper. We added 'Nanoose' to the 'Dataset Title' in Tab. 2 to fix this and modified Fig. 2 to hopefully make it clearer where Nanoose stn is located.

Line 393 – Does "depth strata of the closest HOTSSea grid cell" refer to the depth strata described on line 304 or the depth range of the closest grid cell? – Good catch, it was the latter. We have added clarification to address this.

Lines 434-436 – Very cool! – Agreed!

Line 447 – I'm not familiar with this statistic – how large is "large"? –We made edits to be more specific ('greater than zero' rather than 'large'). The value of the S statistic is constrained only by the number of pairs of data points being compared, so proportional to length of time series.

Line 494   Worthwhile to remind the reader of the difference between v0.12 and v0.14 to reveal what this similarity tells you. We made edits to address this.

Line 497 – Confusing wording since v0.14 and v0.16 both use ORAS5. We made edits to increase clarity.

–

Lines 499-505 – Make this description more clearly comparative between v0.14 and v0.16, at times it is unclear which version you're referring to. Agreed; we did a revision to this paragraph to make it clearer. (Lines 533-)

Line 508 – Perhaps I'm mistaken (and if I am than consider rewording) but I read this line as the higher resolution forcing (HRDPS) leading to lower bias than the lower resolution forcing (ERA5). I don't understand why this was unexpected, please expand. Thank you – this has now been reworded to clarify.

Line 510 – I interpret "across all depths" to mean here that the bias was lower everywhere. However, later in the paper it seems to mean the average over the whole water column. Make sure not to use this phrase for both applications. Changes made throughout (e.g., Lines 308, 610) for clarity ("across all depths" was referring to the depth stratum 0 – max z).

Figure 5 – Impressive changes after bias correction! –Yes, the simple correction made quite a dramatic improvement – it surprised us!

Line 576 – nCRMSE is defined with a capital "N" earlier in the paper. Fixed.

Figure 7a-7b – Add NCRMSE label to the plots. Done.

Figure 7c-7d – NCRMSE instead of NCRMSD; Done.

Lines 615-618 – These sentences seem to contradict. Is SSS variability overestimated or is it good? Thank you, we revised this paragraph to address this.

R2-M26 - Line 623 – In addition to a discussion of the time resolution required for HOTSSea to support ecosystem modelling this paper needs a discussion of what spatial resolution is required. Explain to the reader why 1.5 km is good enough. Thank you, more on this was definitely merited. Justification is now given in paragraph 2 in the intro, paragraph 2 in discussion (Line 790), and Line 149.

Line 630 – PSU? Yes - fixed

Line 630 – remove "be" Done

Line 648 – This seems like a rather large salinity bias at the surface to me. Put it into context. What is it in SalishSeaCast? Do you think that this surface salinity bias will affect model circulation? We expanded on this. The bias is shared with SalishSeaCast, as was apparent in Figure 4 in the SGS subdomain. Yes, it is most likely affecting model circulation. (Line 674, 831-835). We prioritise upgrading our model from NEMO 3.6 to 4.x before further exploration of

–

bias correction, since the newer version offers additional features with respect to vertical and lateral mixing.

Line 689 – How does it correlate? Negatively. Clarification made. (Line 717)

Figure 11 – Could there be a better way to express in the caption that it is a seasonal average of decadal trends? I think solely calling it "seasonal" in the caption could be a bit confusing. Agreed. Change made for clarity.

Line 750   Is it your intension to conduct a similar bias correction on salinity despite the instability problems you mentioned on line 369? Yes, we added a note related to this (Line 822,

Line 771 – Missing space between Figure 9 and Figure 10 Fixed

Table A1 – Something cutoff under "Model" in the top row Fixed

Figure A3 – Just a note that I think that the description of these Taylor diagrams is more clear than what you have in the main body. Thanks, we tried to clarify more in Fig 7 caption.

Thank you again for the helpful review. Please note, we are not sure the line numbers (referencing the revised manuscript) we used here will be that helpful since the final manuscript will not include them. However, we are instructed not to re-submit the manuscript in this discussion section. We would be happy to provide the revised Word file upon request if it helps, though.