

General feedback:

SalishSeaCast

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.

Python packages

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. 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.

Statistical equations

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.

Puget Sound Evaluation

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.

Line-by-line comments:

Line 56 – Feely et al., (2010) and Ianson et al., (2016) are also worthwhile regional acidification papers to reference here.

Line 253 – I found this part confusing, explicitly state the difference between v0.14 and v0.16 to increase clarity.

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.
- I think you are missing the “H” in “RDPS2 (10 km)”
- I liked the addition of approximate resolution next to HRDPS1 and RDPS in the second row. Include this for all the forcings.
- 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.

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

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

Line 287, 293, 297 – Define acronyms.

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.

Line 315 – Accidental paragraph break?

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.

Line 351 – “Described above” at the beginning of a new section is confusing, specify which section it was described in.

Line 367 – Remove “it”

Line 382 – I also don’t see Nanoose station in table 2, I’m a bit confused about why some observations were not included.

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?

Lines 434-436 – Very cool!

Line 447 – I’m not familiar with this statistic – how large is “large”?

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

Line 497 – Confusing wording since v0.14 and v0.16 both use ORAS5.

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.

Line 508 – Perhaps I'm mistaken (and if I am then 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.

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.

Figure 5 – Impressive changes after bias correction!

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

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

Figure 7c-7d – NCRMSE instead of NCRMSE

Lines 615-618 – These sentences seem to contradict. Is SSS variability overestimated or is it good?

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.

Line 630 – PSU?

Line 630 – remove "be"

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?

Line 689 – How does it correlate?

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.

Line 750 – Is it your intension to conduct a similar bias correction on salinity despite the instability problems you mentioned on line 369?

Line 771 – Missing space between Figure 9 and Figure 10

Table A1 – Something cutoff under “Model” in the top row

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