Review of: Moana Ocean Future Climate V1.0: High Resolution Marine Climate Futures For The New Zealand Region

This study reports on dynamically downscaled, ocean-climate scenario simulations (historical and projections) conducted for the ocean regions surrounding New Zealand. These efforts represent an important contribution to translating coarser resolution climate model output to regionally-relevant scales needed to understand risks posed by changing climate on marine ecosystems and resources. However, more rigorous justification and validation, and method clarity and consistency (i.e., major revisions) are needed for publication. I've listed broad issues below and attached a document elaborating on more specific, technical suggestions.

Justification for use of specific forcing products:

- It is not clear why the NZESM and the NZRCM (downscaled atmosphere from the NZESM) were the best choices for lateral boundary and atmospheric forcing conditions, respectively. The authors cite the outlier behavior for equilibrium climate sensitivity and transient climate response of the NZESM (Data, Lines 109-113). Also, with regards to representing the Sub-Tropical Front, the authors note "...known issues with the version of NZESM used as forcing in this study" (Abstract, Line 8). These issues would suggest that these data products are not the best tools to use with this framework.

More rigorous model validation and analysis:

- Validations of the new MOFC simulations presented in this manuscript are primarily comparison against a "well validated hindcast" (i.e., the existing, reanalysis-forced MOH hindcast simulation). Well validated does not mean the reference performs well in all aspects. There should be more comparisons against observations. Additionally, I recommend the authors indicate, when reporting discrepancies between historical MOFC and MOH, whether these departures are more or less consistent with observed ocean conditions. In other words, is MOFC better or worse than MOH?

- Many of the comparisons are between means or trends that are calculated over different numbers of years, which is problematic. This can result in aliasing interannual variability (e.g., ENSO) and misrepresentation of the multi-decadal climate state. Ideally, the historical reference simulation for the MOFC would be the same duration as the future projection simulations (i.e., 30 years). If extending the historical simulation is not feasible, the authors should provide additional context for the shorter historical period, indicating whether the NZESM forcing product exhibited more La Niña vs. El Niño-like characteristics (if ENSO is an important consideration for conditions in this region). This would add transparency to potential biases that may arise from timeframe inconsistencies. If ENSO is not as important, please report on other modes of variability that may be relevant.

Specific Comments:

- Introduction, Lines 76-78: *"In this paper we validate a 16 year period (1994-2010) of MOFC forced with historical NZESM against data from MOH and then explore regional scale medium-term (2030-2060) and long-term (2070-2099) marine climate change under the CMIP6 SSP2-4.5 and SSP3-7.0 emissions trajectories."*

  It is not ideal to be comparing periods that differ substantially in length (i.e., 30yr future vs. 16 historical). A climate normal is defined (World Meteorological Organization standards) as the average of a 30 year period or longer. The risk in comparing against shorter historical timeframe is biasing the comparison to aliased higher-frequency, interannual variability (e.g., ENSO) rather than the lower-frequency climate state. I recommend generating a 30 year historical simulation or demonstrating that the historical period is not overly skewed by a particular phase of ENSO and/or other important modes of variability in this region.

- Data, Line 112-114: *"This implies the warming seen in NZESM future climate simulations and our derived products are likely towards the higher end of warming under each of the selected Shared Socioeconomic Pathways."*

  This seems like an argument against using the NZESM to force future climate simulations. I suggest including a justification as to why this was a more suitable product than other CMIP models. Additionally, there should be some discussion of the implications in the results section (e.g., interpreting the MHW analyses).

- Method, Lines 139-140: *"Atmospheric forcing was primarily drawn from NZRCM with the exception of a narrow band of 1.9° latitude at the south of the domaine where NZESM was used instead; sensitivity experiments demonstrated that this approach did not induce significant bias"*

  Why was the narrow band of NZESM forcing applied? Please explain the need for this modification. And how were the sensitivity experiments posed such that they demonstrated a lack of significant bias?

- Method, Line 145: *" - A historical emissions run between 1990 and 2010"*

  The historical simulation should be of the same duration as the future periods that are compared against it. If feasible, I recommend running a longer historical simulation. Otherwise, I recommend investigating/discussing whether this shorter historical period might be biased (relative to the longer 30yr future periods) due to aliasing of higher-frequency interannual variability modes.

- Model Evaluation, Lines 152-154: *"To evaluate the performance of the present climate downscaling simulations we compared the run for the reference period (MOFC) with a*

well-validated hindcast on which the configuration was based (MOH). By doing this, we can identify possible biases that can contaminate the analysis of the climate change scenarios."

Ideally, model validation would be conducted against independent observation data rather than another model simulation. While the MOH may be "well validated", it is not without its own biases. Comparing the MOFC historical simulation against the MOH informs us on how these two simulations agree and differ but, while the MOH might be a good target, the manuscript doesn't explain whether discrepancies or biases are due to better or worse performance relative to the real world. I recommend including more comparisons against observations and indicating whether discrepancies between MOFC and MOH are indication of better or worse performance by MOFC

- Model Evaluation, Line 156: *"Figure 2 shows temperature and salinity diagrams for the MOH, NZESM and all of our experiments."*

  This figure needs a clearer explanation as to what it shows. Specifically, there is no explanation as to what the color represents. It's not until much later in the text that Figure 2 is described as a "histogram" (Line 232). The units are confusing without further explanation.

- Model Evaluation, Lines 156-157: *"MOFC has a fresh bias relative to MOH of about 0.2 PSU at temperatures between 2°C and 12°C. At lower temperatures the salinity bias becomes smaller, while at higher temperatures the salinity bias increases to about 0.5 PSU (at 18°C)."*

  How are the biases of 0.2 and 0.5 calculated/determined? The bias at 18°C is also a fresh bias but it's confusing as written (salinity bias increases" without referencing direction - fresh or salty). Also, are the geographic regions analyzed held constant between NZESM and MOH/MOFC? A consideration since they are different grids.

- Model Evaluation, Lines 161-168: *"Maps of the mean and standard deviation of the sea surface temperature and sea surface height are shown … MOFC shows variability about 50% larger than MOH"*

  Regarding the sea surface height comparisons, the authors point out the differences between MOFC and MOH but it's not clear if the historical MOFC is an improvement on, or worse solution than the MOH reference. It would help to compare against satellite sea surface height products to give a better sense of the new model (MOFC)'s performance. Additionally, several of the sea surface height features highlighted seem rather subtle compared with the negative SSH values depicted in the MOFC in the southeastern corner of Figure 3b. The coincident high standard deviation (yellow blob) in Figure 3f is also quite striking but there's no mention of this in the text. What's going on there? Is this a boundary artifact?

- Model Evaluation, Lines 188-193 *"The first EOFs … disagreement in finer spatial scale features"*

  It's not clear what distinguishes "good agreement" and "reasonable agreement" - these are qualitative terms and need more quantitative support.

- Model Evaluation, Lines 208-211: *"Linear trends in domain-wide OHC are given in Table 1. Surface ocean (0-200 m) warming is in good agreement across all three models. In the 200-1000 m range, we see strong warming in MOH but little warming in either MOFC or NZESM. Between 1000 m and 2000 m, the MOFC shows about half as much warming as MOH or NZESM, and below 2000 m MOH displays significantly stronger warming than either MOFC or NZESM."*

  I don't think it is appropriate to compare the trends reported in Table 1, specifically commenting on whether they disagree or agree. First, the time periods analyzed are not of consistent duration across modeling products: 27 years (MOH), 20 years (MOFC), 24 years (NZESM). None of these meet the criteria of the World Meteorological Organization of a "climate normal". Trends over shorter time periods can be more strongly biased by interannual variability (e.g., ENSO). The MOH lines in Figure 5 exemplify how the trend for a timeframe can be dominated by smaller stretches of years (e.g., Fig. 5a, blue line, yrs 1994-2000). The authors state that the MOH and historical MOFC simulation are not directly comparable temporally for this reason (Line 186: *NZESM may not produce an El-Nino in the same year as the real world an MOH*). While the trend comparison is not a direct temporal comparison, it does approach that sort of comparison when too-short a time-frame is considered. I recommend removing the Table/ these lines from the text.

- Model Evaluation, Lines 213-214: *"Examining maps of OHC trend by depth range (not shown) shows that the difference in warming between MOFC and MOH is primarily a product of the NZESM boundary forcing"*

  I suggest including the referenced trend maps in the supplement or describing in greater detail the evidence for this conclusion (that the warming in MOFC is a product of the boundary forming). As it is, there is no support or explanation for this conclusion.

- Model Evaluation, Lines 216-229: *"Time-mean vertical sections of temperature... with the exception of the SLC"*

  Again, I don't think it's appropriate to assess/validate a 16-year mean state of the historical MOFC relative to MOH/observations with any sort of statement regarding bias or performance because these aren't climate time frames and the MOFC is not forced with realistic/reanalysis forcing conditions. Per the manuscript Line 186, we wouldn't expect them to reproduce on short time extents since their atmospheric forcing conditions may have different interannual variability. Also, the time spans of the "Other Studies" isn't reported so it's hard to say whether MOH should match closely too.

- Future Climate Projections, Line 232: *"Temperature-Salinity histograms"*

  This is the first mention of these plots being histograms. More clarity is needed in describing how the histogram aspect of the plot is depicted in the historical and difference plots - there is no explanation and the units are difficult to decipher.

- Future Climate Projections, Lines 232-238: *"We see a shift towards warmer and fresher conditions predominately within the Subantarctic Mode Waters and Subtropical Mode Waters… We see no significant changes in the lowest segment of the T-S curves, corresponding to intermediate and deep waters."*

  The TS positions of the mentioned watermasses need to be indicated in the figures, otherwise there's no context for the reader to follow the authors' descriptions of shifts or changes in the watermass representation of the model.

- Future Climate Projections, Line 237: *"We see no significant changes"*

  How is significance quantified/ determined?

- Future Climate Projections, Table 3

  I recommend testing and reporting statistical significance for these and Table 1 trends.

- Future Climate Projections, Lines 252-259: *"Indeed, the time-series of future climate OHC in Figure 5… are comparable to reference period warming"*

  Comparing trends among timeseries comprising different numbers of years can be misleading. As pointed out earlier - the reference period in the MOH simulation is longer than that of the reference time series of the MOFC (26 vs 20 years). A few years can contribute substantially to the overall linear trend - the additional years at the end of the MOH simulation look to contribute considerably to the positive MOH surface trend in Figure 5.

- Future Climate Projections, Lines 261-262: *"In the surface ocean (Figure 8a-d) there is warming over the entire domain"*

  Given the variability in the surface values (Fig. 9a), I recommend testing and reporting whether the warming trends are statistically significant. It's not clear that the warming signal would exceed the noise for all regions shown in the panel.

- Future Climate Projections, Lines 269-270: *"Times-series of OHC (as percentages of the 1991-2010 mean value)..."*

  It's not clear why sometimes the analyzed MOFC reference period is 1994-2010 and other times the full 1991-2010 period is used. Does this have something to do with not including "spin up" years? If so, this should be applied consistently for both reference and future projection simulations. In contrast, the future periods appear to always use the full duration of the simulation, which means they would include the "spin up" effects at the beginning of their respective simulations in any analysis. I recommend clarifying the rationale for this methods. Alternatively, perhaps the 1994 start year is used only for consistency with the MOH simulation? Clarity is needed.

- Future Climate Projections, Lines 269-270: *"At deeper levels we see the ACC warms much more rapidly"*

Does “more rapidly” refer to the slope or the offset of the red line in Fig 9b-d? If it’s slope, it’s not clear to me that it’s significantly steeper than the Tasman Sea (blue line in Fig 9b). I recommend quantifying/ statistically evaluating these trends to support the assertions made.

- Future Climate Projections, Line 281: *“The SLC and FLC show little evidence of change under all future climate.”*

  I think it’s worth nothing that the near-shore bias in the SLC exhibited by the MOFC (relative to MOH) is exacerbated in the 2070-2100 ssp2 4.5 scenario (Fig10f). The change on its own may not be relatively large but, if it’s further exacerbating a bias. This could have important implications for the utility of this model configuration in particle tracking studies.

Technical Comments

- Introduction, Lines 76-78: *“In this paper we validate a 16 year period (1994-2010) of MOFC forced with historical NZESM against data from MOH and then explore regional scale medium-term (2030-2060) and long-term (2070-2099) marine climate change under the CMIP6 SSP2-4.5 and SSP3-7.0 emissions trajectories.”*

  Are both the medium-term and the long-term timeframes 30 years long? Because the referencing convention is not consistent between the medium- and long-term time slice experiments. If the year after the hyphen is included, it would be 31 years for the medium-term and 30 years for the long-term simulations. If the year after the hyphen indicates "up to but not including" said year then the medium-term would be 30 years and the long-term would be 29. Additionally, the convention for indicating the years for the long-term time slice experiment is not consistent throughout the text. This is the only instance in the text where the end year is “2099”. I recommend checking the manuscript text for consistency.

- Model Evaluation, Lines 194: *“We conducted a similar analysis for super-annual frequencies and sea surface height”*

  Should this be “in sea surface height” instead of “and sea surface height”

- Model Evaluation, Line 26: *“Time-mean vertical sections of temperature, salinity and cross-section velocity were computed for the lines shown in Figure 1”*

  I recommend specifying “lines” as “transect lines” for clarity.

- Model Evaluation, Table 2

  It’s difficult to line up the entries in the “Other Studies” column with the model results. I suggest adding some spacing to the rows and use consistent vertical alignment across columns for improved readability.

- Future Climate Projections, Figure 9 Caption: *“Time-series of Ocean Heat Content changes (as a reference period mean OHC) for the Tasman Sea”*

The parenthetical text is confusing, I suggest rewriting

- Future Climate Projections, Figure 9

  Improve figure for readability - the dashed lines for SSP2 4.5 are difficult to make out in all panels but especially in panel a where the shaded overlapping regions make the end-of-21st century simulations blend together

- Future Climate Projections, Figure 10

  I recommend reducing the range of the colormap. Currently, the data only utilizes the paler portion of the colormap and it makes interpretation more difficult.

- Future Climate Projections, Line 288: *"The occurrence of marine heat waves was investigated using a similar approach used to (Hobday et al., 2016)"*

  Check grammar/ sentence structure. Maybe "...used *by* Hobday et al. (2016)" ?

- Future Climate Projections, Line 324: *"The average percentages of the domain experiencing MHWs"*

  I recommend clarifying that this is a time average.

- Figure 1
  - **Color schemes:** Since the bathymetry is portrayed in greyscale, I recommend changing the color schemes for isobaths and the New Zealand landmass.
  - **Labeling conventions:** I recommend changing the labeling convention for fronts so they are distinguishable from currents to avoid confusion. As is, it appears as though the label for the sub-tropical front (STF) refers to the eastern-most blue arrow. This is quite confusing when the text details aspects of the STF region being to the southwest of the Southern Island.