

Dear Dr. Yool and dear Reviewers,

We appreciate your time and effort dedicated to reviewing this manuscript. Below are our point-by-point responses to the reviewers' comments. Please note that new figures were added (Figure 3,17,18, A3), which resulted in the renumbering of subsequent figures. Relevant changes in the text are marked with 'main changes' tags throughout this document.

Regards,

Dalton Kei Sasaki

reviewer 1

This manuscript presents and evaluates MOM6-COBALT-NEUS25 version 1.0, a high-resolution (1/25 deg.) coupled physical-biogeochemical model for the northeastern US designed to support marine carbon dioxide removal assessments and benthic modeling applications. The configuration consists of the MOM6 ocean model coupled with COBALT using a z^* coordinate with 75 vertical layers, incorporating tidal forcing, river discharge, and multiple data sources (Glorys, ERA5, TPXO9, WOA23) for initialization and boundary conditions. A thorough validation against observational databases, satellite products, and climatologies is presented for the simulation period in order to assess model performance. Key findings demonstrate that the model realistically simulates Gulf Stream dynamics, seasonal temperature and salinity structures with small SST biases, and realistic surface chlorophyll and nutrient seasonal cycles with expected spring and fall peaks. The configuration successfully captures regional circulation patterns and the development of cold SST patches through the inclusion of tidal mixing, an improvement over current reanalysis products. The work represents a significant advancement in regional ocean modeling by providing improved spatial resolution and process representation compared to global reanalyses, offering a robust foundation for climate impact studies and fisheries applications in this ecologically and economically important region. Some limitations include persistent biases in the Mid-Atlantic Bight's vertical temperature-salinity structure, a northward bias in Gulf Stream position with reduced meandering that affects tracer field representation, weak coastal freshwater signals leading to positive salinity biases, underestimation of mesozooplankton biomass, and limited skill in capturing subinertial sea surface height variability— however, many of these shortcomings arise from known modeling limitations and not necessarily from poor configurations choices. On the contrary, the authors have done a good job in with their parameter choices to ensure the best representation for their desired applications. The authors also present evidence on how these biases do not precluded the intended purpose of the configuration in its totality. This manuscript would be an excellent addition to the growing body of work on high-resolution regional modeling using MOM6. I would recommend this manuscript for publication after the authors consider the following comments:

1. The authors attribute the warm shallow water temperature biases during summer and fall partly to "overmixing associated with the upper ocean boundary layer scheme". Whereas Ross et al. (2023) (using ePBL alone) found that adjusting the submesoscale restratification front length parameter in the Fox-Kemper et al. (2011) scheme represented a key trade-off between mixed-layer depth and bottom-

temperature bias in their configuration. Why was the hybrid approach selected over ePBL and the Fox-Kemper restratification scheme?. (1) what motivated the choice of the hybrid max(ePBL, KPP) scheme over ePBL with the Fox-Kemper restratification (as in Ross et al. 2023)? , and (2) was the sensitivity of bottom temperatures to the restratification front length parameter identified in Ross et al. (2023) explored in NEUS25? (it is unclear from lines 6210-623 which parameterizations were tested).

Re: Both ePBL and Fox-Kemper were developed for lower resolution models. In Ross et al, the model resolution we used was 1/12th. At 1/25th, we are beginning to resolve some of the features (e.g., Langmuir Cells), for which that parameterization was developed and therefore we took an empirical approach to parameter space exploration without the expectation that what worked in Ross et al is necessarily the best choice at a higher resolution but rather a starting point. Our choice was based on the smallest global errors considering RMSE, mean bias, absolute mean error and pointwise correlation of different configurations that were tested. The configurations were evaluated through surface temperature and bottom temperature fields . During our calibration phase, we tested extensively different model configurations including Fox-Kemper restratification and pure ePBL/KPP configurations, which yield relatively small differences in our results and did not show significant impact in the bottom temperatures. These results were temporary and were purged by our HPC scratch system, hence we are unfortunately not able to present them.

main changes:

- section 3.1/paragraph#4/line 217-232 (diff.pdf)

2. The time step specified in line 201 is inconsistent with the time step in Table 1. Please clarify.

Re: We apologize for the inconsistency, the correct time step is 1800s. We changed the text accordingly.

main changes:

- section 3.1/paragraph#2/line 201

3. Lines 603-605: The authors could compare with a higher resolution SST product like OSTIA which may improve known biases in OISST:

https://data.marine.copernicus.eu/product/SST_GLO_SST_L4_REP_OBSERVATION_S_010_011/description

Re: Thank you for the comment, you are correct. OSTIA is indeed a product that provides more nuances to the spatial description of average SST in the Gulf of Maine and Georges Bank, as shown by the modified plots in Figure 5 and Figure A2. The SST overestimates in MOM6 vs OISST in Bay of Fundy are largely improved, while Glorys vs OISST shows a more profound difference (up to 3°C) in the same locations. One hypothesis for this effect is that tides in the area play a significant role in vertical and horizontal mixing associated with the macrotidal regime in Bay Fundy and tide-topography interactions in Georges Banks. We replace Figure 6 and changed the text accordingly (including caption).

main changes:

- section 4.2.1/paragraph#1-3/lines 383-408 (diff.pdf)
- section5/paragraph#3/lines 654-664 (diff.pdf)

4. The colormap chosen for Figure 3 is inappropriate with the white background as the light blues are hard to distinguish from white.

Re: We apologize for the colormap. We changed it for clarity. We also changed the caption following Reviewer #2 comments.

main changes: section 4.1/Figure 4 (diff.pdf) (former Figure3.)

5. I would suggest the authors to review when/where they are referencing figures. For example, the paragraph starting in line 366 does not reference any figure. I am aware that the paragraph is still discussing Figure 3, but I believe it is good practice to add an additional reference in cases like this.

Re: We reviewed the figure references throughout the text.

6. Several figures are lacking panel numbering which would make in-text referencing (which also needs revisions) much easier and effective for readers.

Re: Thanks for this feedback - we changed the panels throughout the text as recommended and will use this take this suggestion in future works.

7. Figure 7 needs to be revised: top colorbar is unreadable, panels are not numbered and the shared axes don't really work with the offset applied to each column.

Re: Thank you for the comment, we corrected both colorbar and panel numbers.

The number of the figure is now 8. Regarding the axes, this cascade subplot division we selected has lon/lat positions referenced in the left-most plot only. We made this choice to make the figure more compact, appealing to the reader and to make

visual comparison straightforward. For this reason we would like to leave the axes presentation as it is.

8. Mooring/ADCP IDs in Figure 1 are different to Figure 4.

Re: We corrected the IDs and made them coherent across the text and figures.

Figure 4 is now Figure 5.

reviewer 2

This paper describes a new simulation of an ocean model representing the North East US Continental shelf at $1/25^\circ$ resolution, coupled to a biogeochemical model. Most of the paper is devoted to documenting the fidelity of the simulated fields relative to various observational products. These comparisons are fairly comprehensive, quantitative and clear; the figures representing the model-observation comparisons are also clear and attractive. Therefore, the paper will serve as a valuable reference for future work that will then use the model to address scientific questions. I support the publication of the paper after the authors address a few questions/suggestions below.

I just have two somewhat larger comments and then a few more minor points below that.

Major:

First, I think the authors should spend a little more time describing some basics of the model initialization and boundary conditions in Section 3.1. Though I found this section helpful (and Table 1 usefully describes much of the model configuration), I found a few simple and fundamental details absent. For instance, I can see that the model was initialized from Glorys in 1993 and used Glorys to provide boundary conditions. I suppose that means that the MOM6 simulation described here was also run from 1993-2019, but I can't find this simple fact stated clearly. Also, is the 6-month "spin-up" part of this period, but then excluded from the comparisons, or somehow external to this period? More importantly and with respect to boundary conditions, it is important to explain if there is any freshwater or salinity nudging to avoid drift. A couple of decades is plenty of time for the mean salinity (and even sea level) fields in the model to drift if the mass budget isn't balanced (i.e. evaporation + precipitation + net freshwater flux across model boundaries). Section 3 should explain whether/how the freshwater and overall mass budgets are balanced and, if not, whether that has led to domain-wide drift in salinity or sea level.

Re: Our nudging for temperature and salinity fields focuses solely on correcting rim artifacts associated with the open boundaries, as written in Section 3 (Methods). We included a time series evaluation of the weight-averaged salinity for MOM6 and Glorys (the same domain) that demonstrates the salinity drift in both cases is consistent and likely physical (Figure A3). Imbalances in freshwater flux may be present in MOM6, but the similar behavior in both model and reanalysis indicates that the salt budget is balanced for the duration of the experiments. MOM6-COBALT-NEUS25 generally doesn't present any noticeable undesired drift. MOM6 is designed as a finite-volume numerical model that prioritizes the strict conservation of properties to be balanced, which contrasts with other

regional models that require specific volume conservation flags to keep ssh and salinity constrained. We have not included the overall mass budget, since ssh doesn't present spurious trends during the simulation period and salinity has similar patterns to glorys. We also included an extra paragraph with initialization details.

The spin-up of the model is relatively short due to the relatively shallow areas on the shelf and the strong currents associated with the Gulf Stream system. Chlorophyll, nutrients and carbon related fields adjust in roughly three months from the climatological fields. Since the initialization takes relatively little time to adjust, we included this period in our evaluation. The complete period of simulation was also included.

main changes:

- *section 3.1/paragraph#1/line 198*
- *section 3.1/last paragraph/lines 267-271*
- *appendix/FigureA3*

Second, one rationalization given for developing this new simulation was to have a tool to evaluate mCDR processes, but no comparison of any simulated fields against the observed carbonate system was offered. Given that there are GLODAP climatologies of TA, DIC and time-varying surface pCO₂ fields derived from SOCAT data, I think this omission is strange and should be well-rationalized or remedied by including one or more carbon system comparisons.

Re: We appreciate the comments on the mCDR processes. We haven't included direct evaluations of the TA and DIC in the first submission, because our initial assessment was that gridded climatologies with resolution of 1 x 1 deg were too coarse. In a complex and semi-enclosed region such as the Gulf of Maine, these products don't likely capture the inorganic carbon dynamics and could be biased towards open ocean values. After this request we added a DIC comparison between our model and a climatology based on a climatology based on Neural Network product and GLODAP and also a long-term climatology of TA for the continental margin of the U.S. We haven't included the Neural Network product for TA validation because we weren't able to obtain this particular dataset from the database reposit.

main changes:

- Section 3.2/last paragraph/lines 323-331
- section 4.3/ Figure 17 and caption (diff.pdf)
- section 4.3/second-last and last paragraph/lines 642-656 (diff.pdf)

Minor:

Figure 3: This depiction does not at all look like streamlines. Rather, some of the lines seem to end abruptly (not possible for the non-divergent velocity field for which a stream function is defined). Also, the “current speed” given in colors does not clearly correspond to the closeness of the streamlines, as would be expected. I suspect this is not a classically-defined stream function and should be clarified or replaced.

Re: We apologize for using the term streamline loosely as it is a fundamental description of the non-divergent field in physics and Geophysical Fluid Dynamics. We included a more strict and formal definition of the parameter used to define the lines. Figure 3 is now Figure 4.

main changes:

- section 4.1/Figure 4 + caption (diff.pdf)

Figure 1 and Figure 4 - It's not clear which stars correspond to which ADCPs. In other words is A01 = A? Why are there numbers in Figure 4 and not in Figure 1?

Re: We apologize for the inconsistency. 'A' and 'A01' represent the same mooring. We corrected all instances of moorings to be represented by letters only. This correction is made throughout the text and Figures.

Line 385 - Here is where part of my “major point #1” above could be inserted. Are salinities allowed to drift unrestrained? Is the whole domain getting fresher? Saltier?

Re: We included the appropriate figure and text as requested in a previous question. The salinity drift is consistent with Glorys’.

main changes:

- section Appendix A/Figure A3 (diff.pdf)

- section 4.2.1/last paragraph/lines 413-416. (diff.pdf)

Figure 8 caption says “days” for the bottom right panel, while the colorbar is labeled “months.” These should be consistent

Re: We corrected Figure 8 and made minor modifications in the caption.

main changes:

- section 4.2.3/Figure 9 (diff.pdf)

Figure 11 caption and section 4.2.5: I would guess that the data must be detrended before performing EOF analysis, otherwise the warming trend would dominate the PCs. If they are detrended, that information should be added to the caption and/or text.

Re: We acknowledge that EOF analysis formally requires stationary data. Although the warming trend is small relative to interannual variability, we have repeated the analysis with detrended data to comply with this requirement. The resulting EOF maps and PC time series are nearly identical to the original, suggesting that the trend had negligible influence on our findings. We modified the figure and included a change in the Figure caption to clarify that the data are detrended.

main changes:

- section 4.2.5/Figure 12 + caption (diff.pdf)
- section 4.2.5/all paragraphs

Figure 12: I am curious (and suspect other readers would be too) why satellite SSH is not used as an observational field against which to compare simulated SSH. It would be good to explain this choice somewhere.

Re: The main reason for omission is our understanding that the Gulf Stream signature is likely the most relevant process that is evaluated in terms SSH (apart from tides). This was evaluated using the SSH standard deviation. Since the reviewer raised this point, we decided to include a new Figure 3 and included a paragraph in the text.

main changes:

- section 4.1/paragraph#2/lines 347-354 (diff.pdf)
- Section 4.1/Figure 3

Figure 14 discussion, around line 555 - I don't find the explanation explaining the high chlorophyll in the Gulf Stream convincing and it would be good to either back it up with direct evidence (not necessarily a new figure, but at least a clear assessment of Gulf Stream nutrient concentrations and gradients at its edges). At 0.25° resolution, WOA has ~4 points across the 100 km Gulf Stream at its upstream boundary. I've previously looked at the nutrient fields in WOA and their lateral gradients and recall that the fronts at the edges of the Gulf Stream seem to be present.

Re: We appreciate this comment, since it enabled a better understanding of the issue. We reassessed again the model results and the boundary conditions. The global COBALT simulation (Stock et al., 2014) that provides biogeochemical parameters to our domain is averaged in time (1993-2014), which is the same approach used in MOM6-COBALT-NWA12 (Ross et al., 2023). This introduces a constant supply of phytoplankton over time to the area influenced by the Gulf Stream. Since the Gulf Stream is fast, it doesn't allow the phytoplankton to adjust to summer conditions. A second aspect is that Stock et al. (2014) model has a resolution of approximately one degree in the region, which doesn't properly resolve the Gulf Stream across its width and could influence our results as well.

main changes:

- section 4.3/paragraph#7/lines 599-609 (diff.pdf)

Line 607. I appreciate that throughout the paper, the authors mostly discuss comparisons quantitatively, eschewing descriptors. However, here “low bias” is stated with no quantification. Adding a few quantifications in parentheses here would be appropriate.

Re: We appreciate the comment. For the reviewer's information, we tried to simplify the discussion by leaving specific quantities in results (i.e. Section 4.2.4, where the model is compared with NERACOOS). Hence, we removed the 'low bias' statement, so the paragraph is consistent with our text style in the discussion. In any case, the average bias across all moorings and available instruments when comparing with model results are 0.13 ± 0.49 °C for temperature and -0.38 ± 0.47 for salinity.

main changes:

- section 5/paragraph#3/lines (diff.pdf)

Overall, the discussion and conclusion sections are brief, but this is a long paper with a relatively simple set of objectives, so that seems fine with me. On the other hand, the discussion is almost exclusively focused on temperature, and, briefly, the relationship of temperature with chlorophyll and zooplankton. The omission of any discussion here of salinity or sea level seems notable. Adding 1-2 paragraphs to explain the fidelity (or differences) between these salinity and sea level changes with observations would be appropriate. If the paper comes to include carbon system comparisons, as I suggest would be beneficial, then those should be synthesized here too. Finally, the conclusion reflects on the skill in the Gulf of Maine and new opportunities for interrogating the system using this simulations there. It would also be useful to readers to understand limitations of the model and where attention could be focused to improve it.

Re: These comments guided us towards improving our discussion. We included a few paragraphs about salinity and sea level height, DIC and TA. The conclusions were also updated to link limitations with future improvements.

main changes:

- *section 5/paragraphs#7,8/lines 6704-724 (diff.pdf)*
- *section 5/last paragraph/lines 738-749 (diff.pdf)*
- *section 6/second-last and last paragraph/lines765-781 (diff.pdf)*