Response to Reviewer-2 Comments:

This manuscript presents a comprehensive evaluation of the MOM6-NEP10k model's performance over the Bering Sea, along with sensitivity tests on turbulent mixing parameterizations. The work is rigorous, well-motivated, and makes valuable contributions to regional modeling, particularly in high-latitude domains where ocean–ice interactions and cold pool dynamics are essential for ecological forecasting.

The paper is timely and technically sound. However, the manuscript could be strengthened through greater clarity in presentation, more quantitative analysis in some sections, and deeper contextualization of the results, especially with respect to model biases and sensitivity test implications.

We thank the reviewer for his/her insightful comments.

Line 20. Clarify "retreat earlier (later) in cold (warm) years"—this may seem counterintuitive. Explain mechanism briefly.

We have revised the sentence in the abstract to briefly explain the mechanism (Lines 20-22): "*the model captures the mean timing of sea-ice retreat, though it tends to retreat earlier in colder years and later in warmer years compared to observations*". This is probably because the model underestimates melt rate sensitivity to ice thickness. In reality, thicker ice melts slower, thinner ice melts faster.

Line 56. "northeasterly winds in winter" → add citation or mean wind speed for reference.\

We have revised the sentence to include a mean wind speed during winter and cited recent studies (Lines 59-61). "*The strong northeasterly winds in winter, with a mean speed of ~9 m s^{-1} (Danielson et al., 2014; Stabeno & Bell, 2019), carry cold air from the Arctic, resulting in the formation of sea ice in the polynyas*".

Line 86. Abbreviate NOAA

We revised (Line 90-91) it to "*National Oceanic and Atmospheric Administration (NOAA)* " and used "NOAA" for the rest of the manuscript.

Line 94. Please provide citation that reports the excessive shear-driven vertical mixing found in MOM6-NEP default configuration.

We added a new figure (Fig S2,) in the supplementary material to show the excessive shear-driven vertical mixing in the control simulation.

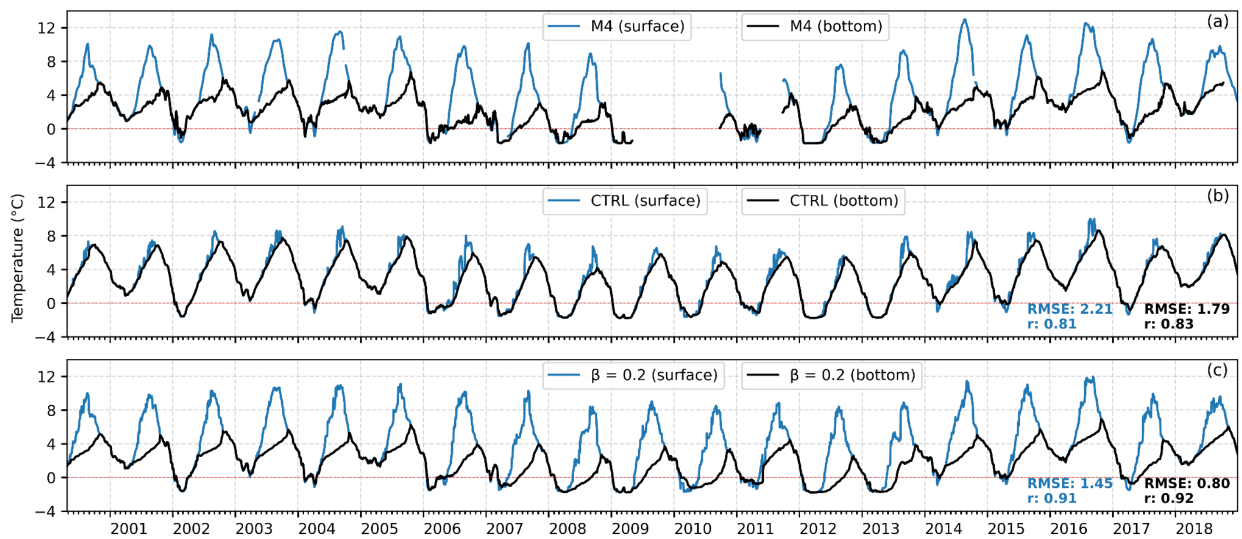

Figure S2: Comparison of surface and bottom temperatures at (a) M4 mooring location, against (b) CTRL (default MOM6-NEP) and (c) β =0.2 simulations. This comparison shows the excessive shear-driven vertical mixing in the CTRL simulation.

Line 95. The physical justification for the particular choices of scaling factors could be expanded. What guided the specific values tested? Clarify whether these fall within observed oceanographic ranges or are purely numerical tuning.

In the revision, we inserted a reference to "Section 3.1" which provides more information (including a new reference to observed diffusivity). In summary, it is numerical tuning guided by moored water mass observations and estimates of vertical diffusivity in the world's oceans from limited observations. See more details in Lines 241-250, 267-269.

Lines 118-119. It could be confusing to the readers to mention the boundary that connects to the Pacific as "western" when most of it appears to be a "southern" boundary.

You are correct - on the model's native grid, this side appears to be the southern end of the model domain, but this is feature of the curvilinear coordinate and that the model grid is not oriented along constant longitudes/latitudes, "western" and "southern" become ambiguous. To clarify, we changed the text (Lines 133-134) to "*the longest of which arcs through the Pacific Ocean and is here referred to as the western boundary as it is adjacent to the north-western Pacific".*

Line 125. GLORYS12 is said to perform well in coastal areas—consider supporting with performance metrics from the cited studies.

We added a statement describing GLORYS performance metrics. The revised text (Lines 143-145) now reads: *"In Amaya et al. (2023), GLORYS12 was shown to have SST RMSE values of ~0.25–0.6 °C and monthly SST anomaly correlations > 0.9 at nearshore locations along the west coast.".*

Line 140. Can you mention the reasons/justifications for using a 3-hourly product vs 1-hourly product of ECMWF for atmospheric forcing?

We suspect that atmospheric variability on hourly versus 3-hourly time scales don't differ significantly, and both of which capture the diurnal cycle. We have clarified the reason for using JRA55-do in the revised manuscript (Lines 160-162) as *"JRA55-do (Tsujino et al., 2018) is specifically designed for ocean–ice modeling applications. It provides consistent atmospheric variables for bulk flux calculations, applies bias corrections to prevent artificial trends, and is compatible with CORE/OMIP forcing protocols".*

Lines 305–360: "...model tends to underestimate sea-ice cover...melts all ice at the same rate..."

This is an important insight into model limitation. I'd suggest incorporating references to ongoing work or methods that include age-dependent melt rates. Could this be easily incorporated in future iterations?

We clarify that SIS2, the sea-ice component used in this study, does include a multi-category ice thickness distribution scheme. Hence, the melt rates are not modeled uniformly; instead, thinner ice categories can melt out faster, while thicker ice persists through the melt season (Adcroft et al., 2019). The remaining biases in the simulation, as shown in Fig. 8(a, c), suggest still imperfect melt rates in the model, and potential limitations in other physical processes such as snow insulation, melt pond dynamics, or atmospheric forcing errors. We have revised the text to reflect this clarification (Lines 389-395).

"While SIS2 uses a multi-category ice thickness scheme that allows thinner ice to melt more quickly than thicker ice (Adcroft et al., 2019), some biases still arise due to unresolved processes such as melt pond formation, snow effects, or errors in atmospheric forcing. Addressing these processes may improve model performance in future configurations".

Line 310: "mean ± std dev of areal ice extent" → good summary; you might also include coefficient of variation to show relative interannual variability.

Thank you for the suggestion. We have added the coefficient of variation (CV) alongside the mean and standard deviation to highlight relative interannual variability. The CV is similar for MOM6 and satellite data (~21.5% and ~19.7%, respectively), indicating similar relative variability despite the difference in mean ice extent.

Revised text now reads (Lines 338-341):

The long-term (1993–2018) mean ± standard deviation of areal ice extent in March (the month with maximum ice) is $0.65 \pm 0.14 \times 10^6$ km² in MOM6-NEP (CV = 21.5%) and $0.71 \pm 0.14 \times 10^6$ km² in satellite observations (CV = 19.7%). These similar CV values suggest comparable relative interannual variability in both datasets.

Lines 365–400: "...agreement in summer was poorer...SST is highly sensitive..."

Consider plotting correlation maps of SST biases against ice retreat dates to support the hypothesis. A brief regression analysis could further validate the sensitivity relationship.

We removed the sentence in the revised manuscript.

Lines 405–415: “MLD biases in winter increase significantly…”

This is a critical point. MLD overestimation in winter could lead to errors in nutrient entrainment and ecosystem models. Include a brief discussion on whether this is driven by JRA55 biases or turbulent mixing choices.

This indeed is an intriguing result. Note that the large biases primarily happen along the shelf-break with steep topographic gradients (Fig. 11f). Winter observation data in this region are very limited, so the observational data used in deBoyer calculation have large uncertainties. We already discussed this in the original manuscript: *“MLD biases can be linked to errors in surface forcing, model physics, inaccuracies in numerical algorithms, and/or uncertainties in observations, but which factor is the main contributor is unknown. MLD and its seasonal evolution affect nutrient distribution and primary production, and is crucial to the marine ecosystem dynamics of the region. Further quantifying the modeled MLD and its spatiotemporal variability, and understanding the mechanisms contributing to its biases will be a research priority in the future.” (*See lines 439-443*).*

Lines 440–460. It would help to add a histogram of thermocline depth differences (model – obs) across seasons. Also clarify the threshold logic—what happens when vertical gradients are weak or noisy?

"Figure S3 presents the seasonal distribution of the thermocline depth bias, calculated as the model depth minus the observed depth. There is a distinct seasonal cycle in model performance. In winter (DJF), the bias is effectively zero, which is attributed to the detection algorithm assigning a consistent default depth (70 m) to both the model and observations under well-mixed conditions. As seasonal stratification develops, the model exhibits a slight shallow bias in spring (mean: -3.64 m) and summer (mean: -1.25 m). The largest spread in model error occurs during summer. By fall, as the water column destratifies, the bias becomes minimal (mean: +0.44 m), and the error distribution narrows considerably."

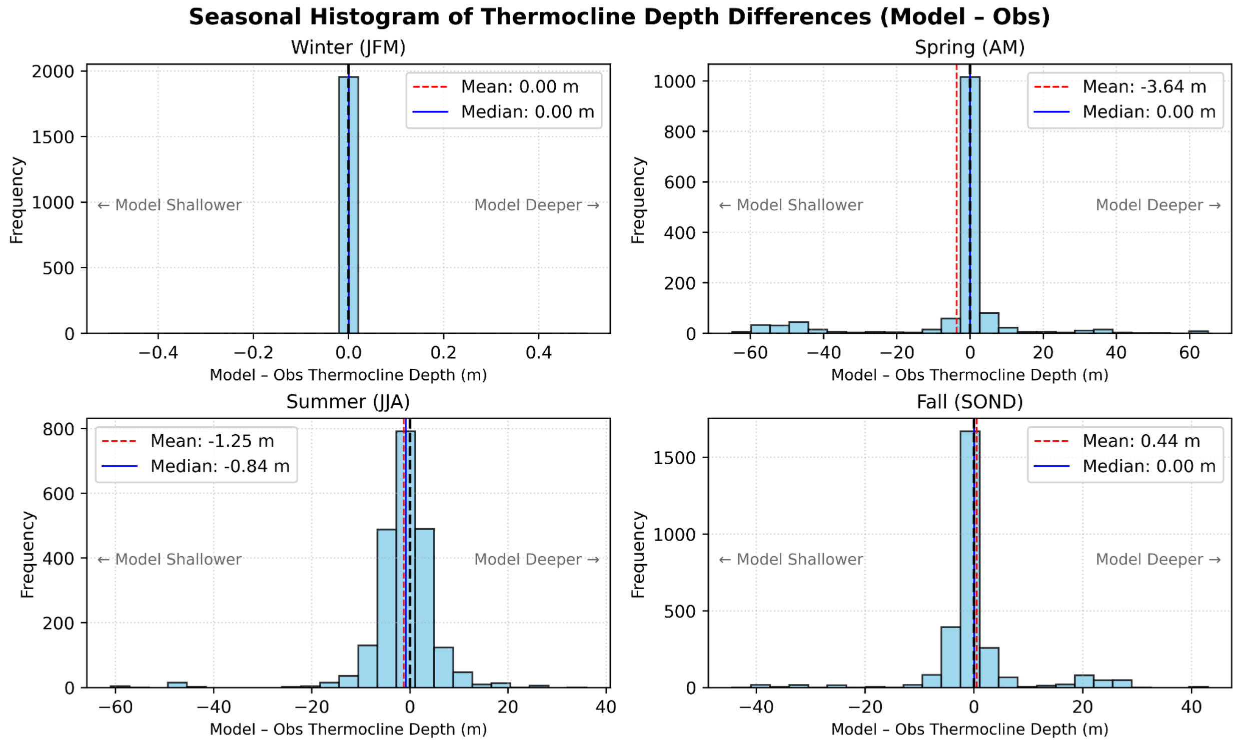

Figure S3: Seasonal histogram of thermocline depth differences (model minus observation) at M2 mooring location.

Regarding methodology, "*the thermocline depth is defined as the depth at which the maximum vertical temperature gradient (threshold > 0.1 °C m^{-1}) occurs; we consider temperature gradients within the 5-70 m depth range. If the gradient is weaker than 0.1°C m^{-1} throughout the water column, we assign 70 m as the thermocline depth*". This threshold logic has now been clarified in the revised manuscript Section 2.3 (Lines 186-189).

Lines 510–530. Consider adding a short discussion of potential ecological implications (e.g., fish spawning habitat or match/mismatch hypothesis) to emphasize why this matters.

In revision, we added the text (Lines 545-548): *"The cold water mass that persists near the bottom as a result of sea ice melting and insulation from the surface heating during the summer is referred to as the "cold pool". This feature limits the distribution of commercially important species such as pollock and pacific cod. It also provides a corridor for arctic species to move southward".*

New References:

Adcroft, A., Anderson, W., Balaji, V., Blanton, C., Bushuk, M., Dufour, C. O., Dunne, J. P., Griffies, S. M., Hallberg, R., Harrison, M. J., Held, I. M., Jansen, M. F., John, J. G., Krasting, J. P., Langenhorst, A. R., Legg, S., Liang, Z., McHugh, C., Radhakrishnan, A., Reichl, B. G., Rosati, T., Samuels, B. L., Shao, A., Stouffer, R., Winton, M., Wittenberg, A. T., Xiang, B., Zadeh, N., and Zhang, R.: The GFDL Global Ocean and Sea Ice Model OM4.0: Model description and simulation features, J. Adv. Model. Earth Syst., 11, 3167-3211, https://doi.org/10.1029/2019MS001726, 2019.

Amaya, D. J., Alexander, M. A., Scott, J. D., and Jacox, M. G.: An evaluation of high-resolution ocean reanalyses in the California current system, Prog. Oceanogr., 210, 102951, https://doi.org/10.1016/j.pocean.2022.102951, 2023.

Danielson, S. L., Weingartner, T. J., Hedstrom, K. S., Aagaard, K., Woodgate, R., Curchitser, E., and Stabeno, P. J.: Coupled wind-forced controls of the Bering–Chukchi shelf circulation and the Bering Strait throughflow: Ekman transport, continental shelf waves, and variations of the Pacific–Arctic sea surface height gradient, Prog. Oceanogr., 125, 40–61, https://doi.org/10.1016/j.pocean.2014.04.006, 2014.

Stabeno, P. J. and Bell, S. W.: Extreme conditions in the Bering Sea (2017–2018): Record-breaking low sea-ice extent, Geophys. Res. Lett., 46, 8952–8959, https://doi.org/10.1029/2019gl083816, 2019.

Tsujino, H., Urakawa, S., Nakano, H., Small, R. J., Kim, W. M., Yeager, S. G., Danabasoglu, G., Suzuki, T., Bamber, J. L., Bentsen, M., Böning, C. W., Bozec, A., Chassignet, E. P., Curchitser, E., Dias, F. B., Durack, P. J., Griffies, S. M., Harada, Y., Ilicak, M., Josey, S. A., and Yamazaki, D.: JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-do), Ocean Model., 130, 79-139, https://doi.org/10.1016/j.ocemod.2018.07.002, 2018.