Response to Reviewer-1 Comments:

This manuscript is focused on demonstrating the performance of an implementation of MOM 6 sea-ice ocean model in the Bering Sea. The authors provide a detailed comparison of the model with available observational data from 1993-2018. The manuscript is well written and organized. The figures are clean and easy to interpret. As a report on 'Development and deployment of MOM 6 in a regional configuration' the authors have done a good job, but as a manuscript on geophysical model development, it is not as clear. The principal model 'development' is an adjustment to the vertical mixing parameterization, but no physical justification for the change is provided. Numerous model/observation comparisons are performed, but for several of them where the model doesn't perform well, the authors just note that it is unresolved, rather than offer an explanation. I think for someone interested in using MOM6 in the Bering Sea with their configuration all the content is well presented, but a general reader would get frustrated particularly in the latter half of the manuscript, where a lot of detail is given mostly to say that in some places the model does quite well, in others it doesn't and we don't know why. In addition to my specific comments to address below, I would suggest the authors streamline the presentation by simply summarizing where more work is needed, maybe putting some of the figures that show unexplained, disfavorable comparisons in an appendix, (if they really want to keep them). But, I leave it to the editor's discretion to give guidance for this particular special issue, on how much of this content should be in the body of the paper.

We thank the reviewer for his/her insightful comments. As this reviewer has alluded to, some model "development" works are broader in scope, including those building a new physical parameterization, while others are more specific, such as those testing and improving an existing parameterization scheme. We think that detailed evaluation and sensitivity tests as presented in this study are important steps in model "development". We have carefully considered the reviewer's comments and streamlined our materials where possible. These revisions are detailed in our point-to-point responses below.

Scientific questions/comments

As mentioned above, my principal concern with this paper relates to the tuning of vertical mixing to better reproduce stratification and bottom temperature on the Bering Sea shelf. The law-of-the-wall scaling that is used to specify the mixing length near the bottom has quite a robust basis in observation and theory, The authors really need to point to some physical process that could cause a factor of 5 deviation from this and offer some supporting evidence. Altering the slope of the profile of shear driven vertical mixing at the bottom boundary impacts the profile throughout the water column (Figure 3) so the authors need to justify the change over the whole profile.

You are correct, the Jackson-Hallberg-Legg shear-driven mixing scheme (JHL scheme) considers non-local effects, as can be seen by its two physical equations:

$$\frac{\partial^2 \kappa}{\partial z^2} - \frac{k}{L_d^2} = -2SF(Ri). \quad (10)$$

$$\frac{\partial}{\partial z}\left[(k + v_o)\frac{\partial Q}{\partial z}\right] + \kappa(S^2 - N^2) - Q(c_N N + c_S S) = 0. \quad (11)$$

In this framework, eddy turbulent kinetic energy (Q) and eddy diffusivity (kappa) are interactive and feedback on each other, and the differential form of the equation suggests that changes in one vertical location can influence results in other locations (as can be seen in Fig. 3 of the manuscript, as the reviewer has pointed out). Within the bottom boundary layer where the water is unstratified, decay length L_d in (10) is controlled by L_z, the distance to the nearest solid boundary (i.e., the bottom of the ocean). By reducing the amplitude/scale of L_z, we still obey the "law of wall", which dictates that eddy diffusivity in this region increases linearly with distance away from the boundary, albeit at a slower rate (i.e. we attenuated the shear-driven mixing of the bottom boundary layer, and this attenuation helped to maintain the stratification between the bottom boundary layer and surface boundary layer). Eliminating L_z entirely would indeed violate the "law of wall". In the JHL scheme, within the bottom boundary layer, the vertical gradient of kappa is approximated as $\frac{\partial \kappa}{\partial z} \approx \sqrt{F_O}\, u*$; however, there are uncertainties in both F_o (shape equation when Ri=0) and u*. u*, the stress velocity, is a function of tau_o, stress within the boundary layer. It is conceivable that the highly tidally energetic regimes of the Bering Sea shelf generate large tau_o in our control simulation and sensitivity tests with tides, contributing to large kappa and shear induced mixing in both the bottom boundary layer and the interior. The default JHL scheme was developed for coarse resolution ocean models with large time steps where modeled ocean velocity and stress are likely underestimated, and it did not consider tides. When applied in a high-resolution regional configuration with tides, the scheme is found to produce too strong mixing which erodes the well observed sharp thermocline between the surface and bottom boundary layers. In addition, this gradient is typically considered proportional to the von Karman constant (as described in JHL), but is in fact influenced by surface roughness and/or the presence of suspended sediment, and hence allows for some tuning in our oceanographic setting. By rescaling L_z, we have not violated the "law of wall" but only reduced its vertical gradient. We have now clearly mentioned this in the discussion (section 4), the revised manuscript (lines 600-612)

A related question is, how does altering the vertical mixing scheme impact model performance beyond the eastern Bering Sea shelf? What impact does this have on the Aleutian throughflows or even along the US and Canadian NorthEast Pacific coasts? If the change only improves things on the Bering Sea shelf, is it sensible for the modified parameterization to vary spatially?

Rescaled L_z is also used in the MOM6-NEP configuration in Drenkard et al. (2025), who examined watermass structures in other regions across the NEP domain, including in the Gulf of Alaska and along the California Coast (see Fig. 17, in that paper). Results of Drenkard et al. (2025) suggest that the uniformly rescaled L_z has not caused water mass degradation in other regions of the NEP. A more in-depth analysis of flow through the Aleutian passes are currently underway, but it is beyond the scope of this study. It is possible that a spatially varying L_z could further improve model results. These research questions remained to be explored in the future.

A second scientific concern I have relates to the discussion of circulation. The entire section only seems to say that currents qualitatively agree with what we expect to see. There is no comparison with observation so I think this should require only a sentence or two at most, rather than a whole subsection. Additionally I question the robustness of the Alaska Coastal Current on the Bering Sea Shelf. The model mean circulation shows no clearly identifiable currents flowing northward over the 50 and 100m isobath in the only figure in the paper related to currents (figure 2b), nor do I know of any recent observational datasets that clearly delineate such currents.

We opt to keep this subsection (3.2) because it provides a circulation context over which the water mass structures at the mooring locations are discussed. We are planning a dedicated follow-up manuscript that will focus specifically on the Bering Sea circulations, where we can examine its characteristics and compare our model results with available observations in detail.

Fig. 2a is adopted from Stabebo et al. (1999, 2005), where the Bering Sea circulation pattern was constructed from observations. On Fig. 2a, the relatively weak ACC along the 50-m isobath (3-5cm/sec), flow along the 100-m isobath, and other major currents in the region, such as the Bering Slope Current (BSC) and Aleutian North Slope Current (ANSC) are outlined and named. A visual inspection and comparison of Fig. 2a vs. Fig. 2b shows that there is a close correspondence between each of these named currents in observations and modeled ocean currents. One might repeat the current names on Fig. 2b but doing so obscures the velocity vectors so we omitted the names on Fig. 2b.

My third scientific concern relates to the sea ice concentration comparison with observations. The authors offer a detailed description of sea ice comparison at two mooring locations. But as there are no sea ice concentrations measured at those moorings, why not provide a more comprehensive spatial analysis of model error and biases over the domain? Does the behavior at the two moorings exemplify the range of variation found over the domain?

Thanks for the suggestion. While we focused on the two moorings to highlight daily ice-concentration variability in detail, we agree that a domain-wide spatial validation is needed to place those results in context. To address this, we have added a new figure in the supplementary (Figure S1) showing the modeled versus NSIDC satellite mean sea-ice concentration climatology (December–June, 1993–2018) across the entire Bering Sea shelf. This spatial comparison demonstrates that the mooring locations capture variability and biases that are representative of much of the domain, while also revealing areas of slight overestimation along the northern shelf break and underestimation near the eastern shelf.

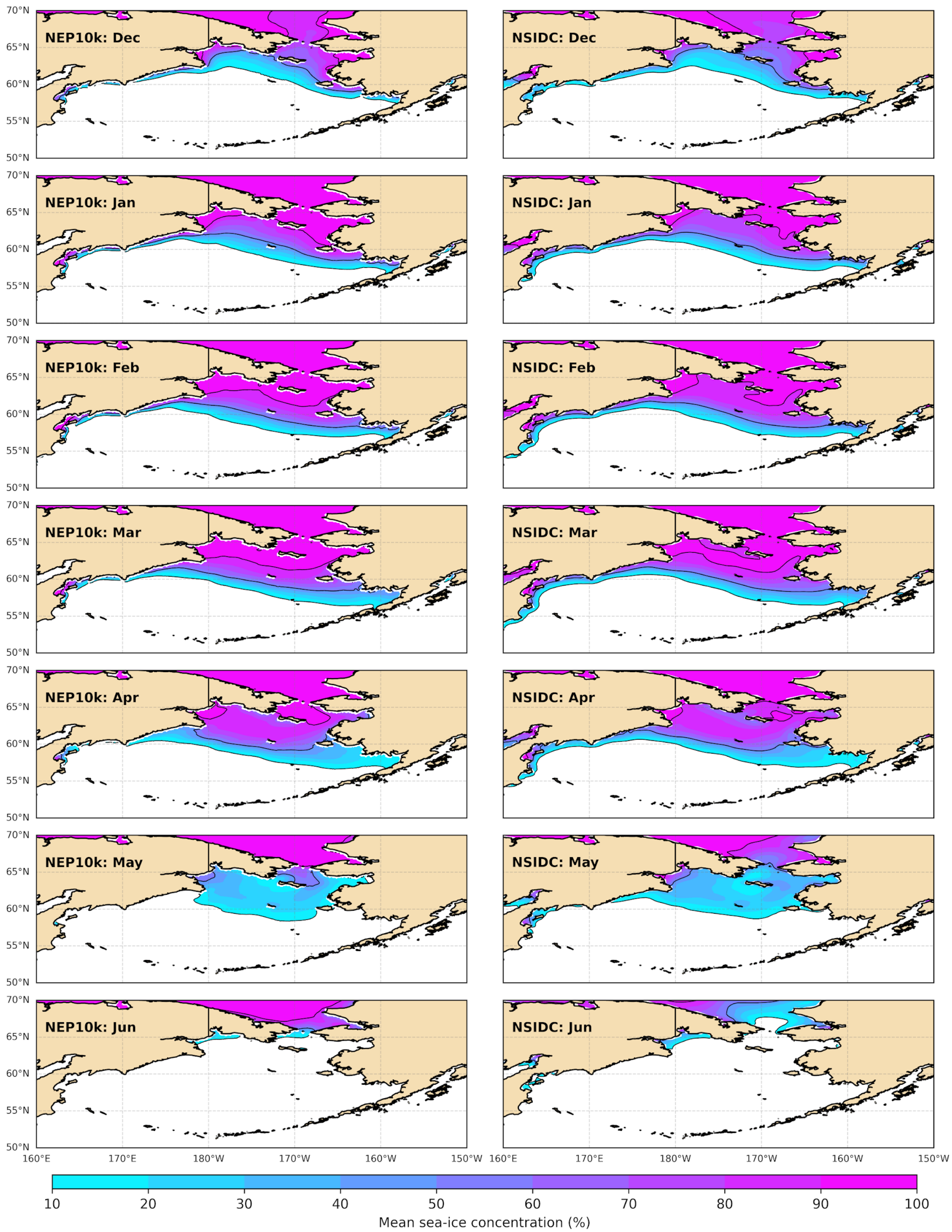


Figure S1: Bering Seasonal Sea Ice Concentration and Spatial Extent. Comparison of spatial patterns in Bering Sea monthly mean (1993-2018) NEP10k sea ice concentration against NSIDC Satellite output.

Particulars and clarifications (by line):

Line 30: More than 10% of the word's fish, half US.. etc. This statistic references publications from over a decade ago, is there a more recent assessment that confirms it is still the case?

Line 32: 37 million poinds Same reference as line 30?

We thank the reviewer for pointing this out. We have replaced the references with the latest data. The revised text now reads (Lines 31-34):

"More than 10% of the world's fish and shellfish and about 60% of the U.S. commercial seafood harvest come from these ecosystems (National Marine Fisheries Service, 2024; Alaska Seafood Marketing Institute, 2023). In addition, subsistence harvests are important for the livelihoods and cultures of many communities in Alaska. Each year, approximately 36.9 million pounds of foods are harvested by rural communities (National Marine Fisheries Service, 2024)".

Line 43: more oceanic in its vertical structure. What does 'more oceanic' mean?

To improve clarity, we have revised the sentence (line 45) to describe the outer shelf domain (~100–180 m depth) as *"which is more gradually stratified".*

Line 45 'but the domains exist...' I don't understand what this means.

To improve clarity, we changed the text to (Lines 46-48):

"These domains are most clearly expressed on the southern shelf, where tidal currents are roughly twice as strong as in the north, but similar cross-shelf structures extend from the Alaska Peninsula to St. Lawrence Island, a distance of ~1000 km."

Line 53 Western passes exit to the Pacific I think, not the Gulf of Alaska

Text in the revised manuscript is changed to *"North Pacific".*

Line 60 (Stabeno and Cheng, in prep) I would list this as 'personnel communication' rather than in prep., at least until their manuscript is submitted and accepted for review.

Agreed. The text is changed to "*personal communication*".

Line 93 A paragraph describing briefly (not necessarily comprehensively) what regional modeling has been done previously for the Bering Sea seems necessary in this section, so that the reader can put into context how this work adds to that.

Thanks for the suggestion. We added the following text and references to the revised manuscript (Lines: 98-105):

"*During the last 20 years, hydrodynamical modeling of the Bering Sea has been primarily based on the Regional Ocean Modeling System (ROMS). Hermann et al. (2013, 2019, 2021) used coupled physical–biological models, incorporating benthic components, to examine lower trophic level dynamics and ecosystem responses to climate variability. Kelley et al. (2020) used a high-resolution configuration of ROMS to explore the drivers of ocean temperature and sea-ice* *variability across the shelf. Simulations from these efforts revealed strong linkages between physical forcing and key ecosystem components such as large crustacean zooplankton. More recent studies have extended the scope of Bering Sea modeling toward climate downscaling and scenario projections, using coupled physical–biogeochemical models based on ROMS frameworks (Cheng et al., 2021; Hermann et al., 2021; Pilcher et al., 2022).*"

Line 110 maybe mention the distance the not-open boundary is from Bering Strait to further justify why it's not of great concern for this study?

We added the following text to the revision (Lines: 123-125):

"Since the primary region of interest in this study, the Bering Sea, is ~1200 km away from the model's lateral boundary, and sea ice in the Bering Sea is formed mostly locally, treatment of sea-ice boundary conditions is not expected to affect our results significantly."

Line 137 How are GLOFAS river discharges implemented? Presumably they could provide a freshwater flux at every coastal point. But one might also only consider major rivers.

GloFAS freshwater discharges at ocean-adjacent "pit cells" are remapped to the nearest MOM6 coastal ocean grid cells. "Pit cells" are GloFAS grid cells where the local drain direction indicates that only inward water flow occurs and is therefore a point of accumulation (e.g., lakes) or a point of egress to the ocean via either ocean adjacency or connectivity through other "pit cells" (e.g., wetlands). "*The full implementation is described in Denkard et al. (2025)"*, and we have now added a citation to the revised manuscript (see lines 154-155).

Line 143, The switch from ERA5 to JRA55 could use further explanation. What is it that makes JRA55 'specifically tailored' for driving ice-ocean models?

In the revision, we clarified the reason for using JRA55 (see 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".*

Line 143. Does Large and Yeager specify buoyancy and momentum fluxes over both water and sea ice?

We used the bulk formulae developed in Large and Yeager (2004) to calculate surface turbulent buoyancy and momentum fluxes from MOM6-NEP simulated SST, sea-ice concentration, and imposed atmospheric state variables from JRA55-do. In other words, we do not specify surface turbulence fluxes (ie, sensible and latent heat fluxes, evaporation, and momentum fluxes). This is described on Lines 163-164 of the revised manuscript.

Line 208 Provide a few more words describing the Jackson et al 2008 scheme. It's a scheme designed for climate models. Why is it deemed appropriate at these scales? Are other schemes available in MOM6? Were they tried?

Thanks for the suggestion. In the revision (Lines 229-231), we briefly introduced the Jackson et al. 2008 scheme by the following statements:

"Shear-driven vertical mixing in the interior ocean is modeled following Jackson et al. (2008). In this parameterization scheme, the turbulent kinetic energy (TKE) and diffusivity are determined by vertically nonlocal steady-state equations. The steady-state assumption is applicable in models with time steps longer than the characteristic turbulence evolution time scale."

At 10-km horizontal resolution and the corresponding time steps, eddies are not fully resolved in our model configuration, and parameterization of shear-induced mixing by eddies is necessary. MOM6 also includes other mixing options, such as the K-Profile Parameterization (KPP) (via the CVMix library) for upper-ocean and interior mixing, and specified background diffusivity. These alternatives were tested during early development of global MOM6 but produced less satisfactory results (Adcroft et al. 2019).

Line 216 unnecessary sentence.

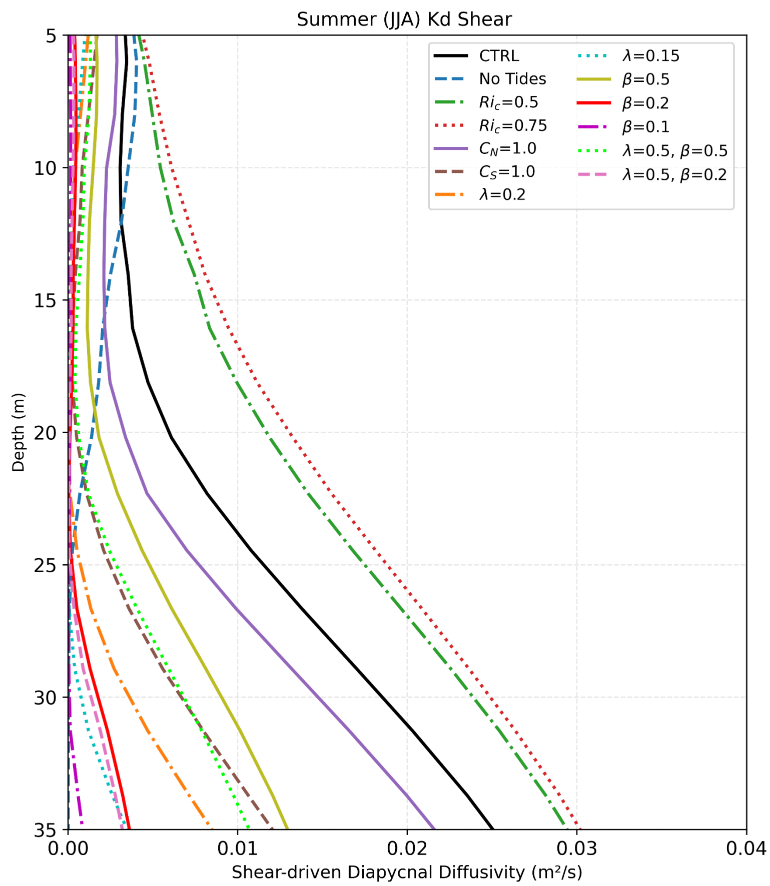
This was due to a paragraph formatting error. It's corrected now.

Line 227 Ric hasn't really been introduced. Presumably it's a parameter in the Jackson scheme. Why not decrease Ric?

On Line 243 of the revision, we described that "*Ri_c is the critical Richardson number above which turbulent mixing does not occur*". Jackson et al. (2008) provided a detailed discussion (see its Section 4b) on the range of Ri_c based on theoretical argument and direct numerical simulation. There is no strong justification for decreasing the value of Ri_c to being smaller than 0.25, so we didn't. But we agree that a lower Ri_c could potentially decrease mixing strength, since it would require a stronger shear to overcome stratification to produce mixing.

Line 235 A zoom in to the pycnocline depth might be informative. Often the slope of the mixing coefficient profile here plays a critical role, even though the absolute magnitude of the coefficient is smaller. It matters more because the T and Z gradients(curvature) are larger.

Below is a zoomed in version of the mixing coefficient profiles between 5 m and 35 m depth, which brackets the typical summer pycnocline depth of 20 m at the M4 location. This plot is included for your reference and helps to highlight the vertical structure and gradient differences among the sensitivity experiments in the region of strong stratification. Note this is shear-induced diffusivity only.



Line 265. I can only discern a few data points above 13 deg..is this correct?

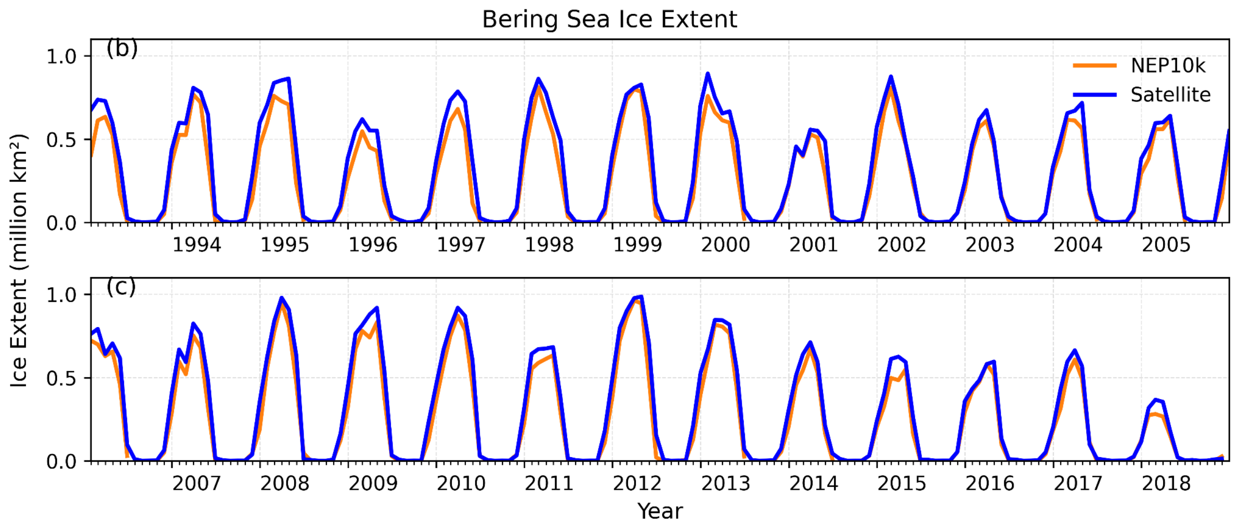
You are correct - summer bottom temperature on the Bering Sea shelf is rarely warmer than 13 °C, and it primarily occurred in the later years ~2016-2018. Note that the dots on Fig. 5 may be on top of each other.

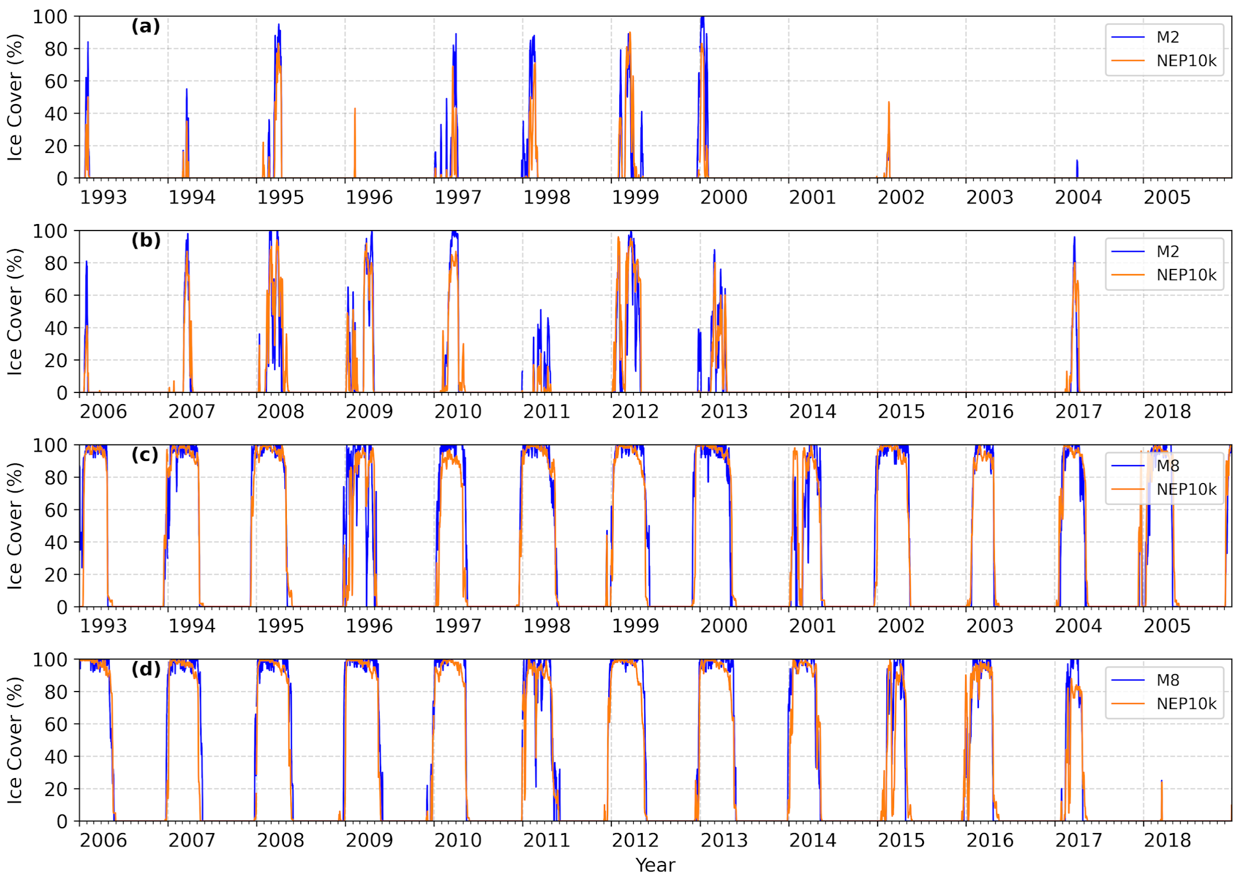
Line 315 clarify why those particular days in March were used for comparison. Were they the days of maximum sea ice extent?

"The March dates used for comparison were selected based on satellite observational data as reported in Stabeno and Bell (2019)". Specifically, March 20, 2012, was identified as a period of near-maximum sea ice extent over the eastern Bering Sea shelf, while March 17, 2018, represented a record low extent. These dates allow us to assess model performance under extreme ice conditions. We have clarified this in the revised manuscript (Lines 337-338).

Line 315. There are so many years squeeze onto this figure (and later ones) that it's hard to really distinguish the model and obs lines in panel b. Perhaps stretch over two rows split into first and second half of years modeled?

We split the years into two rows (1993–2005 and 2006–2018) to better show the model–observation comparison. These plots are included in the main manuscript (see Figure 6b,c, and Figure 7).





Line 352 the ‘discrepancy’ as determined by what measure?

We explained that *“The discrepancy refers to the difference (model minus observation) in daily sea ice concentration”,* as seen visually in the (Fig. 8a, c) at mooring locations. in Lines 380-381.

Line 356 M8 simulations?

We revised the text (line 385-386) to “*modeled ice concentration at the M8 location*”.

Line 360. Wasn't 'older' ice in Hunke's paper referring to multiyear ice in the arctic up to 6 years old? Is this applicable for seasonal? Also a high ice year, doesn't necessarily mean an early ice onset year, so it's uncertain that the ice would be older.

We agree that Hunke is not the best citation although thickness-dependent melt rates apply to both seasonal and multi-year ice. Since this feature was first discussed in Thorndike (1975), we now reference Thorndike (1975) instead. You are also correct that ice thickness/volume and ice age are not always linearly related. We modified this discussion to reflect the complexity of ice melt. See lines 388-391

Line 368 ...Cold Siberian Chukchi maybe related to lack of sea ice OBCs to flush sea ice?

Yes, the cold bias in the Siberian/Chukchi region may be influenced by the absence of sea ice open boundary conditions in our MOM6-NEP configuration. Without proper sea ice exchange across the northern boundary, the model may retain more ice in this region, leading to excessive surface cooling. We have added a sentence in the revised manuscript to reflect this limitation (see Lines 399-401).

The revised text now reads:

"The largest cold biases occurred in the Chukchi Sea and were primarily associated with summer months (Fig. 10c). This may be partly due to the absence of sea ice open boundary conditions which occur only in the Chukchi, which can lead to excessive sea ice retention and surface cooling by limiting ice exchange across the northern boundary."

Line 375. Discrepancies seem to suggest that the Yukon and Anadyr river outflows may be underestimated?

You're right, the discrepancies in freshwater distribution near the Yukon and Anadyr River suggest that the GloFAS based river discharge used in our model underestimates freshwater input from these sources. We have noted this possibility in the revised text and plan to investigate it further in future sensitivity experiments.

Revised text now reads (Lines 425-426):

"These discrepancies may reflect underestimated freshwater input from major rivers such as the Yukon and Anadyr in the GloFAS forcing, which could affect coastal salinity and stratification patterns."

Line 391. Looks like a 'warm' bias south of the Yukon outflow, maybe clarify where is being referred to.

There isn't a "warm bias" just south of the Yukon outflow. Fig. 9c shows overall cold annual mean SST biases over the shelf; Fig. 9f shows annual mean SSS (not SST) bias, which has a positive value south of the Yukon outflow. Fig. 10 a-d shows seasonal SST biases, which are positive along the entire coastal region in JJA. So, we think our original texts are accurate (see Lines 417-426).

Line 455. If Observations are more well mixed, does that mean that the adjustment of Beta parameter makes the solution worse here?

The "*well mixed*" was referring to that on Fig. 12 f, there are more dots near the top of the figure (where the modeled mld values are 70 m) than at the right edge (where the observed values are 70 m), However, since the dots on this figure can be on top of each other, it's hard to tell the exact difference. A slight timing mis-match between the model and observations can lead to such discrepancy. During winter, the entire water column on the middle shelf is well mixed; the thermocline analysis is primarily concerned with summer stratification. To eliminate confusion, we removed this statement.

Line 456 ..leftover comment from internal review probably

Removed.

Line 483. Examples of specific times or places the model didn't do well but the authors don't know why. I don't know that such discussions provide the reader with anything compelling.

We appreciate this critique. We added the following statements (Lines 514-520) in the revised text) to make the statements more compelling: "*Salinity is harder to measure and model than temperature. There is no natural feedback between SSS and surface freshwater fluxes, and biases of salinity in model simulations can accumulate although advection limits this bias. Despite this, the simulation captured the observed patterns of seasonal and interannual variability of water column salinity at M8, showing the penetration of fresh signal to deeper depth each autumn, and a relatively fresh period from late 2009 to 2015, and salty periods prior to 2009 and after 2016. 2014-2015 stands out as an anomalously fresh period in both OBS and model. These years had extensive ice cover at M8 in winter (Fig. 7c, d), though it is not significantly different from other years. Whether anomalies in ice thickness/volume contributed to the fresh signal in these years remains to be investigated".*

Line 501: A sentence fragment maybe meant for the conclusions? I'm not sure why it's here.

Removed.

Line 508 Temperature plays ... a tautology?

We removed redundancy. The revised text now reads (Lines: 543-544):
"Temperature plays a significant role in water column stratification, affecting the spatial distribution of pelagic and demersal communities."

Line 548- 552 This isn't really demonstrated anywhere in this paper.

We have revised the text to clarify that this is a qualitative assessment based on past experience. Revised text now reads (Lines:585-586):
"The model simulation appears to better capture weak shelf circulation features, such as flow along the 50-m and 100-m isobaths, compared to previous ROMS-based efforts (Kearney et al., 2020)."

Line 566 The mention of comparisons with Kearney appear a bit out of place in the conclusion. There weren't any quantitative comparisons with Kearney's results in the body of the manuscript to support these statements. Earlier biases in MOM6 freeze and melt were mentioned but I don't know how those compared quantitatively to the 2 weeks for Kearney referred to here.

The 2-week difference in ROMS Bering Sea model was systematic and happens every year but the MOM6 discrepancy discussed in Fig. 8 varies interannually, and there are no systematic biases in the long-term mean. We have revised the text to clarify this.

Revised text (Lines 615-618) now reads:

"...as well as the long-term mean sea-ice arrival and retreat timing in both the southern and northern shelf. This is also an improvement over prior ROMS Bering Sea simulations, where modeled sea-ice concentration shows a systematic bias in its long-term mean, tended to arrive and melt later than observations by ~2 weeks (Kearney et al., 2020)".

Line 576. There hasn't been any quantification of flow through Bering Strait vs. Local production, and presumably even if there were significant flow through the strait, the open boundaries may be far enough away not to matter.

Although not shown in this paper, the simulation captures the observed Bering Strait Sea water volume transport. We are not aware of any solid sea-ice transport observations through the Bering Strait. Regardless, we agree that the closed sea-ice boundary condition is far away from the Bering Sea and has not affected simulated sea ice in the latter. We have revised the sentence (Lines 626-627) to say that "*likely because most ice in the Bering Sea forms locally, and the missing sea-ice open boundary is far to north of the Chukchi Sea and does not affect our results."*

Line 583: There are several differences between this and Drenkard, based on my reading here. Is it clear that the differences are more likely caused by the difference in atm forcing than by the difference in mixing parameterization (beta)? - or did I misunderstand and Drenkard used the same adjustment to Jackson's scheme?

Drenkard et al. (2025) used the same adjustment to the Jackson et al. (2008) shear mixing scheme with the same β (lz_rescale = 0.2) as in this study. We have clarified this in the revised manuscript to avoid confusion (line 634).

Final comment: Careful review takes time. With the number of established authors on this manuscript I would expect that numerous issues I found in my reading should have been identified and addressed by the main author or their co-authors before submission.

An unexpected personnel situation happened when we were preparing for the submission, and we ran out of time. We apologize for the mishaps. We appreciate the time and effort you put into reading our manuscript and your insightful comments, which helped improve this manuscript. We have now added an acknowledgement to your review.

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.

Alaska Seafood Marketing Institute (2023). Economic Impacts Report https://www.alaskaseafood.org/wp-content/uploads/MRG_ASMI-Economic-Impacts-Report_2023_WEB-PAGES.pdf

Cheng, W., Hermann, A. J., Hollowed, A. B., Holsman, K. K., Kearney, K. A., Pilcher, D. J., Stock, C. A., and Aydin, K. Y.: Eastern Bering Sea shelf environmental and lower trophic level responses to climate forcing: Results of dynamical downscaling from CMIP6, Deep-Sea Res. Pt. II, 193, 104975, https://doi.org/10.1016/j.dsr2.2021.104975, 2021.

Drenkard, E. J., Stock, C. A., Ross, A. C., Teng, Y.-C., Morrison, T., Cheng, W., Adcroft, A., Curchitser, E., Dussin, R., Hallberg, R., Hauri, C., Hedstrom, K., Hermann, A., Jacox, M. G., Kearney, K. A., Pages, R., Pilcher, D. J., Pozo Buil, M., Seelanki, V., and Zadeh, N.: A regional physical-biogeochemical ocean model for marine resource applications in the Northeast Pacific (MOM6-COBALT-NEP10k v1.0), Geosci. Model Dev. Discuss, https://doi.org/10.5194/gmd-2024-195, 2025.

Hermann, A. J., Gibson, G. A., Bond, N. A., Curchitser, E. N., Hedstrom, K., Cheng, W., Wang, M., Stabeno, P. J., Eisner, L., and Cieciel, K. D.: A multivariate analysis of observed and modeled biophysical variability on the Bering Sea shelf: Multidecadal hindcasts (1970–2009) and forecasts (2010–2040), Deep-Sea Res. Pt. II, 94, 121–139, https://doi.org/10.1016/j.dsr2.2013.04.007, 2013.

Hermann, A. J., Gibson, G. A., Cheng, W., Ortiz, I., Aydin, K., Wang, M., Hollowed, A. B., and Holsman, K. K.: Projected biophysical conditions of the Bering Sea to 2100 under multiple emission scenarios, ICES J. Mar. Sci., https://doi.org/10.1093/icesjms/fsz043, 2019.

Hermann, A. J., Kearney, K., Cheng, W., Pilcher, D., Aydin, K., Holsman, K. K., and Hollowed, A. B.: Coupled modes of projected regional change in the Bering Sea from a dynamically downscaling model under CMIP6 forcing, Deep-Sea Res. Pt. II, 194, 104974, https://doi.org/10.1016/j.dsr2.2021.104974, 2021.

Jackson, L., Hallberg, R., Legg, S., Jackson, L., Hallberg, R., and Legg, S.: A parameterization of shear-driven turbulence for ocean climate models, J. Phys. Oceanogr., 38, 1033–1053, https://doi.org/10.1175/2007JPO3779.1, 2008.

Kearney, K., Hermann, A., Cheng, W., Ortiz, I., and Aydin, K.: A coupled pelagic–benthic–sympagic biogeochemical model for the Bering Sea: documentation and validation of the BESTNPZ model (v2019.08.23) within a high-resolution regional ocean model, Geosci. Model Dev., 13, 597–650, https://doi.org/10.5194/gmd-13-597-2020, 2020.

National Marine Fisheries Service (2024). Fisheries of the United States, 2022. U.S. Department of Commerce, NOAA Current Fishery Statistics No. 2022. Available at: https://www.fisheries.noaa.gov/national/sustainable-fisheries/fisheries-united-states

Pilcher, D. J., Cross, J. N., Hermann, A. J., Kearney, K. A., Cheng, W., and Mathis, J. T.: Dynamically downscaled projections of ocean acidification for the Bering Sea, Deep Sea Res. Part 2 Top. Stud. Oceanogr., 198, 105055, https://doi.org/10.1016/j.dsr2.2022.105055, 2022.

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.

Thorndike, A. S., Rothrock, D. A., Maykut, G. A., and Colony, R.: The thickness distribution of sea ice, J. Geophys. Res., 80, 4501-4513, 10.1029/JC080i033p04501, 1975.

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.

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.

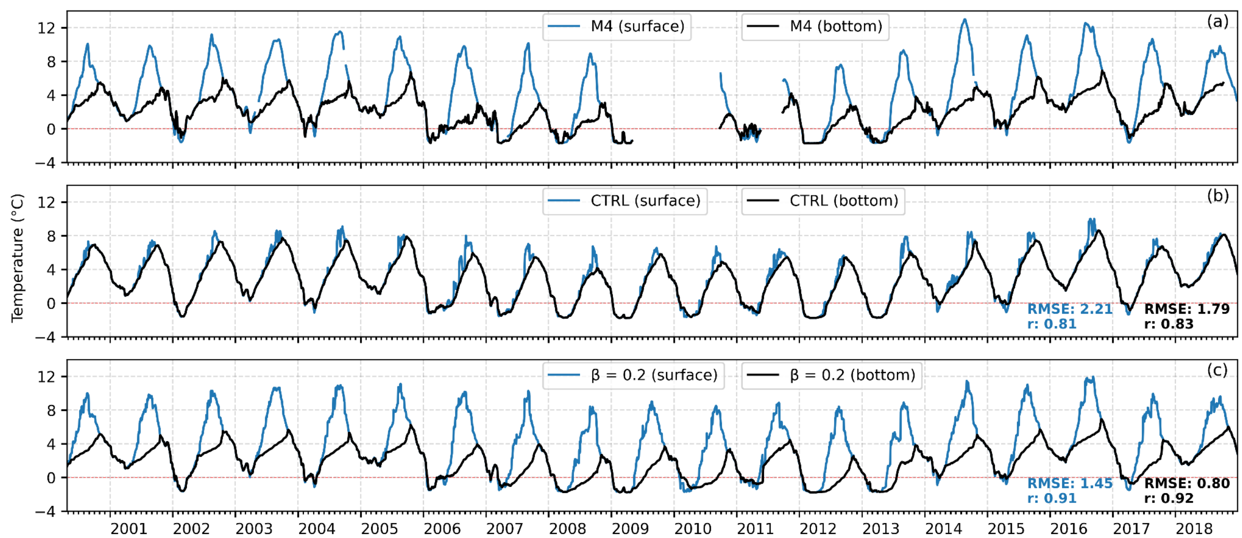


Figure S2: Comparison of surface and bottom temperatures at (a) M4 mooring location, against (b) CTRL (default MOM6-NEP) and (c) β =0.2 simulations. This comparison shows the excessive shear-driven vertical mixing in the CTRL simulation.

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."

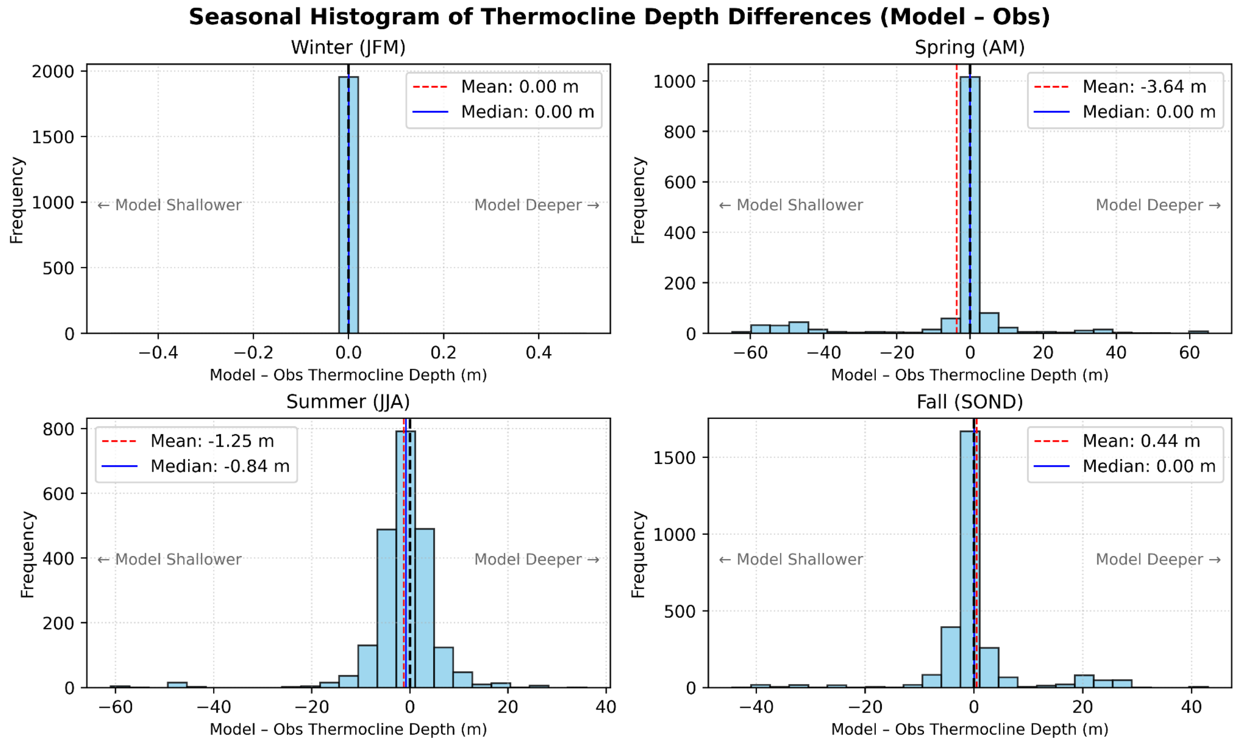


Figure S3: Seasonal histogram of thermocline depth differences (model minus observation) at M2 mooring location.

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.