

Response to comments from Reviewer #2

Reviewer's comments are shown in black, and the authors' responses are provided in blue below each comment.

We thank the reviewer for the careful, thorough and constructive assessment of our manuscript. We agree with many of the reviewer's points and have revised the manuscript accordingly to improve clarity, methodological transparency and scientific rigor.

We wish to emphasize that the objective of this study is strictly to document and evaluate the performance of the existing TOPAZ5 high-resolution HYCOM-CICE configuration in free-running mode over 2009-2019, rather than to present a new model sensitivity study or a new observational synthesis of the Arctic Ocean. TOPAZ5 is evaluated using the observational and reanalysis products and hydrographic sections available within the established validation framework. The hindcast contains no data assimilation and the reported results therefore describe the intrinsic behaviour and limitations of the coupled HYCOM-CICE configuration. We agree, however, that the manuscript should more clearly distinguish demonstrated model skill from limitations imposed by observational coverage and by the dependence between some forcing and validation datasets.

Below we respond point by point.

Major Comment 1: *“The model domain is focused on the Arctic and North Atlantic, however there is no validation whatsoever for the Arctic Ocean hydrography and the T/S sections used for validation are all in the Nordic Seas outside of the Arctic Ocean. While in-situ data is indeed sparse in the Arctic Ocean, other studies have provided extensive model validation for the different basins in this region using for example PHC3 or WOA18 climatology (Khosravi et al., 2022; Shu et al., 2023; Wang et al., 2024). A similar analysis is the minimum requirement for presenting a new model for this specific region. There are also existing CTD, ITP and mooring observations for the Arctic Ocean, which can be used for validation.”*

Response: We agree with the reviewer that subsurface observations within the Arctic Ocean would provide an important additional constraint on the model evaluation. However, the present hydrographic-section analysis was specifically designed to evaluate the representation of the main Atlantic Water inflow into the Arctic system. The selected Svinøy, Gimsøy, and Barents Sea Opening (BSO) sections are the major gateways through which AW enters the Nordic Seas and subsequently the Arctic Ocean. This is explicitly stated in the manuscript.

The present study therefore does not claim to provide an exhaustive observational validation of subsurface hydrography throughout the entire Arctic Basin. We agree that some statements in the manuscript were too broad in this respect and have qualified them accordingly. Specifically, statements referring to the model as capturing the "pan-Arctic hydrographic structure" have been revised so that they refer specifically to the basin-scale surface fields and the validated AW inflow pathways.

We respectfully maintain that the absence of an additional Arctic-wide CTD/ITP/mooring analysis does not invalidate the present evaluation. Such an analysis would constitute a separate study aimed at evaluating the full 3D structure of the Arctic basins. The present manuscript provides a comprehensive multi-parameter evaluation of:

- Arctic surface hydrography (SST, SSS) against OSTIA and WOA,
- Arctic sea-ice thickness and concentration against CS2SMOS and OSI-SAF,
- Transports through the Bering Strait, Fram Strait and Barents Sea Opening,
- Vertical structure along the critical AW inflow branches.

In addition, we have added T and S maps at 300 m against the WOA climatology to provide an additional subsurface constraint on the Arctic Ocean hydrography.

We have added a statement in the Discussion explicitly acknowledging that a comprehensive Arctic Ocean subsurface validation, while highly desirable, is a distinct and extensive exercise and remains an important avenue for future work.

Major Comment 2: *“The model validation in this paper is done in several cases against observational data that is also assimilated into the reanalysis products used at the lateral boundaries (GLORYS12V1) or the atmospheric forcing (ERA5). While some circularity in this kind of setup might not be completely avoidable, it should be minimized where possible, and any remaining circularity must be explicitly stated and discussed. Specific instances: (a) SST vs. OSTIA — ERA5's prescribed SST boundary condition is the operational OSTIA product from September 2007 onward (i.e. for the entirety of the 2009-2019 validation period, Hersbach et al., 2018), so the model is in part being validated against the same SST fields used to force it. (b) SLA and EKE vs. GLORYS — GLORYS12V1 assimilates along-track sea-level anomalies from satellite altimetry, making it a dynamically interpolated product of the same observations used for the “independent” altimetry comparison. Since GLORYS also supplies TOPAZ5's lateral boundary conditions, validating TOPAZ5's SLA/EKE performance against GLORYS mixes a real model comparison with a comparison against the boundary condition's own data source. Section 2 states that “surface relaxation of sea surface salinity (SSS) towards the corresponding monthly climatology was applied”. It is not clear from this wording whether the SSS surface relaxation is applied only during 1992 or throughout the entire 1992-2019 simulation (I assume it is the latter). This needs to be stated explicitly, as it directly affects the validity of the SSS results, since WOA 2018 is also used as the initial condition and as the sole observational reference for SSS validation. If the surface SSS relaxation is applied for the full simulation period, the SSS validation would be fully circular, since the model is then being validated against the same field it is restored toward.”*

Response: We thank the reviewer for raising this important methodological point. The reviewer is correct that the reference datasets used in this study are not strictly independent of the model forcing or lateral boundary conditions. We have revised the manuscript to explicitly

acknowledge these dependencies and to clarify the interpretation of the corresponding validation metrics.

Specifically:

First, SST vs. OSTIA: The atmospheric forcing is ERA5, and OSTIA contributes to the ERA5 SST product. Consequently, the SST comparison with OSTIA should not be interpreted as a completely independent validation of the model surface temperature. We have explicitly acknowledged this dependence and clarified that the OSTIA comparison primarily evaluates how the TOPAZ5 solution behaves relative to a widely used high-resolution SST analysis. Importantly, however, TOPAZ5 does not directly prescribe OSTIA SST, the SST evolves dynamically within the coupled ocean–sea-ice model and is affected by surface heat fluxes, ocean circulation, mixing, sea-ice processes and horizontal/vertical transport. Thus, while the comparison is not fully independent, it remains a meaningful consistency assessment.

Second, regarding SLA and EKE versus GLORYS, we acknowledge that GLORYS12V1 assimilates along-track satellite altimetry. We therefore agree that the comparison between TOPAZ5 and GLORYS for SLA and SSH variability is not independent of the direct comparison with satellite altimetry in the vicinity of the model boundaries. At about 1000 km from the boundary, the TOPAZ5 mesoscale activity evolves differently from the GLORYS data, and the dependence vanishes. We note that the manuscript did not previously use the word "independent" for this comparison; the concern is nevertheless valid, and removing GLORYS from the validation addresses it directly.

Third, SSS relaxation toward WOA18, we agree that the temporal applicability of this relaxation must be stated unambiguously. The relaxation toward WOA18 monthly climatology was applied only during the 1992 initialization year, and not during the 2009–2019 validation period. The model uses surface SSS relaxation: The relaxation timescale is 30 days. The relaxation is only in the open ocean, not below the sea ice, and a maximum salinity difference of 0.5 is used. In addition, lateral-boundary relaxation is applied only during the initialization year. We therefore agree that SSS in the open ocean is not completely independent of the climatology used for relaxation. We have revised the manuscript to avoid implying that the comparison is unconstrained. The revised text now states that the open-ocean comparison is partly constrained and should be regarded as a consistency check, while under sea ice the model SSS is not directly relaxed and the comparison is therefore less directly constrained by the WOA18 climatology.

Major Comment 3a: *“The bathymetry used in this setup is GEBCO_2014, but the authors state in the discussion that “Refining these transports will require higher-resolution grids and improved bathymetric representations.”. Why not make use of the up-to-date bathymetry data GEBCO_2026, which is substantially improved compared to the 12-year-old version, especially in the Arctic region.”*

Response: We acknowledge that newer bathymetric products, including GEBCO_2026 and updated IBCAO versions, provide improved representations of Arctic bathymetry. However,

the present paper evaluates a specific TOPAZ5 model configuration and its existing hindcast for 1992-2019. The bathymetry is a fixed component of the configuration whose performance is being documented. Replacing GEBCO_2014 with a subsequently released bathymetric product would not constitute a simple correction to the present experiment; it would define a different TOPAZ5 configuration and would require preprocessing all model input data according to the new bathymetry (by far the most time-consuming task), rerunning the entire 28-year hindcast to isolate the impact of bathymetric changes on circulation, transports, sea level and sea-ice properties. Such a sensitivity experiment is beyond the scope of the present evaluation study, which is focused on documenting the performance of the existing configuration. While an update to GEBCO_2026 would be beneficial, the associated costs of updating the bathymetry and rerunning the hindcast are substantial, and this update is therefore planned as part of a subsequent model upgrade, coordinated with a broader project team. We agree, however, that the wording in the discussion could be interpreted as suggesting that improved bathymetry is unavailable. We have revised the text to state this.

Major Comment 3b: *“Similarly, the authors acknowledge the limited data in the MLD climatology by de Boyer Montégut et al. (2004), but they do not use the newer Holte et al. (2017) climatology that also includes Argo data.”*

Response: We agree that the Argo-based climatology of Holte et al. (2017) provides valuable additional observational information, with substantially improved coverage in the subpolar regions. We also acknowledge that the limited observational coverage of the de Boyer Montégut et al. (2004) climatology represents an important limitation. Following the reviewer's suggestion, we have replaced the de Boyer Montégut et al. (2004) climatology with the Holte et al. (2017) climatology as the observational reference in the revised manuscript. All comparisons of mixed layer depth are now based on Holte et al. (2017), which provides a more robust and better-constrained assessment of model performance, especially at high latitudes. The corresponding text, figures, and discussion have been updated accordingly.

Major Comment 4: *“There are several instances when the paper seems to contradict its own findings. For example: the abstract and conclusions describe “warm and saline” surface biases, but the biases shown in figures 1/2 and summarized in table 1 are cold and fresh. Also, the differences in the seasonal cycle between TOPAZ5 and WOA 2018 SSS are described as “smallest during winter and increase toward late summer and early autumn”, but figure 4 shows the smallest differences in June.”*

Response: We agree that the wording in these locations was inaccurate and, in some cases, inconsistent with the spatially averaged statistics. The reviewer is correct that the sign and magnitude of the bias depend strongly on region and season.

For SST: The spatial maps show both warm and cold regional biases, while the basin-averaged statistics relative to OSTIA give a small negative mean bias (-0.11 °C). Therefore, we have replaced the "warm" by a "cold" SST bias.

Similarly, the SSS field shows positive biases over parts of the Beaufort and Laptev shelves (exceeding 2 in summer), whereas the ice-free domain-averaged comparison with WOA gives a negative mean bias (-0.21). This distinction is already discussed in the results, where the apparent difference between the spatial and basin-averaged behaviour is attributed to the different spatial masks. We have made this distinction clearer in the abstract and conclusions.

For the SSS seasonal cycle: We have revised the relevant sentence to describe the observed/modelled seasonal relationship precisely as shown in Fig. 4. The differences are now described accurately, including the fact that the smallest differences occur in June.

Response to Detailed Comments

Figure 1: *“It would be helpful to see a map of horizontal resolution. This could possibly be added as figure 1b.”*

Response: We appreciate the suggestion and we have added a panel to Figure 1 showing the horizontal resolution of the model domain.

Line 86: *“The target densities for the lower 40 hybrid z-layers were adjusted to reflect the water masses in the focus region.” Please specify which water masses were targeted, and provide the actual target density values. It would also be useful to state whether this included Arctic water masses (e.g. Pacific Water, the Atlantic Water core, cold halocline density range) or was focused on the Nordic Seas / North Atlantic. How well are these water masses simulated in TOPAZ5?”*

Response:

We thank the reviewer for this comment. The range of target densities was set to include all the major water masses in the region. The target densities were initiated from the linear-exponential density scale recommended in HYCOM and then adjusted after successive test runs to correct instances of too thick or too thin layers. Sensitivity experiments were conducted to determine the optimal target densities for the 50 hybrid layers (range: 0.1 to 28.12 σ units), refining the configuration to better represent the water-mass structure across the domain, including the Atlantic Water layer. The 10 upper layers are always z-levels with the top layer at 1.2 m, ensuring sufficient vertical resolution in the upper ocean for simulations including biogeochemistry. Alternative configurations, including setups with different target density profiles, were tested, but the chosen configuration provided the best balance between near-surface stratification fidelity and isopycnal water-mass tracking. To illustrate the density coverage, the target densities for layers 12-21 are 25.72, 26.40, 26.90, 27.13, 27.19, 27.25, 27.29, 27.34, 27.42, and 27.50 σ units, respectively. This range encompasses the density domain typically associated with Pacific Water (approximately 25.5-27.5 σ units), while Atlantic Water densities (~27.7-27.9 σ units) are represented in the overlying layers. The configuration therefore spans the full density range relevant to the North Atlantic, Nordic Seas, and Arctic system without being explicitly tuned to any single water mass. The original

sentence in the manuscript has been revised accordingly to avoid the misleading impression that specific water masses were targeted.

We added the paragraph below to the manuscript:

“To ensure a dynamically and hydrographically consistent vertical coordinate across the extended North Atlantic–Nordic Seas–Arctic domain, the target isopycnal densities for the hybrid mixing scheme were prescribed based on the climatological water-mass distribution of the entire system rather than being tuned to any single water mass. The model uses 50 hybrid layers with carefully chosen reference potential densities (Appendix Table A1). The 10 uppermost layers are z-levels, with the top layer at 1.2 m, ensuring high vertical resolution near the surface where air-sea exchange processes are most critical. Below these, the remaining 40 isopycnal layers span the density range 24.05–28.12 σ units, which is relevant to the North Atlantic, Nordic Seas, and Arctic. In particular, layers with target densities of 25.72–27.50 σ units cover the density domain typically associated with Pacific Water, whereas Atlantic Water densities are represented by the 9 layers with target densities of 27.72–27.90 σ units. Sensitivity experiments were performed to determine an appropriate distribution of target densities. The chosen configuration provided the best balance between near-surface stratification fidelity and isopycnal water-mass tracking.”

Line 87: *“I mentioned the bathymetry dataset in my major comments. GEBCO_2026 is available, including the updated International Bathymetric Chart of the Arctic Ocean (IBCAO), improving the important Arctic Ocean bathymetry.”*

Response: Please see our response to Major Comment 3a. A change to GEBCO_2026 would be beneficial, but the associated costs are very high, and this is therefore planned for a subsequent model upgrade.

Line 100: *“WOA 2018 is used for initialization in 1992. Is there a specific reason for not using PHC3 or another Arctic Ocean-focused climatology? Please also specify which WOA 2018 product was used for initialization (e.g. a specific decadal average, or the multi-decadal mean). If it is the same 2005–2017 climatology used later for validation, this would introduce a possible warm-biased state compared to 1992.”*

Response: The simulation was initialized in September 1992 using WOA18 temperature and salinity fields. This product represents a multi-decadal climatological mean covering 1955–2017 period; it is not the 2005–2017 climatology used subsequently for validation. Consequently, the initialization does not preferentially impose the relatively warm conditions of the 2005–2017 period and does not introduce the specific warm-state bias suggested by the reviewer on that basis.

The manuscript has been revised to specify the exact WOA18 product and its climatological period and to distinguish the initialization from the subsequent validation period. The purpose of the long spin-up is particularly important here: the 2009–2019 evaluation is not intended to

represent the direct persistence of the initial 1992 climatological state. The model is allowed to dynamically adjust for 17 years before the period used for validation. We therefore do not consider the use of WOA18 for initialization to constitute a direct validation of the initial state, except in the very deep ocean.

Regarding the choice of climatology, we acknowledge that PHC3 or another Arctic-focused product could provide an alternative representation of Arctic hydrography. However, a key consideration is temporal coverage. The PHC product primarily represents conditions prior to 2000, meaning its observational basis is largely contemporaneous with, or even older than, the 1992 initialization date. While this could be argued as an advantage for historical fidelity, it also means the product does not incorporate the more recent observational data (including the 2000s and 2010s) that are included in WOA18. We selected WOA18 because it is a well-established, globally consistent product that benefits from a longer and more recent observational record, which we judged to be more appropriate for constraining the model's initial state in a consistent manner with the rest of our experimental design.

We have added a clarification of the specific WOA18 product used and the rationale for its selection to the revised manuscript.

Line 120: *"Please see my first major comment. Excluding the ice covered Arctic Ocean from validating an Arctic Ocean model is not acceptable."*

Response: We thank the reviewer for this critical and constructive observation. Regarding the surface fields: we retain the sea-ice mask for surface temperature and salinity comparisons only, because (i) OSTIA SST values beneath sea ice are not observations but a gap-filled interpolation, and (ii) available Arctic SSS climatologies are not reliable in the ice-covered near-surface layer. We agree this means surface validation in the ice-covered Arctic is not meaningful, and we do not claim it as such. However, we fully agree with the reviewer that this does not excuse the absence of subsurface validation. We have therefore added comparisons of model temperature and salinity against WOA18 climatology at 300 m.

Line 141: *"Standard deviation of SSH is used as a proxy for EKE, which is a valid approach. However, in several instances the metric is simply referred to as "EKE," which is not what is actually being computed. Please either compute true geostrophic EKE from SSH gradients, or full EKE from model velocities, or consistently label the metric as SSH variability/standard deviation throughout the text, figures, and abstract."*

Response: We agree with this terminology point. To avoid ambiguity, we now consistently refer to this quantity as "SSH standard deviation" throughout the manuscript, including the abstract, figures, captions and results.

Figure 2: *"There are biases in the under ice SST when comparing TOPAZ5 to OSTIA, indicating that the calculation of T_{freeze} in the model does not match T_{freeze} in the observations."*

Response: We thank the reviewer for the insightful comment. Yes, under the ice cover, the observed SST may be uncertain due to freezing temperatures. In TOPAZ5, the freezing temperature is determined by a linear function of salinity: $T_f = -0.054 \cdot S$. Multiplying -0.054 by a salinity of 35 gives -1.89°C, which closely aligns with the standard -1.8°C to -1.9 °C range. The OSTIA product also provides sea ice concentration, which uses a conventional freezing temperature of about -1.8 °C, intended for other uses than model validation. Although these two freezing temperatures are very close but not identical, Figure 2 still provides a visual overview of the deviation pattern.

To quantify the SST deviation excluding sea ice cover, we further calculate their RMSE and Bias in the text at Line 176: “Basin-averaged statistics, excluding the ice-covered Arctic, for the period 2009-2019 indicate that SST in TOPAZ5 is closer to OSTIA dataset (RMSE = 0.17 °C, Bias = -0.11 °C) than to the WOA climatology (RMSE = 0.43 °C, Bias = -0.35 °C) (Table 1).”

Figure 4: *“The differences that are referred to in the text are very small compared to the full seasonal range and therefore hard to see in the plots. As mentioned in major comment 4, the description in the text (and the conclusions drawn from that) don’t match the results shown.”*

Response: Agreed.

Revisions made: We have revised the text so that the description follows the plotted seasonal differences directly.

Figures 5 and 6: *“(and corresponding appendix figures): The x-axis is unlabeled - please specify what it represents (e.g. distance along section, longitude, or station number). Additionally, the axis range and tick values differ between the observation and model panels, making it unclear whether the two panels share a common horizontal reference frame; this should be corrected so that features can be directly compared across panels. It would also be helpful to add a third panel showing the model-observation difference, as in the earlier bias figures (e.g. Fig. 2).”*

Response: We agree that the section coordinate should be clearly identified so that the observational and model panels can be compared unambiguously.

We have:

- Clarified the horizontal coordinate in the figure captions and ensured that the existing panels are interpreted using the same section reference.
- Added model-minus-observation difference panels for the hydrographic sections, following the reviewer's suggestion. These panels now appear alongside the paired observational/model cross-sections.

Figure 7: *“As stated in major comment 2, GLORYS assimilates along-track satellite altimetry, so the “independent” altimetry and GLORYS comparisons are not independent of each other.”*

Response: We agree that including GLORYS alongside the independent altimetry product in Fig. 7 could give the misleading impression of an independent multi-platform validation, since GLORYS assimilates along-track satellite altimetry. To avoid any ambiguity, we have therefore removed GLORYS entirely from Fig. 7.

Line 311: *““TOPAZ5 also underestimates positive trends in the Barents and Greenland Seas relative to both observations and GLORYS.” - This understates what Fig. 7a actually shows: in part of the Greenland Sea TOPAZ5 produces a negative SLA trend, whereas both altimetry (Fig. 7b) and GLORYS (Fig. 7c) show consistently positive trends there. This is a discrepancy in trend sign, not merely magnitude and should be described and discussed as such.”*

Response: We agree that the wording "underestimates positive trends" is insufficiently precise for the region identified by the reviewer. Where TOPAZ5 produces a negative trend while the reference datasets show positive trends, this represents a difference in the sign of the trend rather than merely an underestimate of its magnitude. We have revised the text to describe this explicitly: "TOPAZ5 fails to reproduce the observed positive SLA trend and produces locally negative trends in part of the Greenland Sea."

Line 322: *“differences in peak amplitude are evident, with altimetry recording higher values, suggesting limitations in the resolution of TOPAZ5” - Please clarify the mechanism being proposed here and why it is influenced by model resolution. Fig. 8 shows regionally averaged, monthly-mean SLA time series. This should largely remove any resolution-dependent (i.e. mesoscale/submesoscale) signals. It is not clear how horizontal grid resolution would produce a difference in regionally averaged, monthly-mean peak amplitude. Please justify this explanation or offer an alternative.”*

Response: We agree with the reviewer that the original wording attributed the difference too directly to horizontal resolution. Figure 8 represents a spatially averaged, monthly mean SLA signal, and therefore the difference cannot be interpreted simply as unresolved mesoscale variability. We have removed the specific attribution of the Fig. 8 peak-amplitude difference to horizontal resolution and instead state that the discrepancy may reflect differences in the simulated large-scale circulation, water-mass redistribution, imperfections of the boundary forcings and the absence of data assimilation.

Figures 7/9: *“Why is the Arctic Ocean basin masked out entirely in these figures (likely due to unavailable/unreliable altimetry under sea ice), while the same region is shown with data in Figs. 2 and 3, despite being excluded from the corresponding statistics there? Please state the reason for masking in the captions, and consider using a consistent treatment (e.g. showing the ice mask boundary) across all spatial figures in the paper.”*

Response: We thank the reviewer for this critical and constructive observation. We agree that the inconsistent treatment of the Arctic Ocean basin across the figures was confusing and required clarification. To directly address this concern, we have completely removed the ice mask from our validation procedure and re-evaluated all model outputs accordingly. This ensures a consistent spatial treatment across all figures in the manuscript. We have also revised the relevant text, figures, and discussion throughout the manuscript to reflect this change.

Removal of either altimetry or GLORYS: *“Since altimetry and reanalysis are not fully independent data sets, the authors could consider removing one of the two and instead showing the biases, like in figures 2 and 3.”*

Response: We thank the reviewer for this suggestion. While altimetry and GLORYS are not fully independent, they provide complementary information: altimetry offers direct observational constraints, while GLORYS assimilates a broader range of observations. Removing one would weaken the comparison. Crucially, by extending our analysis below the ice mask, GLORYS provides a reference against which the altimetry standard deviations can be evaluated. This broader context shows that the larger deviations in certain regions reflect the inherent challenges of altimetric measurements rather than a systematic issue.

Line 331: *“TOPAZ5 successfully captures the large-scale spatial distribution of mesoscale variability across the North Atlantic and Arctic Ocean. This is not well supported by Fig. 9. Beyond the systematic amplitude underestimation described in the following sentences, there are real differences in spatial pattern.”*

Response: We agree that the original statement was too strong; we replaced it with wording indicating that TOPAZ5 reproduces some of the major regions of enhanced SSH variability but does not fully reproduce their spatial structure or amplitude. This change reflects better Fig. 9 and is consistent with the limitations already discussed in the manuscript.

Line 333: *“western North Atlantic” should be rephrased here. The “western North Atlantic” conventionally refers to the US/Canadian seaboard region, including the Gulf Stream separation zone, none of which is part of the model domain.”*

Response: Agreed.

Revisions made: Using the term “western North Atlantic” was imprecise and potentially misleading. To address this, we have removed the incorrect designation and rephrased the text to refer explicitly to the subpolar North Atlantic, Nordic Seas, and Arctic regions, which accurately capture the high-latitude areas analyzed in our study.

Line 348: *“The MLD climatology by de Boyer Montégut et al. (2004) is used for validation, but this data set predates the Argo era and has comparatively sparse coverage in the subpolar*

North Atlantic, Nordic Seas, and Arctic. The authors reference Holte et al. (2017), but it does not seem like the 2017 climatology is used. Additionally, while the constant density threshold of 0.03 kg/m³ is widely used, I recommend using the more recent, Argo-based Holte & Talley (2009) hybrid algorithm and/or the Holte et al. (2017) climatology as the observational reference, given both the more recent data coverage and its more robust handling of convective regions."

Response: Please see our response to Major Comment 3b. We have added the Holte et al. (2017) climatology as an observational reference.

Figure 10: *"Same masking issue as for figures 7/9. Additionally, the colour scale in the winter panels (0-800 m) appears to be saturated in many regions and it is not possible to tell whether these values are close to 800 m or well in excess of it. What is the range of the data? Please choose a color scale that does not hide the actual range of the data."*

and **Line 359:** *">700m, but the color scale in figure 10 shows that TOPAZ5 MLD at least exceeds 800m if not more. Again, what is the actual range of the data?"*

Response: We thank the reviewer for this constructive feedback. We fully agree with both issues raised regarding Fig. 10 and Line 359.

1. Masking: We have removed the mask from Fig. 10 to ensure consistency with the revised Fig. 7 and 9, eliminating the visual discrepancy noted by the reviewer.

2. Color scale saturation: The reviewer is correct that the original 0-800 m color scale does not show the maximum mixed layer depth which in these regions reaches approximately 1100 m. We have revised the color bar to span 0 to 1200 m. This new scale ensures that no saturation occurs and allows the full spatial variability to be clearly visible across all panels.

3. Textual consistency: In line with this revision, we have updated the corresponding text to report the actual maximum MLD values quantitatively, replacing the ">700 m" threshold. We have also cross-checked the figure caption and surrounding discussion to ensure all reported values are fully consistent with the revised color scale.

Line 364: *"The authors acknowledge that the climatology's "limited availability of hydrological profiles... particularly in the subpolar North Atlantic and Arctic-adjacent seas" affects its representativeness in exactly the regions this study focuses on. Given the authors have identified this limitation themselves, why not use a more recent, better-sampled climatology that addresses it (e.g. Holte et al. (2017) which is already cited elsewhere in the manuscript, see also the comment on line 348)? Separately, note that this section states TOPAZ5 MLD exceeds 1000 m in the Labrador Sea, which confirms that the Fig. 10a colour scale (capped at 800 m) and the ">700 m" figure text (line 359) both understate the actual model values."*

Response: Please see our response to Major Comment 3b. We have added the Holte et al. (2017) climatology as a reference.

Line 368: *"spatial patterns align: this is not supported by Fig. 10. TOPAZ5 has the deepest winter convection in the Irminger Sea and a large patch of deep MLD in the Labrador Sea, whereas the climatology shows the deepest winter MLD in the Nordic Seas. Similarly, the summer patterns (including the locations of local minima and maxima) are not well reproduced."*

Response: We agree that "spatial patterns align" is too strong wording when applied to the detailed regional structure. We have replaced "spatial patterns align" with wording indicating that TOPAZ5 reproduces the occurrence of deep winter mixing in several major convection regions but differs from the climatology in the detailed spatial distribution and magnitude.

Tables 2 and 3: *"These could be very easily combined. The last column in 2 is identical to the second column in 3. The section position is not very helpful here, and if included should also carry units (like °E)."*

Response: We appreciate the suggestion. We have combined the tables for clarity and ensured that section coordinates have appropriate units.

Line 425: *"Despite general agreement at the basin scale" is not well supported by Fig. 11 or Fig. 12. The thickness bias covers nearly the entire Arctic basin in positive and negative regions with substantial magnitudes. The September concentration bias likewise extends over a large fraction of the central basin. Additionally, while the text does mention both observational uncertainty and model bias as possible causes of the thickness overestimation, this mismatch is primarily attributed to observational uncertainty.*

Response: We agree that the wording "despite general agreement at the basin scale" could understate the magnitude and spatial extent of the regional biases shown in Figs. 11 and 12. We have therefore revised the wording to clarify that, while the model reproduces the broad spatial distribution of thick and thin ice, substantial regional biases in both ice thickness and September ice concentration remain. We have also revised the discussion of the thickness overestimation to avoid attributing the mismatch primarily to observational uncertainty. The revised text explicitly recognizes that both observational uncertainty and model-related processes, including model physics and ice dynamics, may contribute to the simulated thickness biases. This interpretation is also consistent with our discussion of EVP-related ice convergence and its potential influence on the regional ice-thickness distribution

Line 436: *"Comparisons between seasonal sea ice concentration (Fig. 12) further support the model's ability to reproduce observed patterns, although with varying levels of accuracy.": A*

large region corresponding to the Beaufort, Chukchi, and East Siberian Seas shows bias of around 1, i.e. close to the physically maximum possible bias. Given the magnitude, describing this as "varying levels of accuracy" considerably understates the problem.

Response: We agree that "varying levels of accuracy" understates the magnitude of the errors in parts of the Beaufort, Chukchi, and East Siberian Seas. We have revised the text to state directly that "The seasonal sea-ice concentration comparison (Fig. 12) reveals substantial regional biases, particularly in the Beaufort, Chukchi, and East Siberian Seas, where the simulated summer sea-ice concentration differs markedly from observations. In contrast, the central Arctic is comparatively better represented, indicating pronounced spatial heterogeneity in the model's sea-ice performance."

Line 461: *"TOPAZ5 effectively captures the pan-Arctic and North Atlantic hydrographic structure,": this is a strong statement, given that no subsurface hydrographic validation was presented for the Arctic Basin itself (see major comment 1). Please clarify what this statement is based on, or qualify it accordingly.*

Response: We agree that this statement should be qualified. We have revised the statement to distinguish between the demonstrated basin-scale surface hydrographic structure and the more limited subsurface validation along the Atlantic Water inflow sections.

Line 467: *"Refining these transports will require higher-resolution grids and improved bathymetric representations.": better bathymetry, especially in the study region of the paper, is available. See main comment.*

Response: Please see our response to Major Comment 3a. The statement has been revised so that it does not imply that improved bathymetry is unavailable. Rather, improved bathymetry is identified as a potential future development of the TOPAZ5 configuration.

Line 470: *">700m", see comment above*

Response: Agreed. Answered above in **Line 364** answer.

Revisions made: The wording has been made consistent with the actual range of modelled MLD values. The manuscript now uses the specific values (exceeding 1000 m in the Labrador Sea and approximately 800 m in the Irminger Sea) rather than the generic ">700 m" description where appropriate.

Line 472: *"I don't understand the reference to de Boyer Montégut et al. (2004). That paper describes the observational climatology, but here it is cited in the context of overly diffuse mixing schemes in ocean models.*

The second part of the paragraph makes claims about the stratification (“weak thermocline/halocline gradients reduce vertical stratification, particularly during melt seasons”), but this has not been shown anywhere for the Labrador, Irminger, or Greenland Seas (the regions under discussion in this section). Thermocline/halocline smoothing has only been documented at the hydrographic sections located on the Norwegian coast.”

Response: We agree that the citation was inappropriate in the context identified by the reviewer. We have revised this passage and removed the inappropriate attribution. More generally, the discussion of vertical mixing is now restricted to what is supported by the hydrographic sections and MLD comparison presented in the manuscript. In particular, we avoid extending the specific thermocline/halocline smoothing diagnosed along the Norwegian sections to regions where it was not directly demonstrated.

Line 514: *“TOPAZ5 realistically captures large-scale ocean circulation, sea ice variability, and key seasonal hydrographic patterns, reproducing observed trends including surface warming, upper-ocean freshening, and changes in volume transport and circulation.”: I disagree with this characterization. Large-scale ocean circulation was not analyzed, sea ice variability is not captured realistically, and upper-ocean freshening is not validated (table 1 reports no observed SSS trend to compare against). Given the number and magnitude of the issues raised throughout this review, “realistically captures” considerably overstates the model’s demonstrated performance.*

Response: We agree that the original conclusion was too strong compared to the evidence presented. We have revised the conclusion to provide a more balanced statement of demonstrated capability and limitations. We have also avoided describing the simulated SSS tendency as an independently validated observed freshening trend, because the reference dataset used for SSS does not provide a corresponding observed trend in Table 1. The central conclusion is now that TOPAZ5 provides a useful high-resolution coupled framework for North Atlantic-Arctic Ocean and sea-ice modelling and forecasting, while the validation identifies specific limitations that remain important for interpretation and further development.

We thank the reviewer again for the thorough and constructive review that has substantially improved our manuscript. We think that these revisions address the scientific concerns raised and that the manuscript now provides a clear, transparent and appropriately scoped evaluation of the TOPAZ5 configuration.