Authors point-to-point responds Community Comment #2 to egusphere-2025-2665

Please find the author's responses in black below the reviewer's comments in blue. The italicized text within quotation marks indicates the proposed revisions in the revised manuscript.

1. Why do the authors particularly reply on the two time periods of 1960-1980 and 1995-2000? Line 160: I dont understand how it can be verified through overlapping period consistency checks (1995–2020).

(1) Regarding the choice of time periods (1960–1980 and 1995–2020):

We appreciate your question regarding the selection of time periods. The division into these two specific periods is based on the following scientific rationale:

- 1995–2020 serves as the core validation period in this study. This interval represents the satellite era, which provides a wealth of remote sensing sea ice data, diverse reanalysis products, and in situ observations, allowing for robust validation of the model's performance. Furthermore, this period covers a phase of accelerated global warming, making it critical for evaluating climate model capabilities.
- 1960–1980 is used as a representative historical period prior to the satellite era, enabling an examination of the model's performance under earlier climatic conditions. Additionally, comparing this period with 1995–2020 helps illustrate the model's response to differing climatic forcings, facilitating a comparative analysis.
- Simulating the entire period from 1960 to 2020 would require prohibitive computational resources, which are beyond the current capacity of our project. Therefore, conducting segmented simulations represents an optimal strategy for balancing scientific objectives with feasibility.

(2) Regarding the question you raised near L160

Regarding the mention of "overlapping period consistency checks," we apologize for any lack of clarity in the original phrasing. We aimed to convey that, under computational constraints, we prioritized the satellite era (1995–2020) for high-confidence validation due to the abundance of observational data (satellite, reanalysis, and in situ) available during this time. The term "consistency check" specifically refers to cross-validation among these multi-source observations. We have revised the manuscript to state this more precisely:

"*Given the prohibitive computational cost of a continuous high-resolution simulation from 1958 to 2020, we adopted a strategic two-period integration scheme to prioritize computational resources for our core analysis period (1995–2020). The model's climatological fidelity during this satellite era is verified using multi-source observational data, ensuring a reliable assessment of both sea ice and ocean variability.*" (P6, L165-168)

2. Line 187: any SSS data below sea ice? Is there any justification?

Thank you for raising this important technical limitation. We confirm that OISSS and other satellite-derived SSS products (e.g., SMAP) cannot provide valid salinity observations under sea ice cover. Their retrieval algorithms actively mask areas with SIC greater than 15%, which are flagged as missing values (NaN) in the datasets.

In the revised manuscript, we continue to use OISSS for validation but explicitly state that SSS comparisons are performed only for open-water areas (SIC < 15%). A spatial mask is applied using the native missing value flags (NaN) from OISSS. At each monthly evaluation time step, statistical calculations are performed only on grid points where valid data exists in both the model output and the observations; regions with sea ice cover (OISSS missing values) are excluded from all analyses.

The vague description in the original Section 2.2.2 has been replaced with a precise statement (P8, L233):

"*For sea surface salinity (SSS), …*" has been changed to:

"*For open-water sea surface salinity (SSS) validation (SIC<15%), …*"

3. The major issue is the explanation of the causal analysis throughout the manuscript. For example, the major conclusion of "These systematic biases may be attributed to three principal sources: inadequate representation of eddy dynamics, limitations in mixing parameterizations, and insufficient resolution of cross-scale interactions in key gateways (e.g., Fram Strait) " is not convincing. Any sensitivity experiments can be considered to support the findings?

We sincerely thank you for this critical comment. Regarding the statement in the original abstract ("These systematic biases may be attributed to three principal sources: inadequate representation of eddy dynamics, limitations in mixing parameterizations, and insufficient resolution of cross-scale interactions in key gateways (e.g., Fram Strait)."), we provide the following detailed clarifications and context, which have been incorporated into the relevant sections of the revised manuscript. We also acknowledge the limitation of not being able to conduct definitive sensitivity experiments.

(1) "inadequate representation of eddy dynamics"

1) E3SMv2-MPAS failed to capture the observed significant decadal freshening signal in the upper ocean of the Amerasian Basin during the 2000s–2010s (Fig. 13 in the revised manuscript). We have added supporting evidence and references at the relevant location (P23, L502-510):

"*The simulated salinity biases may be related to the use of an inappropriately high and constant isopycnal diffusion coefficient (κ=300 m²/s) in the GM parameterization. This high diffusion coefficient likely results in excessively strong along-isopycnal mixing, which oversmooths horizontal salinity gradient fronts formed by freshwater accumulation (e.g., from melting ice and increased runoff). During the 1970s, when background freshwater signals were relatively weak, the effect of strong diffusion was less pronounced. However, under the strongly increased freshwater input in the 2000s–2010 (Polyakov et al., 2013; Wang et al., 2019), the persistently high κ value continuously and excessively diffused the simulated upper-layer low-salinity anomalies, hindering their realistic accumulation and maintenance in the basin upper layer. As a result, the model significantly underestimates the magnitude of decadal freshening observed in the region.*"

2) The model did not reproduce the observed seasonal variation of the Atlantic Water, characterized by a warmer and thicker in winter compared to summer (Fig. 12). Corroborating discussion has been added in the text (P21, L468-474):

"*This discrepancy may be attributed to the GM parameterization scheme, which models mesoscale eddy effects on heat and salt redistribution through bolus advection and Redi diffusion. In general, the Arctic winter features greater mixed layer depth and weaker stratification due to brine rejection during sea ice formation and wind-driven stirring (Peralta-Ferriz and Woodgate, 2015). These processes promote eddy penetration, increasing the efficiency of vertical heat transport. In contrast, strengthened stratification in summer restricts the vertical scale of eddies and reduces heat transfer. However, the GM scheme employs a fixed diffusion coefficient, which prevents it from capturing the seasonal variability modulated by stratification changes.*"

(2) "limitations in mixing parameterizations"

1) Co-located biases in SST, SSS, and SIC in the Barents and Greenland Seas. Supporting discussion has been added (P17-18, L404-413):

"*In the Greenland and Barents Seas, systematic underestimation of SST and SSS (Figs. 8c and 9c) coincides with overestimation of SIC (Figs. 3 and 4). These regions are situated within the marginal ice zone, where strong surface wind stress facilitates the transfer of energy to deeper ocean layers through the excitation of near-inertial oscillations and associated turbulent mixing processes*

(D'Asaro, 1985). This discrepancy may be attributed to the model's potential overestimation of this downward energy transfer. Similarly, Zhu et al. (2022) reported that in the equatorial Pacific cold tongue region, the KPP scheme overestimates downward turbulent heat flux, leading to a cold bias in both upper-ocean and sea surface temperatures. A primary reason for these biases lies in the scheme's reliance on a single Richardson number (Ri) relationship for parameterization. Although this approach captures instability conditions in stratified shear flows, it is insufficient to uniquely determine turbulent states and mixing intensities (Zhu et al., 2022), thus limiting its performance in complex dynamic environments."

2) Misplaced warming layer in the Eurasian Basin: Observational data indicate warming occurred primarily in the upper~500m on decadal scales, whereas the model simulated the warming predominantly in the 200–1000m layer. Relevant evidence is now discussed in the manuscript (P23, L489-497):

   "*These discrepancies may be partly attributed to biases in the representation of vertical processes. As indicated by sensitivity experiments such as those of Liang & Losch (2018), enhanced vertical mixing could promote upward heat transport from AW, potentially causing cooling at intermediate depths (200–900 m). Our model uses a relatively low background diffusivity (1.0×10⁻⁵ m²/s), which remains constant across time periods despite evidence that Arctic amplification and Atlantification in the 2000s–2010s (Polyakov et al., 2017, 2025; Rantanen et al., 2022; Richards et al., 2022; Shu et al., 2022) may have strengthened vertical mixing compared to the 1970s. The model's failure to represent this temporal increase in mixing efficiency might have limited upward heat transfer, confining warming mainly to intermediate and deeper layers—consistent with the underestimation of shallow warming and exaggerated deep response seen in our simulations.*"

(3) "insufficient resolution of cross-scale interactions in key gateways (e.g., Fram Strait)"

We acknowledge that the original phrasing might have been ambiguous. Our intended meaning was that shortcomings in both vertical mixing and mesoscale eddy parameterizations could potentially be mitigated by increased resolution, particularly in narrow yet critical gateway regions like Fram Strait.

To address this more thoroughly, we have added a new section in the Discussion (Section 5.2: Sources of Systematic Biases and Trade-offs Between Resolution and Parameterizations; P37, L775-798):

"*Analyses in Section 3 not only discussed the simulation biases of E3SMv2-MPAS but also traced their potential origins. For most biases, the primary causes can be attributed* *to inadequacies in physical parameterizations. First, the inadequate representation of eddy dynamics is a key source. For instance, the underestimation of freshening in the Amerasian Basin may result from the use of a fixed eddy diffusivity (κ=300 m²/s in the Arctic), which oversmooths salinity fronts. Similarly, the model's failure to capture the seasonal variability of the Atlantic Water layer likely stems from the invariant κ in the GM scheme, which cannot respond to the seasonal cycle of sea ice retreat and associated changes in stratification. Second, limitations in vertical mixing parameterizations act as another key source. The coordinated biases in SST, SSS, and SIC in the Greenland and Barents Seas, for example, may arise from the inherent limitations of the KPP scheme's single Ri-based approach in defining turbulent states and mixing intensities within complex dynamic environments. Additionally, the misrepresentation of the warming layer in the Eurasian Basin could be linked to inappropriate background diffusion coefficients within the KPP framework.*

Increasing model resolution presents an effective pathway to reduce reliance on empirical parameterizations by more directly resolving key physical processes, such as mesoscale eddies. Enhanced resolution can, to some extent, mitigate the inaccuracies of existing schemes. For instance, studies have shown that higher resolution improves the simulation of the Atlantic Water layer's temperature, thickness, spatial distribution, and its decadal warming trends (Wang et al., 2024). However, the small Rossby radius of deformation (often ≤3 km) in the Arctic (Veneziani et al., 2022) implies that even with computationally feasible resolution increases, critical processes (e.g. mesoscale eddies, vertical mixing, and ice-ocean interactions) may remain under-resolved (Chassignet et al., 2020; Wang et al., 2018). Therefore, the development of more advanced physical parameterizations remains imperative. It is noteworthy that resolution increases have proven effective in improving the simulation of volume, heat, and freshwater transports through critical gateways such as the Fram Strait and Davis Strait (Wang et al., 2024). The Fram Strait, in particular, serves as a pivotal channel for Atlantic heat influx into the Arctic Ocean (Herbaut et al., 2022; Pnyushkov et al., 2021). In conclusion, we propose that a cost-effective strategy involves targetedly increasing resolution in key gateway regions while concurrently refining parameterizations for mesoscale eddies and vertical mixing."

Due to significant constraints in computational resources, we have been unable to perform the sensitivity experiments that would be ideal for conclusively proving the sources of the biases. Therefore, the attributions discussed above are currently supported by evidence from the model-observation comparison and references to existing literature. Acknowledging that these attributions remain suggestive rather than proven, we have removed the specific claim regarding the three sources from the abstract. The supporting evidence and reasoning related to these potential sources have been showed in the discussion of biases within Section 3, and in the new Section 5.2 of the Discussion, where we present them as plausible explanations and areas for future investigation, not as definitive conclusions.