Referee #2

We would like to thank the reviewer for the constructive comments. Our responses are given below in black type (reviewers comments in blue italics).

The manuscript presents a global ocean modelling study using the NEMO framework to evaluate the performance of an Eddy-Diffusivity Mass-Flux (EDMF) parameterization for vertical mixing. The study compares EDMF with traditional eddy-diffusivity-based schemes (e.g., EVD and TKE-based closures) in simulating ocean temperature, salinity, and mixed-layer depth (MLD), with a particular focus on convective processes across different regimes.
The work is timely and relevant. Representing convection and nonlocal mixing remains a major challenge in ocean and climate models, and the implementation of a mass-flux framework inspired by atmospheric convection is a promising direction. The manuscript is generally well structured and clearly presents the modelling framework, experimental setup, and evaluation strategy. I have identified several major and minor issues that the authors should address. Addressing these points will significantly improve the clarity, physical interpretation, and overall quality of the manuscript.

Q2.A)
General Comment (Convective Regimes): The manuscript does not clearly distinguish between different convective regimes (e.g., shallow tropical vs deep polar convection), despite their fundamentally different plume dynamics and forcing. It remains unclear whether the EDMF formulation appropriately captures these regime-dependent processes or applies a uniform parameterization across all conditions. Please clarify how the scheme represents these contrasting regimes and whether its performance varies between them.

In the Introduction (lines **49-59**), we provide definitions of the two convective regimes in terms of their dynamics and forcing. In section ***2.2 Vertical parameterization options***, we present the equations and coefficients of EDMF, which are applied uniformly across convective regime. The scheme does not use regime-specific formulations; the resulting convective fluxes vary with the local stratification and surface forcing, leading to different convective responses under shallow-diurnal and deep-seasonal conditions.
We then assess the model performance separately for the two contrasting convective regimes in sections ***3.1.1 Deep-seasonal convection*** and ***3.1.2 Shallow-diurnal convection***. Using model outputs at 1-hour resolution, we show that EDMF simulation captures the timing, penetration depths and intensities of both convective regimes. The ability of EDMF to reproduce these two distinct regimes is further summarized in section ***5 Discussion and Conclusion***, lines **730-738**. A quantitative assessment of whether EDMF performs better in one convective regime than in the other would require long-term, high-resolution observational datasets identical for both regimes, which are currently not available.
We have now clarified in the manuscript that the EDMF parameterization uses a unique formulation across convective regimes, while its response to local forcing and stratification allows it to represent both shallow diurnal and deep seasonal convection, lines **160-163**, **726**.

Q2.B) Finally, the manuscript does not discuss the role of double-diffusive convection, which can be important in stratified regions of the ocean where temperature and salinity gradients compete. It is unclear whether the EDMF formulation is capable of representing double-diffusive processes (e.g., salt fingering or diffusive convection), or whether these effects are entirely neglected. Please clarify *whether the scheme accounts for double-diffusive mixing and include a brief discussion of its potential impact on the results.*

In the ocean, double-diffusive convection is a phenomenon in which water movements are generated by the difference in diffusion rates between heat and salt. Heat diffuses (Kh) approximately 100 times faster than salt (Ks). This disparity creates “salt-fingers” or “thermohaline steps” types of instability, which occur within stable regimes. EDMF inherently accounts for the double-diffusion phenomenon through its ED component, where distinct values for Kh and Ks can be specified. However, under unstable conditions, deep convection simultaneously transports heat and salt during turbulent mixing. As a consequence, there is no separate convective vertical velocity (w_p) for heat and salt. We have now addressed the issue of the interaction between convection and double-diffusion, lines **777-780**.

Specific comments :
Q2.C) Fig. 4: Figure 4 shows a distinct structure around ~60 m depth that is not captured by any of the models. This feature may be influenced by processes such as inertial oscillations or internal waves, which are not explicitly represented in the current parameterizations. Please clarify the physical origin of this structure and discuss whether its absence reflects limitations of the vertical mixing schemes or missing wave-driven processes.

We suspect that these structures are associated with tidal variability, which can generate oscillations near the thermocline. Explicit tides are not represented in our simulations, and this missing process may therefore contribute to the inability of the model to reproduce the observed structures. This point has now been clarified in the revised manuscript (lines **400–401**).

Q2.D) Fig. 5: Figure 5 reveals substantial and systematic temperature biases in the subsurface (thermocline/pycnocline) across all parameterizations, suggesting overly strong vertical mixing below the mixed layer and an unrealistic downward transport of heat that weakens stratification. Given the central role of the thermocline in regulating heat storage and vertical exchange, it remains unclear whether the EDMF scheme provides a physically meaningful improvement in vertical structure, despite improvements in mixed layer depth. Please clarify the origin of these subsurface biases and discuss whether they arise from excessive mixing across the base of the mixed layer or from missing ***restratification processes****, as addressing this would strengthen the overall interpretation of the results.*

The comparison between parameterizations suggests that the subsurface (~ 50 m depth) cold biases cannot be attributed systematically to excessive mixing below the mixed layer (Fig. 5c,e,g). In some cases, increasing turbulent transport below the mixed layer actually reduces the bias.

In the tropical ocean, both EVD simulations exhibit similar cold subsurface bias structures, although the addition of the MLP parameterization reduces the magnitude of the maximum bias by approximately 0.25 °C (Fig. 5c). Since the MLP parameterization introduces an additional TKE source below the mixed layer, the reduction of the cold subsurface bias in the TKE+MLP+EVD simulation suggests that insufficient vertical mixing below the mixed layer contributes to the tropical subsurface bias, rather than excessive mixing. By increasing the penetration of turbulent mixing below the mixed layer, MLP allows a greater downward redistribution of heat and consequently reduces the cold subsurface bias.
A similar mechanism can explain the improvement obtained with EDMF. By allowing the penetration of mixed warm waters to greater depths during nighttime, and therefore enhancing nighttime

destratification to greater depth (Fig. 4), EDMF increases the downward extent of convective mixing and reduces the subsurface cold bias (Fig. 5c). The daytime restratification is comparable in the three simulations, despite their different nighttime destratification depths (Fig. 4b–d). Therefore, the comparison does not provide evidence that differences in daytime restratification between the simulations are the primary cause of the tropical subsurface cold biases.

A similar behaviour is found in the northern and southern hemisphere oceans (Fig. 5e,g), although the magnitude and vertical structure of the biases differ from those in the tropical ocean. In both regions, TKE+EVD exhibits a cold subsurface bias, which is also present in the EDMF simulation. The addition of the MLP parameterization substantially reduces this cold bias, indicating that the additional TKE source below the mixed layer can improve the representation of vertical heat redistribution in these regions as well. This result is again more consistent with insufficient, rather than excessive, turbulent mixing below the mixed layer as one contributor to the subsurface cold bias.
However, the comparison (Fig. 5e,g) also shows that increasing mixing below the mixed layer does not systematically lead to a better representation of the subsurface temperature structure. In the northern hemisphere ocean, the additional mixing associated with MLP largely compensates the cold bias and produces a warm bias (Fig. 5e), suggesting that the enhanced turbulent transport may become excessive at some depths. This highlights that insufficient mixing can contribute to the cold bias, whereas excessive mixing can result in an over-compensation.
For the northern and southern Hemisphere oceans, we do not have diagnostics that allow us to quantify the contribution of restratification to the remaining biases. Therefore, we cannot exclude a contribution from differences in the balance between mixing and restratification in these regions.

Overall, these results suggest that the subsurface temperature biases cannot be attributed systematically to excessive vertical mixing below the mixed layer. In the tropical ocean in particular, the reduction of the bias when additional turbulent mixing is introduced below the mixed layer indicates that insufficient mixing contributes to the cold subsurface bias.
In any case, the goal of EDMF is to move away from EVD and MLP tricks in order to physically represent turbulence at the base of the mixed layer. We recognize that this is a challenging task, as the model was originally tuned using EVD and MLP. Furthermore, we know that convection interacts strongly with eddy-induced restratification. In the future, we plan to use the GEOMETRIC parameterization to represent sub-mesoscale restratification in the NEMO code. Consequently, the next step will naturally involve parameter tuning using EDMF, GEOMETRIC, and likely other new parameterizations. Text was added lines **764-771**.

Q2.E)Fig. 7: The manuscript does not adequately discuss model performance in western boundary current regions (e.g., Gulf Stream, Kuroshio), where noticeable MLD biases appear in Fig. 7. These regions are strongly influenced by lateral processes (eddies, fronts, advection), and not solely by vertical mixing. Please clarify whether the reported differences with EDMF in these regions reflect improvements in vertical physics or are influenced by unresolved horizontal dynamics at the given resolution. A brief discussion addressing this point would strengthen the manuscript.

We thank the reviewer for highlighting the specific challenges associated with western boundary current (WBC) regions. We already mention in the manuscript (lines **485–487**) that both vertical mixing parameterizations exhibit similar warm temperature biases in these regions. These biases are likely related, at least in part, to deficiencies in the representation of low-level clouds in state-of-theart climate models, which affect the associated longwave and shortwave radiative fluxes (Zhang et al., 2023).

At the global scale, the difference in temperature biases between EDMF and the two EVD-based simulations is predominantly negative when considered in absolute terms (Fig7i,j,k), indicating an overall reduction of temperature biases with EDMF. However, positive differences remain locally in the Gulf Stream and Kuroshio regions. These residual biases should be interpreted with caution, as WBC are strongly affected by horizontal processes, including mesoscale eddies, fronts, and lateral advection, which are not fully resolved at the 1/4° horizontal resolution of our simulations. Consequently, the remaining positive anomalies do not necessarily indicate a deficiency of the vertical mixing scheme, as they may partly reflect errors in the representation of horizontal dynamics and air–sea interactions.

Further calibration of the EDMF parameters may improve the overall model performance in WBC; however, tuning the scheme specifically to compensate for biases associated with the unresolved dynamics of WBC could lead to over-compensation and a loss of physical consistency elsewhere. We have therefore added a discussion of this limitation in the revised manuscript, lines **487-491**. This point is also consistent with our response to Reviewer #1, Q1.D, regarding the need to avoid over-interpreting reductions in model biases as evidence of improved physical realism.

Q2.F) Justification of MLD criterion: The authors should (i) justify the use of the Holte et al. method more explicitly, and (ii) assess the robustness of their conclusions by comparing with at least one alternative MLD definition (e.g., a fixed density threshold). This would strengthen confidence that the reported improvements with EDMF are physically meaningful rather than diagnostic-dependent.

Following the reviewer's comment and that of Reviewer #1 (see our answer to Q1E), we decided to remove the comparison to the Holte et al. (2017) datasets, which are not adequate for comparisons with climatology of global circulation models. Please see our new Fig. 8 and text lines **526-536**.

Q2.G) Fig. 8: While Fig. 8 demonstrates improved seasonal MLD variability, this metric alone does not uniquely validate the underlying plume dynamics. Similar improvements could arise from compensating errors in parameter tuning. Additional diagnostics (e.g., vertical buoyancy flux, entrainment rates, or plume energetics) are required to confirm that EDMF captures the correct physical processes.

In the present study, the three simulations use the same model parameters, with the exception of those defining the respective vertical mixing parameterizations. Thus, the differences between the simulations arise from their different representations of vertical mixing rather than from differences in parameter tuning. Following this comment and that of Reviewer #1 (Q1.C), we have added Fig. 2b and Fig. S1, which provide additional diagnostics of the vertical heat fluxes. Figure 2b separates the diffusive and convective contributions to the total vertical heat flux in the EDMF simulation, thereby directly illustrating the contribution of the mass-flux component to vertical heat transport. Figure S1 compares the intensity and vertical extent of the total vertical heat fluxes across the three simulations under the same surface forcing. These diagnostics provide additional evidence that the improved MLD variability in EDMF is associated with a distinct representation of convective transport, rather than solely with compensating errors from parameter tuning. Please refer also to lines **214–220** and **316–339**.

Q2.H)Fig. 9: In Fig. 9f,g, only model results are shown without comparison to observational data. Given that these panels appear central to interpreting the behavior of the EDMF scheme, it is unclear how well the simulated patterns reflect reality. Please clarify why observational data are not included for these diagnostics, and whether suitable datasets exist for comparison. Including or discussing observational constraints would strengthen the credibility of the results.

We thank the reviewer for this comment. We agree that observational datasets are important for assessing the realism of the simulated MLD. However, the purpose of Fig. 9 is not primarily to demonstrate that EDMF produces a more realistic MLD than the other parameterizations, but rather to provide insight into the mechanisms underlying the differences in MLD.
Observational estimates of MLD are available in the Labrador Sea. For example, Holte et al. (2017) provide an Argo-based MLD climatology as well as MLD estimates for individual Argo profiles. However, these observations are not directly comparable to the quantity shown in Fig. 9. Our simulated MLD time series are calculated as spatial averages over a box encompassing the mooring location and extending over both the deep Labrador Sea and shallower regions over the continental slope. In contrast, Argo observations provide discrete, irregularly distributed measurements at individual locations and times, with the observational sampling being strongly heterogeneous in space and time. In particular, the Argo array samples the deep basin (depths > 2000m), while the shallower regions of the continental slope are not sampled. Consequently, directly comparing the model box-averaged MLD with an Argo-based time series would require a dedicated space- and time-collocation of individual Argo profiles with the model output, including the computation of MLD from each collocated model profile using the same criterion as applied to the observations. We considered that this detailed model–observation comparison was beyond the scope of the present study.
Nevertheless, the Argo-based results of Holte et al. (2017) provide an independent observational constraint on the seasonal and interannual variability of MLD in the Labrador Sea, and we now discuss this observational context in the revised manuscript. We have also clarified in the text (lines **579-583**) that Fig. 9 is intended primarily as a diagnostic of the mechanisms responsible for the differences between the parameterizations, rather than as a direct assessment of the realism of the simulated MLDs.