Summary and Overall Assessment

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.

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.

Specific Comments:

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.

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.

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.

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.

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.

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.

Additional Comment: 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.