Reviewer # 3

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

General comments

#1 First, the motivation of this study could be more clearly stated in the introduction. This study seems to be a follow up study of the authors' previous work (Giordani et al., 2020), in which they implemented the EDMF scheme in NEMO and tested it in a set of 1D simulations. This study tested the same EDMF scheme in a global 3D simulation and compared it with the default NEMO configuration with EVD and MLP. It would be helpful to provide more discussion on the existing issues of EVD and MLP with appropriate references, especially in terms of their performance in global simulations. It would also be helpful to provide some discussion on the additional insights the tests in a global 3D simulation offered than the 1D tests described in Giordani et al., 2020 in both the introduction and later when describing the results. I was surprised to see so little reference to the 1D results of Giordani et al., 2020 when the authors described their 3D results.

We thank the reviewer for this comment.
First, we agree that the present manuscript lacks discussion on existing issues of EVD and MLP. Despite providing satisfactory results in many applications, EVD+MLP has an important conceptual limitation: convective transport remains represented through a local diffusive formulation, while the MLP introduces an empirical correction to enhance mixing below the mixed layer. In terms of model performance, MLP was initially introduced to deepen the mixed layer, which was commonly found to be too shallow in the Southern Ocean during summer under windy conditions in simulations using EVD only (Rodgers et al., 2014). Subsequent studies showed that EVD+MLP can improve the representation of MLD in the Southern Ocean (Heuzé et al., 2015; Rodgers et al., 2014), and that regionally tuned MLP parameters can further reduce near-surface temperature and salinity biases and improve sea-ice representation (Storkey et al., 2018). Similarly, MLP tuned for the Arctic region improves the representation of wintertime MLD, although summer MLD remains underestimated (Allende et al., 2024). In their global simulations, Calvert and Siddorn (2013) also reported reductions in SST and MLD biases in the extratropical ocean when MLP was included, while highlighting uncertainties associated with the choice of MLP parameters.
Importantly, Storkey et al. (2018) described the use of MLP as a "stopgap solution" to compensate for underestimated mixing processes in the UK Global Ocean GO6 configuration, and highlighted the need for more physically based parameterizations in future model developments. This raises the question of whether the improvements obtained with MLP can instead be achieved through a more physically based representation of convective transport. This question motivates the present study, in which we evaluate the EDMF parameterization, which explicitly represents the non-local transport associated with convective plumes, in a global 3D ocean configuration.
This aspect is now discussed in the Introduction lines **80-101**.

Second, we also agree that the motivation for extending the work of Giordani et al. (2020) to a global 3D configuration should be made clearer. Giordani et al. (2020) provided a detailed assessment of the EDMF formulation in idealized 1D column simulations, which allowed the behaviour of the

parameterization to be evaluated in a controlled framework. In particular, Giordani et al. (2020) showed that EDMF can reproduce the observed depth and intensity of convective mixing in both moderate and strong convection test cases, providing a basis for evaluating its behaviour in a global configuration. The present study is intended as a complementary step towards assessing its behaviour and impact in a realistic global ocean configuration. The aim is therefore not to reproduce the 1D assessment, but to investigate how the EDMF formulation behaves when embedded in the full 3D ocean system. Indeed, the ultimate goal is to activate EDMF in the operational ocean forecasting systems operated at MOI, as well as to participate in the upcoming CMIP-7 exercise.
A key limitation of the 1D framework is that horizontal advection is neglected. Although this simplification is useful for isolating the vertical mixing processes, it limits the extent to which the simulated evolution can represent the behaviour of an oceanic water column over long timescales. In a global 3D configuration, horizontal advection and ocean circulation are explicitly represented, allowing water masses and their properties to be transported between regions and enabling the EDMF scheme to be evaluated over longer timescales and across a wide range of dynamical regimes.
This 3D framework therefore allows us to investigate aspects that cannot be addressed with the 1D experiments of Giordani et al. (2020), including the seasonal evolution of convection, its spatial variability, its interaction with the large-scale circulation, and its impact on upper-ocean properties at regional and global scales. In particular, the present study examines both shallow tropical convection and deep convection in the subpolar North Atlantic, and assesses how the different vertical mixing formulations affect mixed-layer depth, temperature biases, and surface heat fluxes over a 16-year simulation. This is now clarified in lines **111-113**, **116-121,** in the objectives lines **124-134**.

#2 Second, the presentation of the results in this manuscript is very confusing and could benefit a lot from a careful revision. Here are some examples (see more in my specific comments below). (1) The first two tests are essentially at a single location (mooring sites in the Labrador Sea and equatorial Pacific). This would be a great place to discuss the differences in a 1D test and a 3D test and reference to the results of Giordani et al., 2020. What is the additional insights about EDMF a 3D simulation provided? (2) The comparison with observational data in these two tests was very little and I wonder why? (3) In the comparison of global mean temperature profiles in Section 3.2, rather than emphasizing on the bias reduction (which is not obvious because the bias is smaller somewhere but larger elsewhere, and subject to changes in likely required retuning of other parameters), it would probably be more helpful to discuss how the changes in the simulation results are consistent with a better representation of convective mixing in EDMF. (4) The presentation in Section 4 is very confusing and I had a hard time following the discussions there.

(1) We thank the reviewer for highlighting the opportunity to better connect these two case studies with the 1D experiments of Giordani et al. (2020). The purpose of the present experiments is not to reproduce the same test cases, but to provide a complementary assessment of EDMF in a fully 3D ocean configuration.
The two sites considered in our study were selected because they represent two particularly important and contrasting convective regimes for the global ocean: deep wintertime convection in the Labrador Sea, which contributes to the formation of deep water masses, and shallow diurnal convection in the tropical ocean, which occurs repeatedly over a large fraction of the tropical ocean and strongly affects the daily evolution of the upper-ocean mixed layer. These regimes complement the test cases considered by Giordani et al. (2020), who focused on the Gulf of Lions and the PAPA station.
The choice of the Labrador Sea rather than the Gulf of Lions also reflects the limitations of our global 1/4° configuration in the Mediterranean Sea. The spatial resolution and, in particular, the representation of the atmospheric forcing are not sufficient to adequately reproduce the strong local

forcing associated with the winter 2012/2013 deep-convection event in the Gulf of Lions. We therefore considered the Labrador Sea a more appropriate region for assessing deep convection in the present global configuration.
We have clarified this connection with Giordani et al. (2020) and the specific added value of the 3D experiments in the revised manuscript (in the Introduction, and in lines **302-307**).

(2) We agree that the observational constraints should be made more explicit. However, the availability and nature of observations differ substantially between the two cases. In the Labrador Sea, direct in situ observations of deep convection are sparse and irregular in both space and time. Available Argo profiles cannot be directly compared with the model time series without a dedicated space–time collocation procedure and the application of consistent diagnostics to model and observations. We therefore compared the simulated timing, occurrence, depth and intensity/speed of convective events with observational estimates and previous observational studies of Labrador Sea convection (e.g. Kieke and Yashayaev, 2015; Piron et al. 2017; Clément et al. 2024).
In the tropical Pacific, we use high-frequency observations from the TAO mooring at 2°S, 140°W to compare the observed and simulated evolution of the diurnal warm layer and nighttime convective mixing. These observations provide a detailed constraint on the temporal evolution of shallow convection and were therefore used extensively in the corresponding analysis. We have clarified and strengthened the observational comparisons in **lines 337-341, 437-439**.

(3) Following the reviewer's suggestion, we added further discussions to the text in Section 3.2 (lines **483-482, 493-494, 554-559, 568-569**). Please also refer to our answer to specific comments below.

(4) This section aims to assess the ocean's adjustment to the EDMF parameterization compared to EVD. In particular, adjustments in mean temperature and surface fluxes are quantified and physically explained by entrainment anomalies. We highlight a surface flux adjustment amounting to 6.5% of the global net surface heat flux climatology, which is a significant result. Following the reviewer's suggestion, we have added an introduction to Section 4 (lines **656-659**) explaining the purpose of the subsequent analysis, and added sub-section headers to ease the reading. We also modified/simplified several sentences throughout this section.

#3 Third, the authors included a case with TKE+MLP+EVD in the comparison whereas the comparison of the results were mostly between EDMF and TKE+EVD. I understand that TKE+MLP+EVD was the default configuration in NEMO so it may be helpful to compare with that. But I think it would be helpful to include more discussion on the role of MLP on the simulation results. The authors mentioned that MLP was introduced to add additional TKE to represent the effects of near-inertial oscillations and surface gravity wave breaking. Would it be necessary to include these effects in the diffusive component of EDMF as well? On the other hand, the authors stated that "EDMF eliminates the need for empirical adjustments such as MLP" in Section 5, which makes it even more confusing how EDMF accounted for the effects that was supposed to be accounted for by MLP?

We thank the reviewer for this comment.
First, we would like to clarify that the main comparison in this study is between EDMF and the TKE+MLP+EVD simulation, rather than between EDMF and TKE+EVD. The TKE+MLP+EVD configuration corresponds to the default NEMO configuration and represents the state-of-the-art EVD-based configuration against which we evaluate the EDMF parameterization. We have clarified this point in the objectives of the revised manuscript (lines **129–130** and **138–139**).

The TKE+EVD simulation is nevertheless retained as an additional reference because it allows us to better isolate the effects associated with the different convective mixing schemes. It is also specifically used in Section 4 to quantify the adjustments of surface heat fluxes associated with the change from the EVD to the EDMF vertical mixing scheme.
Following the reviewer's suggestion, and also in response to the specific comment on the regulation of the MLD, we have expanded the discussion of the role and effects of the MLP parameterization in the revised manuscript. In particular, we now explain in more detail how MLP suppresses residual instabilities within the mixed layer, thereby promoting a more homogeneous mixed-layer and reducing the occurrence of conditions triggering the EVD scheme. The resulting reduction in EVD activation limits further deepening of the mixed layer and leads to shallower MLDs (lines **624-627**). We have also added discussion of the MLP parameterization and its limitations in lines **85–101** and further discussed its effects on temperature biases in lines **555-559** and **568-569**. While a detailed assessment of MLP itself is beyond the main scope of this study, these additions clarify its role in the TKE+MLP+EVD configuration and its relevance for the comparison with EDMF.

Second, regarding the role of MLP in representing the effects of near-inertial oscillations and surface gravity wave breaking, we would like to clarify the conceptual difference between the two approaches. The MLP parameterization introduces an additional source of TKE in the mixed-layer intended to represent unresolved vertical mixing associated with these processes. In contrast, EDMF is formulated to respond to density instabilities regardless of their origin. Therefore, when processes such as near-inertial motions or wave-driven mixing contribute to destabilizing the density profile, the resulting convective instability can be acted upon by EDMF without requiring a separate empirical source of TKE (as in MLP). In this sense, the effects of MLP that manifest through the generation of density instabilities are implicitly accounted for by the EDMF formulation. We have clarified this in the revised manuscript, lines **197**. It seems also important to recall here that MLP is used jointly with EVD, which already increases the eddy-diffusivity (Kz) by several orders of magnitude in response to unstable stratification. However, these very large, physically unrealistic Kz values were found to be insufficient to fully deal with these instabilities, and therefore MLP was introduced.
Nonetheless, EDMF does not explicitly represent near-inertial oscillations or surface gravity wave breaking themselves, nor does MLP. Processes that generate mixing without producing a density instability are therefore not necessarily represented by the EDMF formulation. We also modified the statement in Section 5 from "EDMF eliminates the need for empirical adjustments such as MLP" into "The EDMF formulation removes the need for an additional empirical adjustment such as MLP within the convective mixing scheme, as density instabilities are directly used to activate the mass-flux formulation regardless of their origin.", lines **835-836**. By this statement, we mean that with EDMF, no additional MLP-type adjustment is required to trigger or sustain mixing associated with density instabilities, because EDMF responds to convective instability regardless of its origin, including that induced by surface cooling or salinification, Ekman buoyancy flux at fronts,…

Specific comments

L16-17: It would be helpful to point it out that the comparison is between the MF component of EDMF and EVD since in both EDMF and EVD the same TKE closure is used for the diffusive mixing? It would also be helpful to clarify on the reason to include TKE+EVD+MLP in the comparison. Should MLP (or a variant of it) be included in EDMF as well?

We thank the reviewer for this suggestion. We prefer to keep the comparison as presented, as the EDMF configuration is evaluated against both the TKE+EVD and TKE+MLP+EVD configurations.

The TKE+EVD configuration provides a useful reference to isolate the differences associated with the convective mixing scheme, while TKE+MLP+EVD represents the default NEMO configuration and therefore provides a relevant benchmark for evaluating the EDMF parameterization.
Regarding the rationale for including TKE+MLP+EVD and whether the processes represented by MLP should also be included in EDMF, we refer the reviewer to our response to General Comment #3, where these points are discussed in detail. In particular, we clarify the respective roles of MLP and EDMF and explain why EDMF does not require an additional MLP-type adjustment within its convective mixing formulation.

L43: Did the authors mean a single convective event lasting for a time period on a seasonal or interannual time scale? I guess the authors meant that the intensity of convective events can vary on these time scales?

We mean the duration of convection events can range from diurnal (e.g. convection occurs every night in the tropics) to seasonal (e.g. winter time deep convection in Labrador/Greenland Sea) to interannual timescale (e.g. polynia-induced convection in Antarctica).

L74-76: It should be mentioned that Garanaik et al. (2024) also proposed to represent non-diffusive convective mixing using Mass-Flux.

Correction added line **114**

L76-81: It would be helpful to be explicit on what is the shortcoming of these schemes mentioned by the authors (KPP, EVD). In particular, what is the problem of EVD exactly apart from the fact that it is essentially a diffusive mixing model? Why it is absolutely necessary to explicitly present convective mixing by something like a mass-flux model?

Diffusion and convection are different processes, as a consequence it is inappropriate to account for the convection process through diffusion.
The shortcoming of the schemes and the motivations to explicitly represent convective mixing are added lines **74-76**, and **80-83**.

L91-92: What would the authors expect to see as potential differences from testing in 1D or regional 3D cases? It would be helpful to elaborate here on the motivation to test this EDMF parameterization in a global setup given that it has already been tested in 1D and regional 3D scenarios in previous work? An overview of the additional insights provided by this test in a global configuration would further justify this work as a standalone manuscript.

These aspects are now discussed lines **116-121**.

L94-103: An overview briefly outlining how these four objectives will be achieved in this manuscript would be helpful.

An overview of the paper is added lines **136-140**.

L117: What did the authors mean by “XYZ bulk formula”?

“XYZ” was a typo now corrected to “IFS” and reference is added, line **154**.

L118-119: Was this a concern?

These specifications were asked by former Reviewer #1, hence for potential reproducibility purposes, we now provide the parameterizations used in our experiments that may impact the mixed-layer properties.

L120-121: Please be specific on what the prescribed chlorophyll was used for.

Correction added line **158**.

L130-133: Please be specific on which simulation — I assume it is not that with EDMF? Should it be used with EDMF as well? I'm also confused why MLP was included in the comparison if the goal was to demonstrate the effect of MF in representing convective mixing?

Correction added line **168,** and please refer to our answer to General comment #3.

L145: Why "up to"? Does the large constant value depend on other parameters?

Corrected line **181-182**.

L148-149: It would be helpful to be specific on the description of three experiments rather than using something like "two of our three experiments"

Corrections added lines **168**, **185-186**.

L151: What did the authors mean by "bespoke numerics"?

Correction added line **188**.

L156: Perhaps give the experiment a name rather than referring it as "One of our experiments"?

Correction added line **194**.

L156-158: I don't understand this sentence.

This sentence was removed/modified.

L158-160: Please be more specific. Some references would also be helpful.
Correction added lines **196-197**

L162-163: What did the authors mean by "regime-specific parameterizations or coefficients"?

As requested by Reviewer #2, we now clarify that a single, unique set of coefficients is used for the EDMF parameterization throughout the global ocean. The coefficients are not adjusted or regionalized to account for different convective regimes.

L163-164: But how?

Correction added lines **200-202**.

L171-174: I'm confused about the reasoning on the choice of \alpha=1. If \alpha=0.2 produced the best results in 1D tests, why using a different value here? What was the difference?

Coefficients tuning in 1D cases does not necessarily apply to 3D cases. Here, the use of alpha=0.2 generates plumes that were too energetic, so we use alpha = 1 to prevent any damping from the resistance term, lines **210-212**.

L179-181: This sentence is confusing. Increases/decreases with what?

Correction added lines **218-220**.

L199: Eq (6) differs from Eq (5) not only by the coefficient \beta.

Correction added line **236**.

L209-210: Table 1 can be introduced earlier.
Correction added, lines **168, 186, 194.**

L214-215: Pleases elaborate more on this point. What did the authors mean by "an online re - calibration of the parameterizations"?

We thank the reviewer for pointing out that the meaning of "online re-calibration of the parameterizations" was not sufficiently clear. By this, we refer to the recalibration of the parameters of the mixing scheme within the fully coupled model configuration, i.e. while accounting for its interactions with the other parameterizations and processes active in the model.

"Ad hoc" adjustments of scheme parameters can produce simulations that are realistic in comparison with available in situ observations and ocean reanalyses. However, such adjustments may lead to over-calibration, as they can compensate for errors originating from other parameterizations or model processes. For parameterizations that interact strongly with the rest of the model physics, semi-automated calibration approaches are therefore preferable. For example, High-Tune methodology of Couvreux et al. (2021) explicitly accounts for parameter uncertainty and allows the parameter space to be explored systematically, while reducing the risk of over-calibration with ad hoc parameters adjustments.
The sentence was modified into "However, the full potential of the new parameterization may only become apparent after an online recalibration of its parameters within the complete model configuration (Hourdin et al., 2021, 2023).", lines **253-255**.

L231-324: The description here is very confusing. Please be more specific on how the model data is compared with OSTIA and ARMOR and COREA_REP. For example, on what temporal resolution was the comparison done? What did the authors mean by "the model equivalents of the observations"?

For the SST comparison, the hourly model SST is interpolated in space and time to the location and time of the OSTIA observations. Similarly, for the vertical profile comparison, the hourly model temperature fields are interpolated to the location and time of each ARMOR/CORA-REP in situ observation, and to the corresponding observed depth levels. The resulting model values at the observation locations, times and depths are referred to as "model equivalents of the observations". The model–observation differences are then aggregated within 3° × 3° grid boxes to calculate the bias and RMSD, reducing the sensitivity of the statistics to the heterogeneous sampling of in situ observations.

This model-equivalent approach is commonly used in ocean data assimilation systems to evaluate model fields against observations, including in the GLORYS12 near-real-time operational system (Lellouche et al., 2021). We have added this clarification and the corresponding reference in the revised manuscript, lines **270-277**.

L238-240: Did the authors computed the MLD from individual profiles or simply used the MLD climatology? From the wording here it sounds like the former but I doubt it. It would be more important to clarify on how MLD was computed in NEMO. Was it computed online and output monthly or was it computed from the monthly output of temperature and salinity?

These aspects are clarified line **282**.

L243: Was it resampled or averaged?

Correction added line **285**.

L253-261: The analyses on AMOC and sea ice concentration are only provided in the supplement and not discussed in the main text. So maybe move these data description to the supplement?
Correction added to the SI.

Section 2.3: The organization of the data description in this section can be improved. Also, it would be helpful to briefly discuss the reasoning of using each dataset.

We have revised Section 2.3 (lines **271-277, 295-297**) to make the rationale for the choice of each observational dataset more explicit and to improve the logical organization of the description. We have kept the section concise and organized the datasets according to the model variables and processes they are used to evaluate, rather than introducing multiple subsections for this relatively short section. Two paragraphs that were previously included in this section have also been moved to the SI, as suggested by the reviewer, which further reduces the length of the section.
The choice of the datasets is now clarified as follows. OSTIA provides a global, gridded SST product and is therefore used to evaluate the modeled surface temperature field. ARMOR and CORA-REP provide a large collection of in situ temperature profiles and are used to evaluate the modeled vertical temperature structure. Both products are widely used reference datasets for the evaluation of global ocean model simulations. For MLD, we use the climatology of de Boyer Montégut et al. (2024), which is particularly suited to global model evaluation because it provides a spatially complete gridded climatology derived from profile observations and uses the threshold-density criterion consistently with the MLD diagnostic applied in our model. This dataset has also been recommended for the evaluation of global ocean models, including in OMIP (Tréguier et al., 2023). Finally, the TAO mooring at 2°S, 140°W is used for a more detailed evaluation of shallow diurnal convection.

This mooring was equipped with enhanced vertical-resolution temperature sensors specifically providing detailed observations of the upper-ocean temperature structure, making it particularly relevant for evaluating the vertical mixing and diurnal convection processes investigated in this study (Masich et al. 2021).

Fig 1: It may be helpful to compare with the seasonal activity of EVD (perhaps in the supplement)?

A direct comparison of the seasonal activity shown in Fig. 1 with the EVD simulation is not possible using the same diagnostic. Figure 1 shows the occurrence of the EDMF convective plume, identified from the presence of a plume vertical velocity (w_p). The EVD parameterization does not explicitly represent convective plumes and therefore does not provide a corresponding (w_p) field in the model output. Consequently, an equivalent measure of plume occurrence cannot be derived for the EVD simulation.
The activity of the EVD-based simulation in terms of depths and intensity of its diffusive fluxes in the Labrador Sea is provided in Fig. S1, and comparable to those of EDMF in Fig. 2b.

Section 3.1.1: This section showed results at a mooring site in the Labrador Sea but barely compared the model results with the observation (other than the magnitude of plume velocity). How did the simulated temperature and salinity profiles compare with the observation at the mooring?

We thank the reviewer for this comment. We realize that the wording “mooring site” in the original manuscript may have been misleading and may have suggested that the model output was extracted at the location of an existing observational mooring. This is not the case. The coordinates (56.53°N, 52.65°W) correspond to a fixed model grid point, hereafter referred to as a “numerical mooring”, at which model output is saved at hourly frequency. There is no observational mooring at this location, and therefore no continuous observational temperature and salinity profiles are available at this exact site against which the simulated profiles could be directly compared.
For this reason, a point-by-point comparison of simulated temperature and salinity profiles with observations at this location is not possible. Instead, in Section 3.1.1, we evaluate the simulated convection against published observations of Labrador Sea convection, focusing on characteristics that can be compared across datasets, namely the timing, depth and intensity and vertical velocity of convective plumes. We have clarified this point in the revised manuscript and replaced the potentially misleading wording “mooring site” with “numerical mooring”. We have also explicitly stated that no continuous observational time series directly comparable to the model diagnostics is available at this location, lines **335-341**.

L333: Some comments on why this specific date was chosen would be helpful.

We thank the reviewer for pointing this out. March 15, 2015, was selected because it falls within the core of the most intense convective winter over the 2011–2015 period, when the simulated convection reaches its maximum intensity and depth. We have added this explanation to the revised manuscript, line **381**.

L355-360: I had a hard time following the reasoning behind the claims by the authors that various previous observations were consistent with their findings in the simulation. Fig. 1 only showed a very rough climatological estimate (in terms of orders of magnitude perhaps) on the convective depths.

We agree that Fig. 1 provides a climatological, spatially resolved overview of the occurrence and typical depth range of the mass-flux (MF) component, rather than a direct quantitative validation against observations. Our intention was to provide a first, broad assessment of the simulated convective depth range and to show that the shallow convective regime identified by the model in the tropical ocean (30°S–30°N) is generally consistent with the range of convective depths reported in the observational literature.
Specifically, Fig. 1a-h shows that the MF component is active year-round in the tropical ocean and is predominantly confined to shallow depths (<100 m). This depth range is consistent with previous observations from the TAO and PIRATA mooring arrays, which report convective depths ranging from a few tens of metres up to approximately 90 m (Masich et al., 2021; Moum et al., 2023). We agree that this comparison should not be interpreted as a direct validation, given the climatological and spatially aggregated nature of Fig. 1. We have therefore revised the text to make the nature and limitations of this comparison clearer, lines **404-411**.
A more detailed comparison with observations is subsequently provided in Section 3.1.2, where we focus on a specific TAO mooring and directly compare the modelled temperature evolution with the observations at that location.

Fig 3: Is it necessary to show the results of a full month?

We chose to show a full month (July 2017) deliberately, rather than focusing on a shorter period, because this allows us to demonstrate that the simulated shallow convective plumes recur systematically during nighttime cooling events. It also illustrates the substantial variability in both convective depth and plume velocity from one night to another. We therefore consider the one-month time series useful for characterizing the recurrent nature and variability of the diurnal convection, rather than showing a single representative event. We have clarified this motivation in the revised manuscript, lines **414-415**.

L401-402: This is probably an overstatement. EDMF may be a better way to represent the effects of the physical processes (shallow diurnal convection) in this case than EVD (which I think is what the authors intended to convey). However, the physical processes themselves are not resolved in EDMF either.

Correction added line **455**.

L426-429: The same (or similar) argument can be made against EDMF.

Correction added line **483-484**.

L429-430: This percentage change reported here is extremely misleading given that the vertical structure of bias is largely unchanged between these three experiments and the compensating bias argument above can be made here as well — the small global depth-averaged bias results from the compensation of pronounced cold bias above ~100 m and below ~600 m and warm bias in between.

Corrected added line **484**.

L431-439 and Section 3.2 in general: I'd suggest the authors to be careful interpreting the results here and avoid making overstatements that may not be sufficiently supported by the results. Using the bias of global mean temperature profile as a metric, there are certainly regions where EDMF performs

better, but there are also regions where it performs worse — in Fig. 5c the small overall bias in EDMF seems to be a result of compensation of even larger cold bias below ~300 m and the warm bias between 100 and 200 m. Compared to the remaining biases (especially when looking at the maps such as in Fig 7), the improvement brought by EDMF is really incremental.

Following this suggestion, several modifications were added throughout Section 3.2.

L569-570: It would be helpful to elaborate how this regulation happened.

The regulation of the mixed-layer depth by the MLP parameterization is explained in more detail in lines **624-627**. The MLP parameterization suppresses residual instabilities within the mixed layer, thereby homogenizing the mixed layer and reducing the occurrence of conditions triggering the EVD scheme. As a result, the frequency of EVD activation is reduced, which limits further deepening of the mixed layer and leads to a shallower mixed-layer depth. We have clarified this mechanism in the revised manuscript.

L583-585: The reasoning of not including a comparison with observations is confusing. Does the same issue also exist when compared with de Boyer Montégut (2024) dataset?

The issue described here applies to the comparison with individual Argo-based MLD estimates, but not to the de Boyer Montégut et al. (2024) dataset. The latter provides a global gridded climatology of MLD derived from profile observations, with spatially complete fields that can be directly compared with gridded model output. We therefore use the de Boyer Montégut et al. (2024) product for the model–observation comparison presented in this study.
We mentioned in the Methods section that this profile-based MLD climatology is particularly suited to the evaluation of global ocean models and has been recommended for this purpose, including in the Ocean Model Intercomparison Project (OMIP; Tréguier et al., 2023).

Section 4: The presentation here is very confusing and I had a hard time following the discussion in this section.

In essence, this section aims to estimate the adjustment of mean ocean temperature and surface fluxes to EDMF. This adjustment is quantified in terms of temperature and surface flux anomalies relative to EVD. Please refer to our answer to General comment #2.

Eq (8): I don't understand this equation. What are e_1, e_2 and t_{mask}? Also, what is the sum over i? What is F(0)?

We agree that the notation in Eq. (8) was not sufficiently clear. The terms (e_1) and (e_2) are the horizontal grid-cell dimensions and are used to calculate the area of each grid cell. The term t_{mask} is the mask used to select the ocean grid cells belonging to the region considered in the analysis. The index (i) denotes each daily time step within the three-month climatological seasons considered (JFM or JAS), as introduced at the beginning of the paragraph.
We have also removed the second expression involving (F(0)), since it is not used in the subsequent analysis and was unnecessary. The equation and its description have been revised to make the notation and calculation more explicit, lines **669-670**.

Eq (10): Please define the symbols with an arrow

Correction added line **680**.

L627: What did the authors mean by “scaling factors”?

Correction added line **690**.

L731-741: Could these assessments in the shallow diurnal convection and deep wintertime convection regimes done using 1D tests? What are the additional insights brought about by using a 3D global simulation?

We agree that idealized 1D experiments could be used to investigate the mechanisms underlying both shallow diurnal and deep wintertime convection, and would provide a useful complementary assessment of the EDMF parameterization. However, the objective of the present study was to evaluate the behavior of the parameterization in a realistic oceanic environment over a long time period.
The main additional insight provided by the 3D global simulation is that it allows us to assess whether the same formulation can reproduce the spatial and temporal variability of distinct convective regimes under realistic oceanic conditions. In particular, the 3D configuration accounts for horizontal advection and its interactions with the vertical mixing processes, which are neglected in 1D experiments. It also allows the parameterization to be evaluated over a wide range of regions and forcing conditions within a single consistent framework and over a 16-year period. This is now added to lines **799-801**.

L766-767: I think the effects of MLP was not sufficiently discussed in this manuscript and I’m still confused why it is included in this comparison.

We agree that the role and effects of the MLP parameterization should be clarified. We have therefore expanded the discussion of the MLP parameterization and its limitations in lines **85–101**, and added further discussion of its effects in lines **559–560, 564–56**5, and **614–618**. We also now explicitly state in the objectives (lines **129–130** and **138–139**) that EDMF is compared with the TKE+MLP+EVD simulation, which corresponds to the default NEMO configuration.
However, a detailed assessment of the MLP parameterization itself is not the main objective of this study. Its effects and limitations have already been discussed extensively in previous studies, whereas the primary aim here is to evaluate the performance of the EDMF parameterization. We therefore use the TKE+MLP+EVD configuration primarily as a relevant reference representing the standard NEMO configuration, rather than as a focus of a separate assessment of MLP.

L772-773: Does GEOMETRIC represent submesoscale restratification? Could the authors provide some references?

Correction added lines **842-844.**