July 1, 2024

Sofía Allende
ELIC Place Louis Pasteur 3/L4.03.08
1348 Louvain-la-Neuve Belgique
Phone: +(33) 641842014
Email: sofia.allende@uclouvain.be

This studies tests a general modification to turbulence parameterization that redistributes turbulent kinetic energy across the base of the mixed layer to balance existing underestimations of the mixed layer depth, with respect to its performance on surface ocean and sea ice properties in the Arctic by evaluating multi-year runs with different parameter settings. Turbulence parameterizations are an important and critically underconstrained factor in ocean and climate models, and their improvement is vital to reduce uncertainties in climate predictions. The methods are established and sound, and the paper will be a good and important contribution to the improvement of ocean models.

Prior to publication, I believe the following major concerns would need to be addressed:

We thank the Referee for his/her careful report. We have followed all of his/her recommendation and we believe that this has significantly improved the quality of the manuscript.

Hereafter, we give a short list of the implemented changes in view of the Referee's comments. Please find below a point-by-point reply.

(1) Throughout the manuscript, the authors use both the assigned variables (chi, f_r, h_T) as well as the NEMO-internal identifiers (rn_efr, nn_htau etc) to refer to the paramters. This makes the manuscript very hard to follow, and gives the impression of a NEMO-specific technical report rather than a scientific paper. I strongly recommend removing all NEMO-internal identifiers and only use the actual variables. (If needed, a table could be added in the Appendix to relate to the NEMO identifiers, but I assume these can be found in a NEMO-specific documentation. I would also try to reduce the references to NEMO specific models to a minimum, and mention other model frameworks that use TKE MLP to make the study more appealing for readers outside the NEMO community.
We thank the Referee for pointing out the lack of clarity in our parameter references. We have revised the manuscript for consistency by removing all NEMO-internal identifiers and using only the actual variables: χ, f_r and h_τ.

(2) In the results section, surface ocean properties are discussed individually, but these values are directly linked in a very straightforward and known way (more mixing = redistribution of salt = deeper ML = higher surface salinity = lower stratification). This separation leads to repetition and makes it hard to follow the paper. I recommend re-arranging both results sections to present findings on the individual diagnostics (MLD, surface salinity, stratification) side by side.
We have reorganized this section to enhance clarity and flow. We first introduce the mixed layer analysis, followed by the discrepancies in ocean stratification, salinity, and temperature vertical profiles, and end with the results on sea surface salinity and temperature.

(3) It would be great to have a table summarizing which configuration performs good/intermediate/bad for specific regions (the three basins) and processes (seasonal cycle, interannual trends, MLD; surface salinity, sea ice).
We have included a table in the discussion and conclusion section that summarizes the performance of

the MLD and sea ice thickness seasonal cycle and inter-annual variability for all regions studied: the Makarov, Eurasian, and Canada Basins.

(4) The overwhelming part of the discussion and conclusion section is a repetition of the results. The summary should be massively shortened, to one paragraph, and the discussion should address questions like: Why do the different configurations yield the observed difference in simulated conditions? What is your recommendation for the 'best' configuration of the TKE MLP in Arctic Environments? How much better does your identified best choice perform against the state-of-the-art configuration? Is it generally applicable, or do users need to discriminate which configuration to use based on regions? What are the limitations of the parameterization, ie, where does it not perform well? What data is needed to make further improvements? Is this parameterization the way to go ahead, or are there alternatives which might be superior in the long run?
We have revised this section to clarify our message. The table mentioned above has been included to highlight which simulations obtain better results in each region. We agree with the referee that some recommendations should improve the manuscript. Accordingly, we have incorporated these suggestions and discussed the limitations of our results.

Minor concerns:

- The description of the implemented underlying turbulence parameterization (not MLP) used in the methods section (l. 60-71) is very brief. For a manuscript focussing on improving turbulence parameterization, I would appreciate a more detailed description without having the reader go back to the original literature.
We have included the prognostic equation that describes the default TKE formulation in NEMO (lines 77-80):

The prognostic equation is given by:

$$\frac{\partial \bar{e}}{\partial t} = K_m \left(\frac{\partial \bar{U}_h}{\partial z} \right)^2 - K_\rho N^2 + \frac{\partial}{\partial z} \left(K_e \frac{\partial \bar{e}}{\partial z} \right) - \varepsilon$$

It results from the balance between the vertical shear, the dissipation of TKE due to buoyancy, the vertical diffusion of TKE, and the energy dissipation.

- l. 95 I do not understand what "excluding salinity restoration under sea ice" means, please add a brief explanation.
In the NEMO model, there is an option for surface restoring to observed Sea Surface Temperature (SST) and/or Sea Surface Salinity (SSS). The SSS restoring term serves as a flux correction on freshwater fluxes to mitigate uncertainties in the observed freshwater budget. In our simulation, this option is deactivated. We have included the following explanation in line 109: "In our simulations, we configure the setup to exclude salinity restoring under sea ice. This means that no flux correction on freshwater fluxes was applied."

- l. 96 Rathore et al. - year is missing
We added the year to the reference.

- Table 1 is not necessary, is / can be incorporated in the text
We have modified the table to include the values of each parameter. We believe it is useful for the reader to have this information summarized.

- Ocean and Sea Ice variables around l. 115: No need to give the NEMO-internal identifiers, stick to the commonly used symbols.
We have removed the NEMO-internal identifiers in the new text (lines 128-132).

- l. 125. The density threshold criterion is repeated in two different sentences
We have rewritten this part, and the information is now presented in a clear and concise manner. Here is the revised text with the correction (lines 139-141): "The MLD is calculated for each individual profile by employing the threshold density criterion. This criterion is based on the difference in density between a given depth (z) and a reference depth ($= z_{ref}$), which is denoted as $\Delta\rho = \rho(z) - \rho(z_{ref})$. The MLD

is then defined as the depth where this difference in density exceeds the threshold value of 0.03 kg/m^3 (e.g. de Boyer et al. (2004))."

- l. 127 "The choice of the reference depth is impactful" - please provide details on why it is, and also I do not understand why you pick different reference depths for model / obs later on, what was the reason for that (especially as you state the choice does NOT make a difference, around l. 195)

We have added an explanation to clarify this issue. In the study by Treguier et al. (2023), it was shown that changing the reference depth of the model CMCC-NEMO from 0.5 m to 10 m can lead to differences in the MLD exceeding 40 m across large areas of the Southern Hemisphere oceans. In the Arctic region, these differences are smaller, reaching only a few meters, as we have also demonstrated using our results and comparing a reference depth of 0.5 m to 5 m. Specifically, during the summer months of July and August, these differences are less than 3 meters. While these variations are significant given the shallow MLD values during these months, they alone cannot account for the discrepancies observed with the LOPS climatology for the same periods.

The new version is (lines 204-207): "As shown by Treguier et al. (2023), the reference depth significantly impacts the MLD. For instance, altering the reference depth in the CMCC-NEMO model from 0.5 m to 10 m leads to differences in MLD exceeding 40 meters across large areas of the Southern Hemisphere oceans. In the Arctic region, these differences reach a few meters."

- if nighttime convection is not an important issue in the Arctic, shorten the corresponding paragraph

We have shortened the paragraph to (lines 142-147): "The LOPS climatology estimates the MLD as the depth mixed over at least one daily cycle, typically assumed to be no less than 10 m deep (e.g., Brainerd et al. (1995)). This depth filters out possible daily stratification in the top few meters, which is common in the tropics or summer mid-latitudes. However, in the Arctic Ocean, where the diurnal cycle linked to solar heat fluxes is minimal or nonexistent, especially with ice present, the MLD can be shallower than the usual 10 m depth (e.g., Peralta et al .(2015)). Therefore, we have recomputed the MLD for the Arctic region using a more appropriate reference depth of 5 meters."

- Table 2. Needs improvement, I do not understand it.

This table has been removed due to changes in the reference of the simulations. We no longer use the internal NEMO names for the variables.

- l. 137. A citation is needed to the TEOS-10 part, no need to mention a python toolbox was used. Also, consistently using TEOS-10 would make it consequent to use absolute salinity and not practical salinity (which is unitless, not pss) throughout the manuscript. The TEOS-10 framework also includes the Brunt-Väisäila frequency N, which should be introduced here instead of around l. 240 in the results, with a shorter description.

We added a citation for TEOS-10. We agree that our previous text was confusing and not clear about why we talked about the conservative temperature and absolute salinity profiles. Initially, we intended to indicate that vertical potential density profiles were computed using the toolbox that uses such quantities. However, this is unnecessary and adds confusion, especially since our results are based on practical salinity and potential temperature. Therefore, we removed this part, and the revised text is (lines 149-151): "We computed vertical potential density profiles using the TEOS-10 Gibbs Sea Water toolbox (Mcdougall et al (2011), and determined the MLD by applying the threshold density criteria. This was done to compare ITP observational data with NEMO sensitivity experiments, where the surface reference depth for ITP varies from 10 to 0 m depending on the profile."

- l. 157-162: you introduce upper ocean and sea ice variability by describing the seasonal cycle, but there are also huge regional differences in the Arctic Ocean that should be summarized here.

The description of the seasonal cycle has been moved to the introduction. We have incorporated regional differences as follows (lines 54-60): "The Makarov Basin, located north of Siberia, experiences seasonal ice cover and receives freshwater from the East Siberian Sea. Shallow depths (500-1500 m) render it highly sensitive to sea ice variability and freshwater inputs. The Eurasian Basin extends from the Siberian Shelf to the North Pole, featuring extensive multi-year ice cover and significant freshwater discharge from major Arctic rivers. Depths can reach 4000 m, influencing Arctic freshwater storage and sea ice dynamics. The Canada Basin, situated between North America and Siberia, is dominated by multi-year ice influenced by the Beaufort Gyre. Its complex bathymetry, including deep ridges like Alpha and Mendeleev, affects ocean circulation patterns and carbon cycling."

- l. 169: "MLD discrepancies are less pronounced" compared to other NEMO simulations, or any other simulations in general?
We have clarified the text as follows (lines 175-177): "Compared to a large portion of global models forced by CORE-II and JRA55-do, as studied by Ilicak et al. (2016) and Allende et al. (2023), which include both NEMO and non-NEMO models, MLD discrepancies with observational data are less pronounced."

- l. 173: "This would be one point to be improved in a future version of this data set" belongs into the discussion.
We have significantly revised the discussion according to the recommendations of the referee. We have however kept this specific sentence in this part of the text to emphasize the difference between the control run and the LOPS climatology.

- l. 195 (and around): Confusing, why not just use the same reference density? Also, move to methods.
This information is already detailed in the methods section. We utilize the MLD output directly from NEMO, which is consistent with the CMIP6 protocol. This approach enables straightforward comparisons with other models and ensures higher accuracy compared to monthly density profile recalculations. Here, we examine the implications of this approach, particularly regarding the shallow MLD values observed in the summer months. We recomputed the MLD using the same reference depth as the LOPS climatology and found discrepancies of up to 3 meters between the two methods. However, these differences are insufficient to fully account for the discrepancies observed with the LOPS climatology.

- I am not sure if I understand the meaning of the standard deviation comparison: Is that a measure for the internal variability in the three basins? Please expand.
Your understanding is correct. The standard deviation associated to the depth of the mixed layer quantifies its spatial variability in each basin. We have added the sentence "To quantify the spatial variability of the mixed layer within each basin, we measure the MLD standard deviation for each month"

- l. 214-216: Not only sea ice melting affects salinity anomaly, also advection of (modified) river water!
We have modified and moved the text to improve clarity. The new text reads as follows (lines 262-269): "The discrepancies between the sensitivity experiments and the control run for the spatial distribution of sea surface salinity and sea surface temperature exhibit a similar pattern (see Fig. 8). A decrease in the MLD with a strong stratification corresponds to a reduction in sea surface salinity and an increase in surface temperature compared to the control simulation. In contrast, an increase in MLD with a weak stratification is associated with an increase in sea surface salinity and a decrease in surface temperature. This can be attributed to the fact that a shallower mixed layer and a strong stratification result in less mixing during ice melt, leading to a fresh anomaly at the surface and trapping heat at the surface. On the other hand, a deeper mixed layer and a weak stratification allow freshwater to mix more deeply, resulting in higher surface salinity and facilitating vertical heat exchange."

- l. 219-230: This comparison to other model runs feels out of place here, but could be modified and go into the discussion. Also, Atlantic Water temperature at 200m depth should be relatively unaffected by modifications of the surface(!) mixing parameterization, right?
Related to the Atlantic Water temperature at 200m depth, we have added the following clarification (lines 247-249): "While our control simulation demonstrates improvements compared to these models, adjusting the TKE MLP parameters does not improve the representation of the temperature maximum below 200 m, as this maximum is primarily affected by heat advection at that depth."

- Figures 7, 8: specify which temperature is shown (potential, conservative)?
We have added to Figures 7 and 8 that we are using potential temperature, as mentioned in the methods section. This should help clarify any confusion.

- l. 239ff. Higher turbulence reduces stratification is quite basic - I am not sure if this paragraph adds to the story of the paper. Consider to remove?
We apologize for any confusion, but line 239 does not contain a reference to "Higher turbulence reduces stratification." Therefore, we are unable to evaluate the suggestion to remove that paragraph. If there is a specific paragraph or section that the reviewer is referring to, please provide additional context or a specific line number for clarification.

l. 251- l. 257. I cannot follow the FWC part: Mixing only redistributes salt / FW, so is the difference from additional ice melt? The numbers seem high for that. Also, for FWC results, a discussion of the sensitivity to choice of reference salinity is needed.
We agree that the calculation of FWC is somewhat out of context in the paper. Therefore, we have decided to remove it. However, here is the explanation of the calculation:

We calculate the upper-ocean seasonal freshwater content (sFWC) by:

$$sFWC(t) = \int_{Z_{fw}(t)}^{0} \frac{S_{may} - S(t,z)}{S_{may}} dz,$$

where S is the salinity at a month t, and $Z_{fw}(t)$ is the vertical extent of mixing approximated here by the MLD of May. Therefore, $S(t,z)$ is the salinity of the month z at its ML deepening. We use the May-average conditions as a reference salinity value following Rosenblum et al. (2021), different from the commonly used 34.8 ppt, to avoid errors stemming from arbitrary values and ensure a fair comparison between experiments Schauer et al. (2019).

l. 260: Add HOW sea surface properties affect sea ice.
We have included how sea surface properties can modify sea ice as follows (lines 272-274): 'For instance, higher sea surface salinity lowers the freezing point of seawater, delaying the formation of sea ice. Conversely, lower sea surface salinity raises the freezing point, promoting sea ice formation. Additionally, colder sea surface temperatures encourage sea ice formation, while warmer sea surface temperatures contribute to sea ice melt."

l. 266 what is meant by "East coast"
We have clarified this part of the text by adding (lines 280-281): "Specifically, regions near the East coast (Chukchi, East Siberian, Laptev, Kara, and Barents Seas) display sea ice thickness close to zero during this month."

l. 292 (and again later): remove "leads to a large Richardson number, which" - the Richardson number has never been defined here, relation stratification and Ri is trivial, so it adds no information here.
We agree with the Referee. The two references to the Richardson number have been removed.

l. 308. "unrealistic seasonal cycle" - unrealistic in which way? remove "displaying a nearly flat linear regression", that is the same as 'no trend'
We have incorporated the suggested changes to the text as requested. The revised text now reads (lines 321-322): "Consistent with Fig. 5, all experiments simulate a ML shallower than observations, except for the $\chi = 1$ experiment with strong TKE MLP under sea ice, which exhibits a too depth seasonal cycle."

l. 363. "beneficial in the NEMO model" - in which way and where? Open water?
We clarified this part as follows (lines 367-370): "As noted by Calvert et al. (2013), Rodgers et al. (2014) and Storkey et al. (2018), the additional source of mixing by TKE MLP is beneficial in the NEMO model to reach realistic MLD in the Southern Ocean and in open water regions, which we demonstrate holds true for the Arctic region as well (see Table. 2)."

l. 386 "biases are not due to vertical mixing" - what could be likely other reasons for these biases then?
We clarified this part as follows (lines 405-408): "The underestimation of sea ice thickness and sea ice concentration in the control run remains in all sensitivity experiments, which show that these biases are not due to vertical mixing only. One potential reason for this could be the ERA5 forcing, which introduce warmer temperatures in the Arctic. As shown by Batrak et al. (2019), ERA5 has a warm bias in winter, leading to thinner ice and a reduced summer extent in the model. Further investigation is needed to explore this aspect.