Review of “Impact of ocean vertical mixing parameterization on Arctic sea ice and upper ocean properties using the NEMO-SI3 model” by Allende et al., (gmd-2024-49)

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:

(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

(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.

(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).

(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?

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.
- l. 95 I do not understand what “excluding salinity restoration under sea ice” means, please add a brief explanation.
- l. 96 Rathore et al. - year is missing
- Table 1 is not necessary, is / can be incorporated in the text
- Ocean and Sea Ice variables around l. 115: No need to give the NEMO-internal identifiers, stick to the commonly used symbols.
- l. 125. The density threshold criterion is repeated in two different sentences
- 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)
- if nighttime convection is not an important issue in the Arctic, shorten the corresponding paragraph
- Table 2. Needs improvement, I do not understand it.
- 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.
- 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.
- l. 169: “MLD discrepancies are less pronounced” compared to other NEMO simulations, or any other simulations in general?
- l. 173: “This would be one point to be improved in a future version of this data set” belongs into the discussion.
- l. 195 (and around): Confusing, why not just use the same reference density? Also, move to methods.
- 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.
- l. 214-216: Not only sea ice melting affects salinity anomaly, also advection of (modified) river water!
- 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?
- Figures 7, 8: specify which temperature is shown (potential, conservative)?
- 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?

- 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.
- l. 260: Add HOW sea surface properties affect sea ice.
- l. 266 what is meant by “East coast”
- 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.
- l. 308. “unrealistic seasonal cycle” - unrealistic in which way?
- remove “displaying a nearly flat linear regression”, that is the same as ‘no trend’
- l. 363. “beneficial in the NEMO model” - in which way and where? Open water?
- l. 386 “biases are not due to vertical mixing” - what could be likely other reasons for these biases then?