<u>REVIEW #2</u>

" Assessment of the Finite VolumE Sea Ice Ocean Model (FESOM2.0), Part II: Partial bottom cells, embedded sea ice and vertical mixing library CVMIX" by Scholz et al., 2021.

The manuscript of Scholz et al documents a large number of sensitivity tests related to new features for FESOM2 (partial bottom cells, sea ice coupling, and vertical mixing). I was extremely impressed by the vast number of simulated years done for this manuscript and the quality of FESOM developments and the writing. While I found the paper sound and clear for the most part, I have a couple major concerns and more detailed minor comments that I would ask the authors to consider.

We thank the reviewer for his efforts and constructive comments. We tried to thoroughly consider all of his comments or answer his concerns.

First and foremost, I struggled in the manuscript to see the scientific impact to the broader community. This manuscript reads as a FESOM technical report and is no doubt very beneficial to that community.

The scope of the paper is to first provide technical insights to the broader FESOM community about the ongoing FESOM developments but also of course to the broader and general modelling community.

However, it was very difficult for me to draw out points of interest to the broader modeling community. A few examples, I felt the discussion and results around partial bottom cells (PBCs) and sea ice coupling did not come across as of interest to the broader community, but it is quite possible that having all these results in one place will be a useful reference. You could consider trying to address causes of biases more directly and what the role of the change was in the circulation change, or perhaps consider making a recommendation of best practice configuration for FESOM at the very least. These could be ways to improve the broader impact of this manuscript.

We agree here with the reviewer that it would be of benefit for the manuscript to give a recommendation for a best practice configuration based on the presented model options in the discussion and conclusion section.

As of now, for the most part biases are simply noted and then moved on. There are a few exceptions, it is mentioned the way FESOM computes the bulk Richardson number causes the changes in some biases, but plots of boundary layer depth or surface layer average velocity and buoyancy for each method were not plotted.

Going into a deep analysis of some of the biases, might exceed the scope and length of this publication, which already has quite some length. This is also one of the reasons why things like buoyancy, boundary layer depth etc. haven't been shown. However we can try to present some of these quantities

within the supplementary material. Also, the cause of some of the biases regarding partial cells could not fully be clarified in this setup and ask for an own examination in a more simplified or idealized configuration.

Second, I think there were issues with the discussion of vertical mixing with regards to CVMix. It is stated that FESOM_kpp is configured in a different manner from CVMix, with the surface layer averaging noted as a key difference. However, this is not correct. CVMix leaves a number of choices up to the calling model, amongst them is the velocity and buoyancy difference for the bulk Richardson number. Griffies et al 2015 recommends using 10% of the boundary layer depth as the surface layer average, but this is not within CVMix itself. We further discuss the dependence on model choices in Van Roekel et al 2018.

We agree with the concerns of the reviewer. We synchronized our implementation with our project partner models MPIOM and ICON-o and they used MOM6 as a template to implement CVMIX KPP. We will clarify this issue in the manuscript.

As an example, POP chooses the largest shear between a depth and the surface cell making this one step further than what FESOM chooses. It would be interesting if FESOM made other choices in configuration relative to default CVMix, e.g. Monin obukhov/Ekman limiters, matching at the boundary layer base, shape function parameters, etc… and what the impact of these choices might be.

In the implementation phase of CVMIX KPP we played with the options for the Monin-Obukhov and Ekman limiters since in fesom_KPP they were activated as a default. But both Monin-Obukhov and Ekman limiters had only a minor impact on the solution.

Relatedly, I think it is important to note that the FESOM KPP choice of the first layer being the surface layer is not physically consistent with KPP (even Large et al 94). Throughout KPP there are built in assumptions regarding the depth of the surface layer (default to 10% of the boundary layer depth) and assuming the first layer as the surface layer is inconsistent.

The original fesom_KPP implementation was inspired by the KPP implementation in MOM4.

While it is a valid assumption it is important to point out this issue and the consequences of it. It would be interesting to explore the impact of this choice, but that is likely beyond the scope of this paper. Here it shows basic plots of T/S, but would be good to know a finer scale view too. Have you conducted a simulation that uses the CVMix library but the FESOM_KPP choice for the numerator and denominator of the bulk Richardson number? This could clearly show differences associated with the choice. My expectation is that the shear is the more dominant term and using the surface value will deepen boundary layers, but it would be interesting to see clear evidence of this. Again, this is a possible place where you could make broader impacts.

To help clarify and grasp the broad points of the paper, I would suggest perhaps having bullet point take summary somewhere (perhaps the discussion/conclusions) with call outs to key figures. A few other suggestions to help with clarity: (1) a table of differences in KPP vs PP and in CVMix / FESOM versions would help maintain clarity (2) a table of the FESOM KPP configuration, e.g. is matching utilized?

Minor comments:

Throughout, you write CVMIX, the acronym is CVMix.

In numerous places you have things like “southern hemisphere September” which reads odd to me. I’d suggest parentheses around the month.

We will change this in the manuscript.

There are also a number of references that need proofing, e.g. Griffis 2015 and Ilicack2006.

We will change this in the manuscript.

There are also many places with subscripts that didn’t typeset correctly.

Originally we tried to avoid subscripts, since they become rather small in the figures. We also tried to avoid the use of hyphens for the description of our experiments since they could be miss interpreted as minus signs, therefore we decided to use an underscore for the description of our experiments (e.g fesom_KPP, cvmix_TKE ...). If this typeset is not wanted, it could be changed.

Line specific Comments:

Line 27: – Delete “The” → ***will be corrected.***

Line 29: – suggest adding “southern hemisphere” to “sea ice melt season mixing” → ***will be corrected.***

Line 40 and L41: remove “one” after “first” and “second” → ***will be corrected.***

Line 50-57: – by the word “embedded” I expected the sea ice code to be in the ocean code as in MOM6/SIS, is this the case for FESOM?

The fact that the sea ice code is within the ocean code is anyway the case for FESOM2, but the term embedded here refers to the publication of Campin et al. 2008 whos came up with the naming and describes the case where the sea ice

is embedded in the surface ocean and swims according to its density in the ocean by replacing water instead of only levitating on top of the ocean

It is not clear. It may be better to say “non-levitating” or “pressure exerting” for clarity, but no need to change the word if clearly defined if the ice code is in the ocean model or uses a coupler.

We do not use any coupler for the sea ice

Line 57: – you mention that you must compute sea ice at every ocean time step to “embed”, this doesn’t seem desirable and is not actually required in our experience in MPAS/E3SM. Embedding is more dependent on the fidelity of the ice-ocean coupling in our experience. Can you clarify what you mean by this statement?

It's no where written that we “must” compute sea ice at every timestep its only written that we rely on zstar for embedded sea ice, which is the case. Otherwise FESOM2 calls in the moment the sea ice routine at each ocean timestep. If this is really necessary or it’s sufficient to be called at every second or tenth time step will be evaluated in the future to maybe further close critical bottlenecks.

Line 81: – It isn’t clear to me what “prime vertical mixing” means, is this default?

For us “prime vertical mixing” schemes refers to schemes that set the general global ocean wide vertical diffusivity like PP, KPP or TKE. “non prime vertical mixing” focus for us on certain processes that are added to the prime vertical mixing, like the breaking of tidal induced internal waves (Simmons et al. 2004, IDEMIX) or local mixing processes like in the Monin-Obukov mixing (MOMIX).

Line 82: – the phrase “deliver a usable mixing scheme” is confusing to me. Do you have a meaning in mind?

“Deliver” stands here in the meaning of “to create”, “ to build” a mixing parameterisation that has certain validity for the entire global ocean. Exchange: ...others that have the purpose to <u>create a general</u> ~~deliver a usable~~ mixing parameterisation for the ...

Line 129: – delete the comma and which → ***will be corrected.***

Line 145: – It is unclear to me why Pacanowski and Gnanadesikan 98 is discussed. It seems FESOM uses Schepetekin 2003 instead.

Pacanowski and Gnanadesikan, 1998 discusses first the basic concept of partial cells, how to compute and how to use them. That's why it's worth

mentioning them. Schepetekin 2003 provides an alternative way on how to compute the pressure gradient force that was more beneficial for us.

__Line 151-153:__ – Have you tested FESOM without the requirement that the bottom thickness be greater than ½ the layer thickness? As an example, MPAS-O runs stably without this requirement, but I have not looked in depth at the possible biases that may exist even though it is stable. It could be interesting to further examine this choice.

We tested mostly for the condition that the layer thickness can not be smaller than half the layer thickness which turned out to be important especially in the shallow shelf areas to keep the time step in limit and to avoid critical vertical CFL conditions. Anyway MOM6 is using this fully lagrangian layer motion with the subsequent remapping, where critical CFL conditions are not an issue anymore. In FESOM2 we will explore this option in the near future.

__Line 157:__ – vertice -> vertex → ***will be corrected.***

__Line 165:__ – biasin -> bias in → ***will be corrected.***

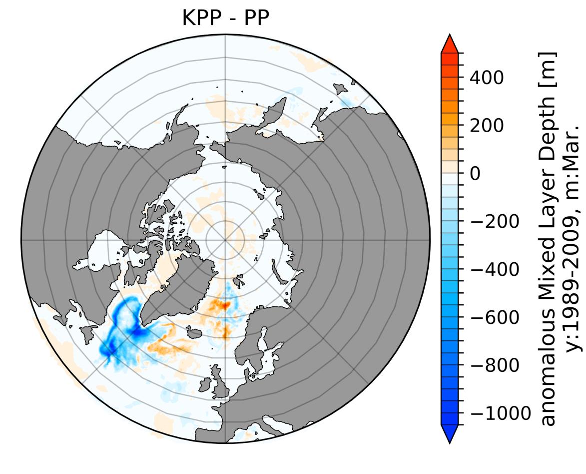
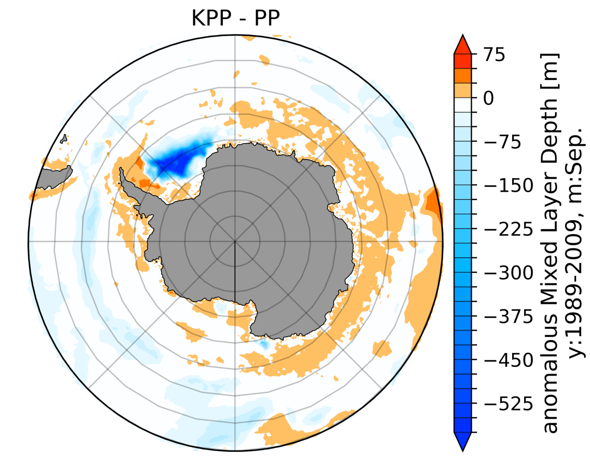
Section 3.2 – In this section you discuss a dependence on sea ice thickness in configuration choices but then only present comparison to sea ice concentration. It would be helpful to plot thickness. This is also more consistent with what is actually measured by satellite.

The results between sea ice concentration and thickness are very similar and there exist more observational derived sea ice concentration estimates (e.g. NSIDC) than thickness estimates, especially with respect to long term climatology. Nevertheless, we can add plots for the sea ice thickness anomaly.

__Line 281:__ – while MOM6 does have a branch with CVMix the original implementation was designed to reproduce the POP formulation, so I would change MOM6 to POP → ***will be corrected.***

__Line 315+:__ - the strong similarity between KPP and PP in the analysis was surprising to me. These schemes are quite different, especially in the near surface. Have you examined boundary layer differences? E.g. Mixed layer depth? I wonder if perhaps the similarities are due to the fields presented, you show averages over fairly thick layers and below the boundary layer I imagine FESOM uses the shear instability induced mixing of Large et al. 1994, it is possible your analysis is only highlighting more deep ocean impacts and the similarities in the LMD SI induced mixing and PP81 mixing make the results seem similar. A simple test would be comparison of MLD between the schemes.

There is a considerable difference between KPP and PP, although the climatological biases with respect to WOA18 is still much larger see Suppl. 2. These differences between KPP and PP can be also seen in the Mixed Layer Depth.

	
March MLD fesom_KPP minus fesom_PP	***September MLD fesom_KPP minus fesom_PP***

Line 398: – ando -> and→ ***will be corrected.***

Line 450-453: – Why not test different background options? Seems like a very easy test to do.

Line 486-488: – this sentence is very confusing to me. When you use ‘except’ but discuss freshening in one part and temperature in the other it doesn’t read easy to me.

There is a typo: it's not supposed to be “warming”. Exchange: “...The depth ranges below indicate a predominant general freshening almost everywhere, except for the Mediterranean outflow and Indian Ocean which indicate a slight salinification~~warming~~....”

Line 527-529: – have you tested combinations of changes? It seems possible (perhaps likely) that some changes have nonlinear interactions and are not as simple as just adding biases.

We did not test this parameter combination in particular. We will redo these experiments to give an adequate answer to this question.

Line 549-552: – any ideas why you see a large change in the gulf stream for the MOMIX + KPP? Is this related to changes in transport (AMOC maybe?)

Fig. 22 shows that with MOMIX the upper and lower AMOC cells become weaker. The weakening of the upper cell leads to a weaker meridional heat transport through Gulf Stream and NAC and could lead to the displayed cooling in the North Atlantic with MOMIX. We will add this explanation to the manuscript.

__Line 582-583:__ – As an MPAS-O developer I confess I agree with your statement here, I'm always deeply impressed by the pace and quality of FESOM developments.

We see MPAS as a rather close competitor that catches up very quickly.

__Line 701-702:__ – add commas around "and to a decrease in the high-latitude" → ***will be corrected.***

__Fig. 14 and 18:__ – the plot titles seem wrong in most panels here. Also in panel (c) of both there is an odd high salinity bias 40N. It is interesting that it is identical in both Fig 14 and 18. Is this a plotting or analysis artifact?

These are not the plot titles, these are the description labels of the colorbar, it might be beneficial to insert here a vertical gap to emphasize this. The high salinity bias at 40° is indeed a bug in the computation of the anomaly. These figures will be corrected and replaced in the manuscript.