

Kiel, December 5th, 2024

Subject: Resubmission of Manuscript: Mixing, Spatial Resolution, and Argon Saturation in a Suite of Coupled General Ocean Circulation Biogeochemical Model Configurations off Mauretania

Dear Editor,

We are pleased to resubmit our revised manuscript, incorporating substantial improvements in response to the very constructive feedback provided by the reviewers. Among the major updates to the manuscript following the reviewers' guidance are a more detailed presentation of the temporal evolution of argon saturation and an analysis of potentially spurious effects of our boundary conditions. The respective updates comprise additional figures and model simulations/analysis. Also, we introduced a new discussion part which discusses the precision of our method and puts our method now better into perspective to other approaches.

Please find the detailed point-by-point response to the reviewers' feedback below. The reviewer comments are marked in bold while our comments are in italics.

Thank you for your time and effort!

Kind regards,
the authors

Point by point response

----- Reviewer #1 -----

We are grateful to reviewer #1 who has obviously spent a lot of time with our manuscript, put the finger on shortcomings of our original presentation and made a lot of very constructive propositions as to how to become more convincing.

R: General comments:

The authors of this manuscript examine the model-simulated inaccuracies in diapycnal mixing that cannot be explained by the horizontal resolution near Mauretania and find that it's comparable to advection numerics and choices of background diffusivity by making use of argon saturation as a proxy for effective mixing. I can appreciate how much time and effort that went into this work but I have some reservations about publishing this manuscript as is. One concern that I have is that there is no comparison of the effective diapycnal mixing with other methods, like that of Holmes et al. (2021) or potentially separating out the spurious contribution to compare with a method like that developed in Ilıcak (2016)'s "Quantifying spatial distribution of spurious mixing in ocean

models". I understand the authors just want to have a relative measure to rank their model configurations with their effective mixing against one another but without knowing whether their method is properly quantifying the effective mixing, the ranking could be misleading. For example, the authors include the mixed layer in their calculations of the effective mixing, despite how there are known confounding effects on the effective mixing. The second issue is that the authors try to connect Ekman pumping/suction to effective diapycnal mixing without explaining the relationship between the diapycnal advection and diapycnal mixing in the Ekman layer. The third issue is that the model spin-up time may suffice for many applications of model experiments, but they are investigating a signal that is so small that the rate at which their model diagnostics are changing at the time they are evaluating them may be relevant to the significance of the small differences they're finding. I suggest major revisions. Specific comments are listed below:

A: Basically, the reviewer asks for a more convincing presentation of the capabilities of our new method along with a more comprehensive embedment of our work in that of others. This makes sense to us.

More specifically we find the reviewer's "issues 1 and 3" are essentially rooted in the same shortcoming of our original manuscript. As we understand, the perception of the reviewer is the following: (1) we present a new method to rank the effect of vertical diffusivity/diapycnal mixing, (2) then we find that our method is affected antagonistically by diffusivity, (3) we run the model/method only for a short period that appears random in length to the reader; such that the reader cannot rule out that the effects we discuss are also random - in the sense that we may have picked a time where the relation is coincidentally in favour of our method/reasoning and finally, (4) we do not make a good job to put the capabilities of our approach into the context of what is already out there.

We revised our manuscript considerably in order to overcome this shortcoming: in the revised version we added configuration Coarse2 to Figure 7 and introduced an additional Figure 8, showing the temporal evolution of all simulations throughout year two and three. Further we added rates to Table 2. A discussion section was added (starting ln. 330), stating that the saturation signals due to mixing are indeed small in amplitude - but nonetheless consistent over time. From this we derive an estimate on the precision of our method. The new discussion section does also put our results into the context of similar contemporary methods - as suggested by the reviewer.

As concerns "issue 2": the Ekman pumping has been put forward in the literature to make the point that resolution matters in terms of vertical (and eventually diapycnal) transports. We realize (and agree with the reviewer), however, that such a relation is neither obvious nor stringent. We changed the text accordingly and state now in line 285: "... We are particularly curious whether the increase in Ekman dynamics with increasing resolution leads to enhanced diapycnal mixing or if its effect is primarily advective. ..."

In summary, the new discussion Section addresses issue #1. Issue #2 is resolved in line 285. The additional Figure 8 and the update of Table 2 address issue #3. The additional appendix B provides another rationale for the chosen spin-up length and the model run time (issue #3).

R: Specific comments:

Line 3: The authors say “unacquainted vertical” by which they are referring to “diapycnal” but that way of describing diapycnal sounds more literary than it sounds like it has specificity; did the authors use AI to help them write this abstract?

A: We deleted "unacquainted".

R: Lines 38-39: The canonical reference for uncertainties in representation of diapycnal mixing (not just vertical) is MacKinnon et al. (2017)'s “Climate Process Team on Internal-Wave Driven Ocean Mixing” even though there's also shear-driven and other mixing

A: Thanks. We included the reference in ln. 39.

R: Line 45: This seems like more “coupled” reasoning than circular because the biogeochemical modules can potentially be developed to be more realistic without developing the diapycnal diffusivity modules, but of course the biogeochemical variables depend upon the diapycnal diffusivities. It's difficult to point to new model innovations in ESMs that don't have this coupling problem. Your argument seems analogous to saying that assimilating sea-ice concentrations with the goal of improving the sea surface properties is circular reasoning because our sea-ice modules are imperfect. In any case, without improvements in the biogeochemical modules, the information propagated by the assimilation of biogeochemical variables may not inform the diapycnal diffusivities as much as the biogeochemical parameters themselves, which is the problem you're pointing out. This would certainly be the case with the Green's function-based approach of ECCO-Darwin (see Carroll et al. (2020)'s “The ECCO-Darwin Data-assimilative Global Ocean Biogeochemistry Model: Estimates of Seasonal to Multi-decadal Surface Ocean pCO₂ and Air-sea CO₂ Flux”) because the physical variables aren't changed in their formulation; only the biogeochemical parameters are altered. I believe the proposal by Trossman et al. (2022) is that the ECCO-Darwin and ECCO setups could be used in an iterative process to incrementally improve the biogeochemical parameters and the physical parameters, respectively, but this is even more computationally expensive than the current process and may require a different process for sequential data assimilation systems, which they present unique issues with in their appendix. Additional evidence that this could work comes from Skakala et al. (2022)'s “The impact of ocean biogeochemistry on physics and its consequences for modelling shelf seas,” where it was shown that assimilation of a suite of biogeochemical variables had an impact on the diapycnal diffusivities beyond that of physical variables.

We removed the "circular reasoning" and toned the paragraph down so that it is no longer so pessimistic in that we merely state that the approach is not guaranteed to succeed in all cases (ln. 44 - 47).

R: Lines 63-64: Will this always be the case? The atmospheric community seems to have come up with a solution to advection numerics in McGraw et al. (2024)'s "Preserving Tracer Correlations in Moment-Based Atmospheric Transport Models". Also, is it possible that spurious diapycnal mixing is the result of spurious vertical advection sometimes? This certainly happens when you perform sequential assimilation (see: Pilo et al. (2018)'s "Impact of data assimilation on vertical velocities in an eddy resolving ocean model").

We included the McGraw (2024) reference in ln. 62 because we agree that it makes sense to point to ongoing developments. As to the applicability of their approach in ocean models (where all vertical layers are touching land somewhere in contrast to atmospheric models where most vertical layers are above orographic features) we are not so sure.

Yes, advection schemes can introduce spurious effects in all dimensions, including vertically. We added "... 3-dimensional ..." to ln. 60.

R: Lines 81-82: Isn't there also the effect of surface disequilibrium at the time of water mass formation (e.g., when subduction occurs) that needs to be accounted for? Both the nonlinearity of solubility in the presence of mixing and surface disequilibrium effect cause apparent oxygen utilization to be different from true oxygen utilization, for example. So the argon tracer needs to be injected below the mixed layer such that it doesn't get obducted throughout the experiment. Also, argon can't be injected into the ocean to derive diapycnal diffusivities from observations (unless the background levels are well-known) because there are significant background levels of argon in the ocean, which will eventually render it another exotic tracer on top of the ones Ledwell used. I mention this because it was noted that argon could be used as a proxy for effective diapycnal mixing in both models and observations in the text.

A: Agreed. We now stress this point more prominently in an additional Discussion Section (starting ln. 337).

R: Line 150: I understand not wanting to compete with Jason Momoa in Google searches by dubbing your model's acronym MOMOA, but you're going to compete with the Museum of Modern Art in New York City instead. In any case, the model and its tracer conservation properties won't be as much of a problem as they would be if the authors chose another model like HYCOM. And the model domain is reasonable, considering problems with boundary conditions that can occur.

A: The model has already been published under this name. As concerns HYCOM: We are really interested in testing our method in an isopycnal model!

R: Line 169: You say that there are 55 geopotential levels in all model configurations here but in Table 1, it says there are 72 grid points in an undefined dimension. The caption of Table 1 only mentions zonal and meridional directions but there is a third dimension listed for the number of grid points. Is that dimension time or is one of your instances of vertical dimensions inaccurate (i.e., in Table 1 or on Line 169)?

A: Thank you for catching this. Ln. 169 was in error and is corrected now (now Ln. 170). There are 72 vertical boxes.

R: Line 186: For future equilibrations of biogeochemical variables, you could consider using another method like the one by Khatiwala (2024)'s "Efficient spin-up of Earth System Models using sequence acceleration"

A: We added the respective citation to the revised version of the manuscript (now Ln. 188).

R: Line 189: I would argue that use of COREv2 forcing is no longer considered standard, as JRA55-do has supplanted COREv2 in OMIP2 and ERA5 is also commonly used

A: You are right. We added the references and changed the wording to: "... a well-tested annual climatological cycle of all the data needed to force an ocean model (note that similar contemporary products exist; Tsujino et al., 2018; Hersbach et al., 2020)." now Ln. 190

R: Lines 201-203: If you're interested in the evolution of each of the model configurations, then you should be assessing metrics that are related to the time rate of change, not averages

A: We added rates of change to Table 2 and added the temporal evolution of simulation Coarse2 to Figure 7d. In addition, we show now the temporal evolutions of all simulations in the new Figure 8. We hope that this additional information makes our approach more convincing.

R: Lines 205-206: One year of spin-up is common amongst published high-resolution model simulations (based on the global kinetic energy metric you seem to be using) because the initial shock is essentially adjusted away but the model's climatological state is still being approached asymptotically. That would need at least several more years to achieve. It may take much longer to equilibrate potential energy, which includes available potential energy relevant to the eddy circulation, but to first order using kinetic energy may be good enough.

A: You are right. What we actually want to say is that a comparison of simulations without any energy would be meaningless, because without movement all resolutions will look alike. So having enough energy, i.e. not falling short of observational estimates, is essential for our study. But, as you pointed out rightfully, it does not prove that energy has equilibrated. We changed the text to (now Ln. 204): "... We compare simulations that have not reached equilibrium yet (after the short integration of only three years at most) which constrains potentially spurious effects of our lateral boundary conditions (c.f. Appendix \ref{app:boundary}) on our results and saves computational resources. Note in this context that \cite{dilma21} showed that the circulation of MOMA is already realistic after a spin-up of only one seasonal cycle. Consistent with this result we find that the energy content of our eddy-field does not fall short of observational estimates after one year already (Figure~\ref{fig:05}). This suggests that our simulation period of 3 years is long enough to

reveal major effects of differing resolution (even though it may take much longer to equilibrate potential energy). ..."

R: Table 2: Continuing on from my previous comments, please include the time rate of change (linear slope of a regression against time, for example) of the Delta Ar[%] metric you include the average of in this table. That would help the reader see whether these values are close enough to what they would be in a more well-spun-up state. Also, why are you including the mixed layer in your calculations when Argon is not a conservative tracer near the sea surface and diapycnal mixing is not very meaningful within a bulk mixed layer? Are you not using a bulk mixed layer?

A: We added respective rates to Table 2. Further, we added the simulation Coarse2 to the temporal evolution in Figure 7. In addition, we introduced Figure A1 which shows the temporal evolution of all simulations (with the exception of Coarse upwind which was off the scale). Further, we excluded the upper 100m from all of our analysis which, indeed, removes some of the noise. Thank you for the suggestion!

R: Lines 234-236: Again, because penetrating solar radiation, bubble entrainment, and possibly other processes affect Delta Ar[%], why are you including the mixed layer in your calculations?

A: You are right. By discarding the upper water column down to the maximum winter surface mixed layer we actually got rid of a lot of noise (mainly the effect of subsurface solar heating accumulating below the summer surface mixed layer). We changed Table 2 and Figure 7d accordingly.

R: Lines 266-268: This is qualitatively accurate. However, it can be noted that there are in-situ estimates of eddy kinetic energy compared with high-resolution model simulations in the literature, like those presented in Luecke et al. (2020)'s "Statistical Comparisons of Temperature Variance and Kinetic Energy in Global Ocean Models and Observations: Results From Mesoscale to Internal Wave Frequencies". Also, along-track SWOT data resolves the higher wavenumbers and there are preliminary results that much more of the spectrum is resolved with SWOT. I'm not suggesting that you perform a comparison with in-situ (e.g., moored) or SWOT data. This is just something to be aware of.

A: Thanks. We added the respective reference (now In. 269).

R: Figure 3: It's curious that the relatively warm SST feature near 18W, 21N in the High configuration is completely absent in Coarse, except for what seems like a separate merging feature near the upper part of the domain plotted here. The rest of the figure comparison looks like the High configuration's SST was coarse-grained.

A: Agreed. We changed the wording to (now In. 251): "... On these scales it is evident that a higher spatial model resolution features more small-scale circulation features such as the relatively warm SST feature near 18°W , 21°N that is completely absent in Coarse. ..."

R: Table 3 and Figure 6: Okay, the higher the resolution, the greater the Ekman pumping/suction there is in its mean and variability, but how come you're spending time on this when you're not relating it to the effective diapycnal mixing? You're relating how resolving the wind stress curl and upper-ocean features to diapycnal fluxes of tracer properties within the Ekman layer, which has more to do with advection. It's possible, however, that, by conservation of volume in your model, diapycnal mixing will counter advective diapycnal fluxes but you are not showing this.

A: You are right. We changed the wording to (now ln. 283): " Hence we expect, consistent with the pioneering work of, e.g., \citet{levy01, mahadevan00, mahadevan06, martin03}, an increase in vertical, fluxes of heat, salt and biogeochemical entities with increasing resolution in our configurations. We are particularly curious whether the increase in Ekman dynamics with increasing resolution leads to enhanced diapycnal mixing or if its effect is primarily advective.

R: Lines 313-319: The differences in domain-averaged Delta Ar[%] amounting to <0.1% with the same back diffusivities (equivalent to increasing the background diffusivity from 20% to 40%) may not be significant/detectable here once you consider the rate at which Delta Ar[%] and/or global kinetic energy is changing as the simulations are spinning-up. It appears that in Figure 7d that the seasonal variability and trend in (Delta?) Ar[%] may be larger than the difference across High and Coarse. So it's unclear whether the comparisons you make with the comparable effect of varying the background diffusivity and/or numerical advection scheme will hold up if your model was spun up more. A decrease in Delta Ar[%] from using a background diffusivity of $0.14 \text{ cm}^2 \text{ s}^{-1}$ to $0.16 \text{ cm}^2 \text{ s}^{-1}$ suggests varying rates of spin-up across different background diffusivities, for instance.

A: We added the Coarse2 simulation to Figure 7d and rates of change to Table 2. Also we introduce the additional Figure 8 which shows the temporal evolution of Ar saturation in all of our simulations. Our argument is that, although the signals are rather small, they are distinguishable and consistent over (quite some) time.

Minor corrections:

A: Thank you for taking the time going through the text so carefully. We really appreciate this and corrected everything accordingly.

We are grateful to reviewer #2! Her/his issue with potentially spurious effects of boundary effects is justified and we think that the manuscript has become stronger now that we added a respective analysis based on a simulation of the time elapsed since water parcels were in contact with the boundaries.

R: Summary:

The main finding of this paper is that diapycnal mixing, a key source of uncertainty in ocean models, is strongly influenced by factors such as model resolution, advection schemes, and background diffusivity. Using Argon saturation as a proxy, the study shows that the effect of differences in resolution is comparable to that of other factors. This study highlights the ongoing challenge of accurately representing ocean mixing in models used for climate projections. I found this manuscript concise and well-written. However, some clarifications in methods and a few minor corrections are needed before publication. I suggest minor revisions. Please see the general comments and specific comments below.

A: Thank you. We found your comments very helpful, especially those related to the impact of the boundaries. We have added respective additional analysis & discussion (discussion In. 330ff, Appendix B: Boundary Conditions In. 447ff) and think the paper has become stronger.

R: General comments:

1. Consider changing the title to highlight the key findings or purpose of this study. The current title is concise, but it could be refined for better focus and impact.

We changed the title to: "Mixing, Spatial Resolution and Argon Saturation in a Suite of Coupled General Ocean Circulation Biogeochemical Model Configurations off Mauretania."

2. While the rationale for using a 3-year simulation period is clear in terms of model spin-up and avoiding equilibrium, I have concerns about the potential influence of spurious effects from the closed boundaries on the region of interest. Could the authors provide diagnostics or an analysis that evaluates how far spurious effects, such as reflected waves or boundary-induced disturbances, propagate into the domain over the 3-year period? This would help confirm whether the region of interest remains unaffected.

A: We added a new analysis to the appendix (starting in now In. 443) showing that it takes, on average, more than three years until material (as opposed to energy) is transferred from the boundaries into the high-resolution domain. In In. 204 we state now: "... We compare simulations that have not reached equilibrium yet (after the short integration of only three years at most) which constrains potentially spurious effects of our lateral boundary conditions (c.f. Appendix \ref{app:boundary}) on our results and saves computational resources. ..."

R: Detailed comments:

1. Line 134: What is the horizontal resolution of MOMSO? Is it 50 km as shown in Fig. 1? It is less than 11km. We added this information to ln. 134.

2. Line 165: "resolutions" should be "resolution."

A: Yes, thank you.

3. Line 169: I have a question regarding the vertical grid description in line 169, where it is mentioned that the model uses 55 vertical levels but 72 vertical grid points in Table 1. Could you kindly clarify this aspect? Is the difference in grid points due to the use of staggered grids, boundary conditions, or another feature of the model's vertical discretization? If there is a specific reason for this configuration, I would appreciate further details as a reader.

A: Sorry. Ln. 169 is wrong. There are 72 vertical levels. Corrected. Now ln. 170.

4. Figure 1 caption: Panel (c) and panel (d) don't show vertical resolution, right? Am I missing anything?

A: You are correct. We deleted "vertical" in the last sentence of the caption.

5. Line 188: I believe "affect" would be better than "effect" here.

A: We replaced "affect" by "drive"; now ln. 190

6. Line 225: It seems to me that equation (1) is incorrect. Since $\Delta Ar = Ar - Ar_{sat}$, why do we need to subtract 100% in this equation? This might be a typo, but if not, please check the calculations presented in this study. Or am I missing anything?

A: No, you are right! This is a typo. Corrected. Now in ln. 226.

7. Line 229: Could the authors elaborate on why mixing always causes an increase in saturation? Wouldn't the sign of the change in saturation depend on the temperature before and after mixing, given that Argon saturation is a non-linear function of temperature?

A: You are right, non-linearity is not enough. We changed the wording to: "... In contrast to concentration the saturation is not conservative: a mixed water-parcel will always carry a higher argon saturation than the arithmetic mean of the original parcels since the saturation curves are convex over the entire range of oceanic temperatures (and salinities). Hence, an increase in $\Delta Ar\%$ may be indicative of mixing. ..." Now in ln. 226.

8. Line 229: There is a typo. Change "his" to "is."

A: Fixed.

9. Line 233: "Allows" should be "allow."

A: Fixed.

10. Line 266: There is an extra parenthesis.

A: Fixed.

11. Line 295: "loses" should replace "looses."

A: Fixed.