Dear Editor and Reviewer,

We would like to thank you for accurately reading and commenting the manuscript and suggesting how to improve it. Answers to your comments are given in details hereafter. We hope that you will find them satisfactory. All authors agree with the modifications made to the manuscript. Reviewer comments are in black and are followed by our response (in blue) that includes changes and/or additions to the text.

For the authors,
Doroteaciro Iovino

RC1: 'Comment on egusphere-2023-469', Anonymous Referee #1

The present manuscript presents a high level overview of the differences and similarities between three configurations of the CMCC ocean model in forced simulations. The comparisons are limited to the simulated state - no of process level explanations of the differences are provided. Further, while the emphasis is on the impact of horizontal resolution, there are a large number of other differences in the models being compared: vertical resolution, topography, the sea ice component model and the sea ice initialization, the salinity restoring timescale, the way that runoff forcing is applied, along with the more typical adjustments to viscous and diffusive parameterizations. So, while the authors have put in a considerable amount of effort in compiling model metrics, the resulting manuscript is rather unsatisfying. We have little new insight into how and why the explicit representation versus parameterization of the mesoscale impacts the simulated ocean state. As a documentation of what was done the manuscript may be adequate with some additional work to correct imprecise descriptions, but I do not envision the manuscript having much of an impact beyond those who might wish to use the CMCC models and it is unlikely to advance the field. Additionally, the manuscript suffers from being poorly prepared with many missing references and poor language constructions.

This study exploited numerical system already implemented and in use and took advantage of ongoing simulations rather than a dedicated set of tests and experiments. The low/medium- and high-resolution should have been configured in tandem but they were designed independently for different scientific goals and developed for distinct applications. So, the models are diverse in their components, numerics and parameterizations and we do agree that the effects of horizontal resolution cannot be completely isolated. We have better addressed this issue in the introduction, model description and conclusions and have explicitly highlighted the need of an improved design of future comparisons.
This work mainly aims to present the performance of the eddy-rich ocean/ice system in relation to the lower resolution configurations in hindcast runs. Specific analysis on how the explicit representation of mesoscale dynamics impacts the ocean state will be shown in dedicated studies for instance on the variability of ocean eddy kinetic energy and the eddy exchanges from narrow boundary currents and basin interior in marginal seas (manuscripts are in preparation).
We closely followed comments and suggestions to improve the quality of the manuscript.

Detailed Comments:

Line 26 "enforces our ability in understanding": poor wording, perhaps is a prerequisite to develop our understanding
We reworded the sentence as follows "An accurate representation of the ocean dynamics within the climate system is crucial to understanding drivers of climate change and variability, and to determining the ocean-ice influence on atmospheric circulation and ecosystems."

Line 29-30: ensemble size is a mother strong trade-off in both prediction and climate modeling.
We reworded this sentence as follows "Despite the ongoing increases in computer power and improvements in techniques, a major challenge in climate model design is the trade-off between the level of model complexity, the length of simulations, the choice of ensemble size and the spatial resolution of different climate components."

Line 31 “start grid spacing”: rather, typical grid spacing. No standard (a specification) exists.
Done

Line 33 “miss key processes”: it is not necessarily the case that key processes are missed. They may be parameterized.
We reworded this sentence as follows “Both resolutions lack an explicit representation of ocean mesoscale dynamics in most of the global domain.”

Line 36 “Despite”: While simulations at this resolution …
We modified the sentence as follows “Simulations of the global ocean domain at this resolution still require significant computational resources, which limits the number and length of runs and the capacity to optimize the model setup. However, thanks to the ever-increasing processing and storage capabilities of the supercomputers, running global models capable of resolving mesoscale dynamics has become feasible for climate simulations. It is now necessary to assess to what extent the enhanced resolution translates into an improved ocean state.”

Line 38: “access to which”: to assess to what extent
Done

Line 60 “resolution dependent”: as discussed at top this is not a convergence study in the sense of numerical analysis. Many other aspects of the simulation besides horizontal resolution are change.
Thanks for the comment. We reworded the sentences as follows “We run OMIP-like simulations with the three models driven by the same forcing dataset, and we compare them in order to identify possible climate-relevant improvements in the ocean response as model resolution increases. It is worth mentioning that the models do not differ only in the horizontal resolution and associated physical parameters since the high-resolution simulation was configured independently for distinct scientific applications and followed a specific development strategy.”

Line 68: Manral et al 2013 ref missing
Reference added

Line 104: suggest starting new paragraph with sentence beginning “While the best …”
Done

Line 154 “in coupled runs”: all the runs described herein are forced not coupled?
We do agree with the referee; this sentence was misleading since all simulations are forced even when the framework of the coupled system is used. We modified the sentence as follows “the initial sea ice properties in ORCA1 and ORCA025 runs are...”.

Line 189: do not correspond to what _was_ found in
Done

Line 203-204: I do not understand this distinction. The difference from the initial state taken from observations is the model bias?

We modified the sentence as follows "This metric shows to what extent and how quickly the modelled 3D temperature deviates from the ocean initial state as the resolution changes. The anomaly for a specific date is computed as the difference between this current value and the WOA13 temperature."

Line 210: Lellouche et al 2021 ref missing

Reference added

Table 2: is the standard deviation stated the inter annual standard deviation of the global mean SST or the mean standard deviation of the global spatial deviation of SST?

It is the interannual standard deviation of the global mean SST as now specified in the caption "Global annual mean, its standard deviation and linear trend of sea surface temperature for the period 1982-2018 (common to all SST datasets), ..."

Line 254: I don't understand the difference between "model physics", which I generally take to be parameterizations and "unresolved processes" which require parameterization?

Thank you for the comment. The opening sentence of the paragraph was inaccurate and redundant. It was deleted.

Line 260: Most of the SST biases are reduced": This is not visually that obvious. State the rms error of annual mean SST at each resolution.

This is correct. The global averaged SST error is similar among models, GLOB16 is not the lowest. There are improvements at local scales. We added the SST RMSE in the manuscript.

Section 3.3.2 and Figures 5-6: This discussion would be much improved by showing panels with the summer MLD (JJA for NH and JFM for SH) as a single plot and winter MLD (vice versa) as a single plot with the color scale appropriate to each season. The discussion of similarities in summer or lower latitude features is completely obscured by the full annual range color scale. This is also a very long section with little insight beyond that they are different. To what extent can we attribute differences to changes in preconditioning of water masses due to differences in large scale flow versus changes in the restartification power of mixed layer eddies in each case? Does the ORCA1 model include a parameterization (e.g. Fox-Kemper e al) of submesoscale mixed laser restratification included in its GM parameterization?

Please note that figure numbering has changed in the revised manuscript from Figure 5 onwards; for consistency with the referee's comments, hereafter we use the figure numbers as in the first submitted paper and indicate the new one in brackets.

Thanks for this comment, which helped to improve the MLD plots. As suggested, we produced a new version of figure 5 [*figure 6 in the revised manuscript*] and figure 6 [*figure 7 in the revised manuscript* that show the winter and summer MLD, respectively, with more appropriate color scale. We decided to keep the March/September spatial distribution of the MLD that is very similar to the JJA/JFM means. The 3-month mean mainly impacts the magnitude of the MLD that is reduced in both observation and models. All plots are attached below for a comparison. The text has been slightly shortened.

A dedicated multi-model study, coordinated by CMCC and based on a larger suite of OMIP simulations at low and high resolutions, will analyze the role that oceanic mesoscale eddies play, in the Labrador Sea, in determining both the location and strength of deep winter convection and the re-stratification of the convected water mass during spring and summer.
The parameterization scheme for sub-mesoscale mixed layer eddies designed by Fox-Kemper et al. (2011) is not included in the CMCC ORCA1 configuration used in this study. This choice is supported by some previous studies based on forced and coupled simulations. The use of the parameterization to improve the mixed layer representation does not consistently result in better performances than models without the parameterization (e.g. Heuzé et al., 2017, Calvert et al.,2020).

- Calvert, D., G. Nurser, M. J. Bell, and B. Fox-Kemper: The impact of a parameterisation of submesoscale mixed layer eddies on mixed layer depths in the NEMO ocean model. *Ocean Modelling*, 154:101678, https://doi.org/10.1016/j.ocemod.2020.101678, 2020.
- Fox-Kemper, B., G. Danabasoglu, R. Ferrari, S. M. Griffies, R. W. Hallberg, M. M. Holland, M. E. Maltrud, S. Peacock, and B. L. Samuels: Parameterization of mixed layer eddies. III: Implementation and impact in global ocean climate simulations. *Ocean Modelling*, 39:61-78, https://doi.org/10.1016/j.ocemod.2010.09.002, 2011.
- Heuzé, C.: North Atlantic deep water formation and AMOC in CMIP5 models. *Ocean Science*, 13, 609–622, https://doi.org/10.5194/os-13-609-2017, 2017.

A)

B)

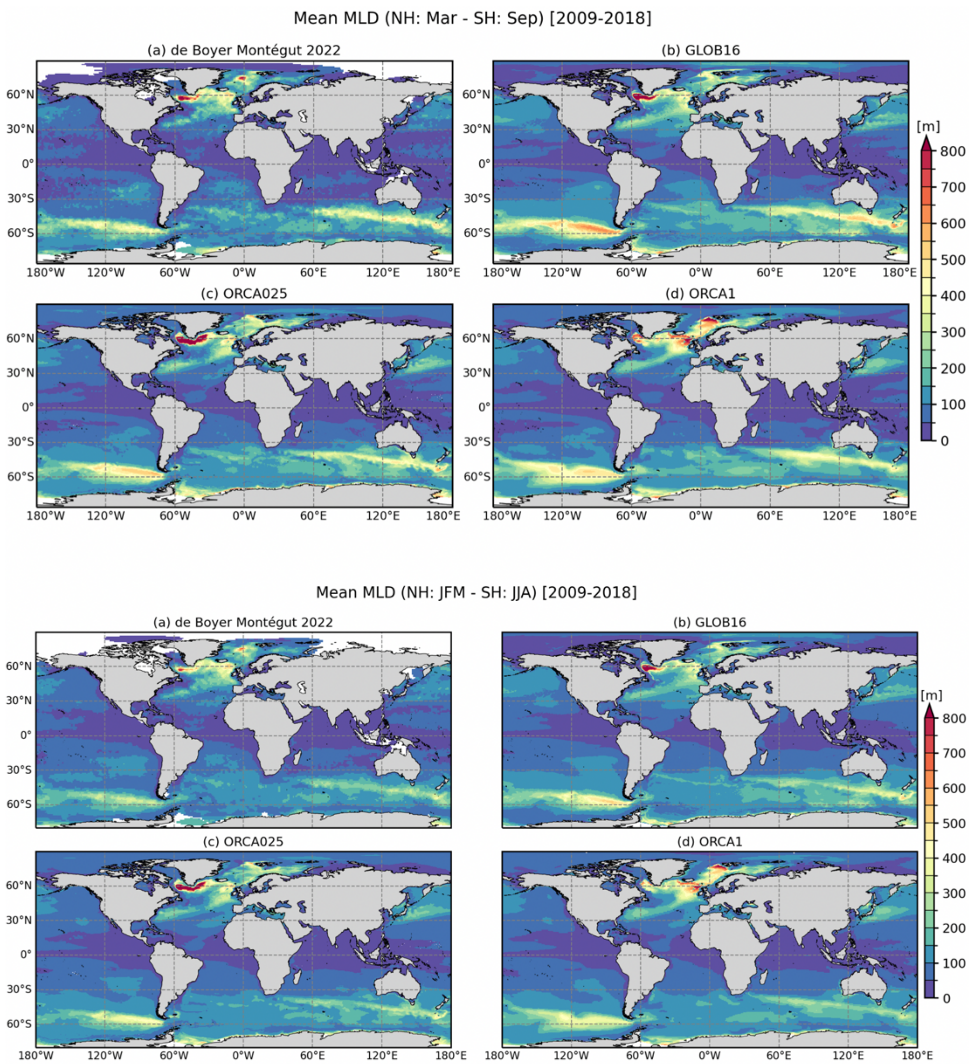
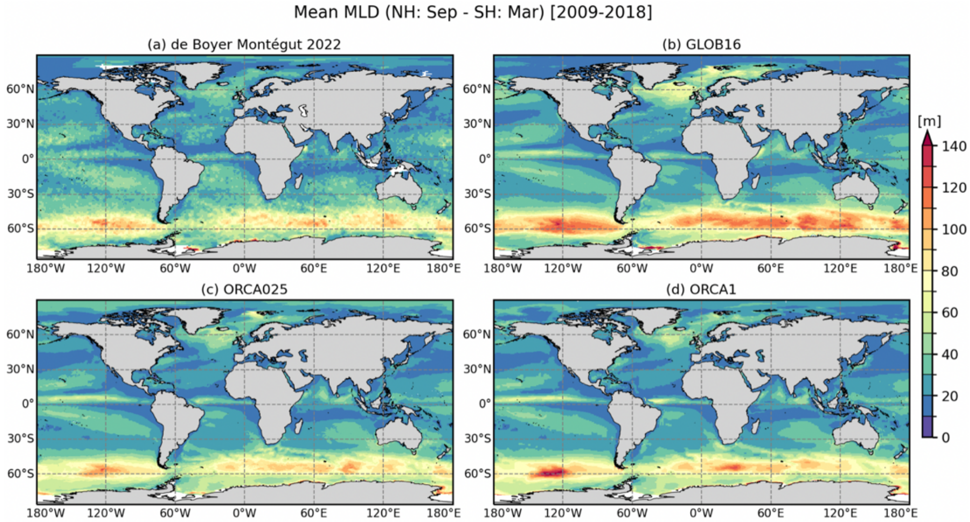


Figure RC1_1. Map of mean mixed layer depth (in m) for

A) March in the Northern Hemisphere and September in the Southern Hemisphere, from (a) observation-based estimates from de Boyer Montégut et al. (2022), (b) GLOB16, (c) ORCA025 and (d) ORCA1. MLD fields are computed as the monthly climatology over last 10-year output.
B) as A but averaged over boreal winter (January-February-March) in the Northern Hemisphere and boreal summer (June-July-August) in the Southern Hemisphere.

a)



b)

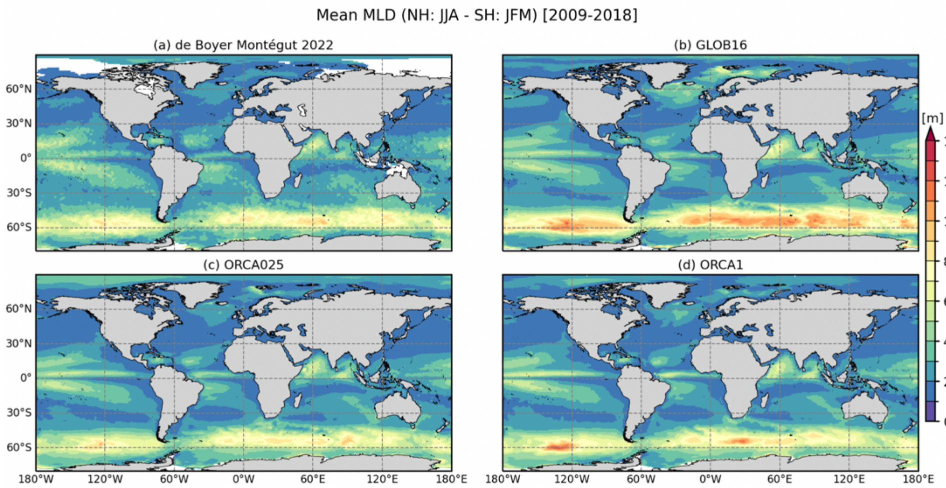


Figure RC1_2. Map of mean mixed layer depth (in m) for
A) September in the Northern Hemisphere and March in the Southern Hemisphere, from (a) observation-based estimates from de Boyer Montégut et al. (2022), (b) GLOB16, (c) ORCA025 and (d) ORCA1. MLD fields are computed as the monthly climatology over last 10-year output.
B) as A but averaged over boreal summer (June-July-August) in the Northern Hemisphere and boreal winter (January-February-March) in the Southern Hemisphere.

Line 338-342 discussion of higher order statistics in Johnson and Lyman 2022: What is the relevance to this study none of these statistics are evaluated in the paper

GOSML dataset by Johnson and Lyman (2022) is constructed as a statistical monthly climatology of the global mixed layer depth, temperature, and salinity determined from the Argo profiles using the density algorithm of Holte & Talley (2009). The dataset includes means and variances, plus additional statistics for mixed layer properties including the median (50th percentile), 5 th, and 95 th percentiles, as well as skewness and kurtosis. Johnson and Lyman (2022) find that the distribution of MLD is non gaussian, with large skewness and kurtosis that vary seasonally and spatially. The

MLD variance displays seasonal variations and depends on the MLD itself (regions with deep ML have a large MLD variance). The properties of the MLD statistics for the OMIP simulations are analysed in Treguier et al. (2023). Studying the mixed layer statistics is not the aim of this manuscript. Here we used the GOSML only as a reference dataset for validating the latitudinal variability of the simulated mean MLD in March and September, together with the observational dataset compiled by de Boyer Montégut et al. that uses a different mixed layer definition.

- Johnson, G. C., and Lyman, J. M.: GOSML: A global ocean surface mixed layer statistical monthly climatology: Means, percentiles, skewness, and kurtosis. J. Geophys. Res.: Oceans 127. doi: 10.1029/2021JC018219, 2022.

- Treguier, A. M., de Boyer Montégut, C., Bozec, A., Chassignet, E. P., Fox-Kemper, B., McC. Hogg, A., Iovino, D., Kiss, A. E., Le Sommer, J., Li, Y., Lin, P., Lique, C., Liu, H., Serazin, G., Sidorenko, D., Wang, Q., Xu, X., and Yeager, S.: The mixed-layer depth in the Ocean Model Intercomparison Project (OMIP): impact of resolving mesoscale eddies, Geosci. Model Dev., 16, 3849–3872, https://doi.org/10.5194/gmd-16-3849-2023, 2023.

Line 349: Johnson and Lyman 2022 ref missing
Reference added

Line 350: again, the use of the full dynamic range in the axis scale makes it difficult to compare the quality of the simulation of shallower mixed layers. The relative error could be just as large as for deeper M.
We modified the MLD plots with color scales that are now appropriate to each season. Please see above the comment on Figure 5 [*figure 6 in the revised manuscript*] and Figure 6 [*figure 7 in the revised manuscript*].

Line 374 “representations is underrepresented”: nonsensical phrase
We reworded the sentence.

Line 377: “unable to represent flow instabilities …”: yes, but the figure is showing mean low speed, not EKE
In accordance with the Referee’s suggestion, we changed the sentence as follows “It captures the major current systems of the global ocean, but it underestimates the magnitude of the surface velocity field and fails to represent mesoscale eddies and meanders.”

Line 385: “dependent on model numerics”: how so? Don’t all of the models use the same numerical methods to solve the equations of motion?
Among the three configurations, the numerical methods are similar but not exactly the same. The western boundary currents differences are also dependent, for example, on differences in the lateral boundary conditions and topography. More details on the physical parameters are now provided in Table 1.

Line 386: “impact of mesoscale dynamics”: This has not been shown. It could simply be the impact of viscous boundary layer dynamics or topography which also differ across configurations
Following the comment, we reworded as “The Gulf Stream simulated by the three models is depicted in Figure 10 (left column).”

Line 404 "passed": past
Done

Line 421 "Figure 10 shows role of mesoscale eddy field …": again, the figure shows the mean flow and no analysis of eddy-mean flow interaction is provided.
Thank you for the comment. We reworded as "Figure 11 shows the complex ocean circulation in the Southern Ocean sector...".

Section 3.3.2 (sigma overturning): A more precise definition of how the stream function was calculated at each resolution is required. Was it computed from Eulerian mean (monthly, annual, climatological?) velocity and density fields? Was the GM eddy-induced velocity included in the ORCA1 result? The authors should write down the integral with averaging operators in the appropriate places for clarity. This is important in trying to understand whether differences seen are related to "bolus velocity", diapycnal processes or surface forcing. The reference to Andrews and McIntyre suggest that we should be interpreting something about eddy induced transport, but the discussion is unclear. One of the major differences is the structure of the strong clockwise cell in the ACC region which I presume is related to the degree of compensation of the Deacon cell. No discussion of this feature is provided, nor are we sure how to interpret the result given the uncertainty about exactly what is being shown.
The MOC in density space is computed following the formulation by Farneti et al 2015 (equation 7 in Appendix A). In section 3.3.2, we added this reference to the computation of the MOC on potential density surfaces (referenced to 2000 dbar) from monthly meridional velocity and density fields.
We also specified that the meridional velocity in output to the ORCA1 model is the sum of the Eulerian-mean velocity and the GM eddy-induced component obtained through GM parameterization (Gent and McWilliams 1990). The reference to Andrews and McIntyre (1978) was wrongly placed.
The use of potential density as the vertical coordinate, rather than depth, results in a better characterization of water mass transport and is more suitable for representing the MOC in the Southern Ocean. In particular, the wind-driven Deacon cell, which normally appears in depth-space MOC, is mostly due to a geometrical effect of the east-west slope of the isopycnals when the zonal and vertical integration is computed at fixed depth levels. No cross-isopycnal flow is associated with it (Döös and Webb 1994, Farneti et al. 2015).
The MOC structure presented in Figure 11 [*figure 12 in the revised manuscript*], from the southernmost boundary to 30ºS, agrees with previous studies (e.g. Farneti et al. 2015). The wind-driven subtropical cell is part of the horizontal subtropical gyres and is confined to the lightest density classes. This anticlockwise cell comprises a surface flow spreading poleward to 40S, compensated by an equatorward return flow. Below, the upper cell is depicted by the large clockwise circulation, which mainly consists of upper circumpolar deep water. The anticlockwise lower cell, in the densest layers, that consists of the poleward lower circumpolar deep water and the deeper equatorward AABW. From 60ºS to the Antarctic continent, the transport represents the contribution of subpolar gyres in the Weddell and Ross Seas. The revised text includes this description and a better characterization of the water masses.
The global MOC for the three simulations is presented in the plots below in depth space (left) and potential density space (right). In all models, south of 30S, the MOC in depth space shows the clockwise Deacon and upper cells, and the anti-clockwise lower, subtropical and subpolar cells. In density-space, the Deacon Cell disappears, and the surface waters recirculates in the ACC anti-

clockwise upper cell. This is in very good agreement with the schematics of the main Southern Ocean cells (see Figure 16 in Farneti et al. 2015).

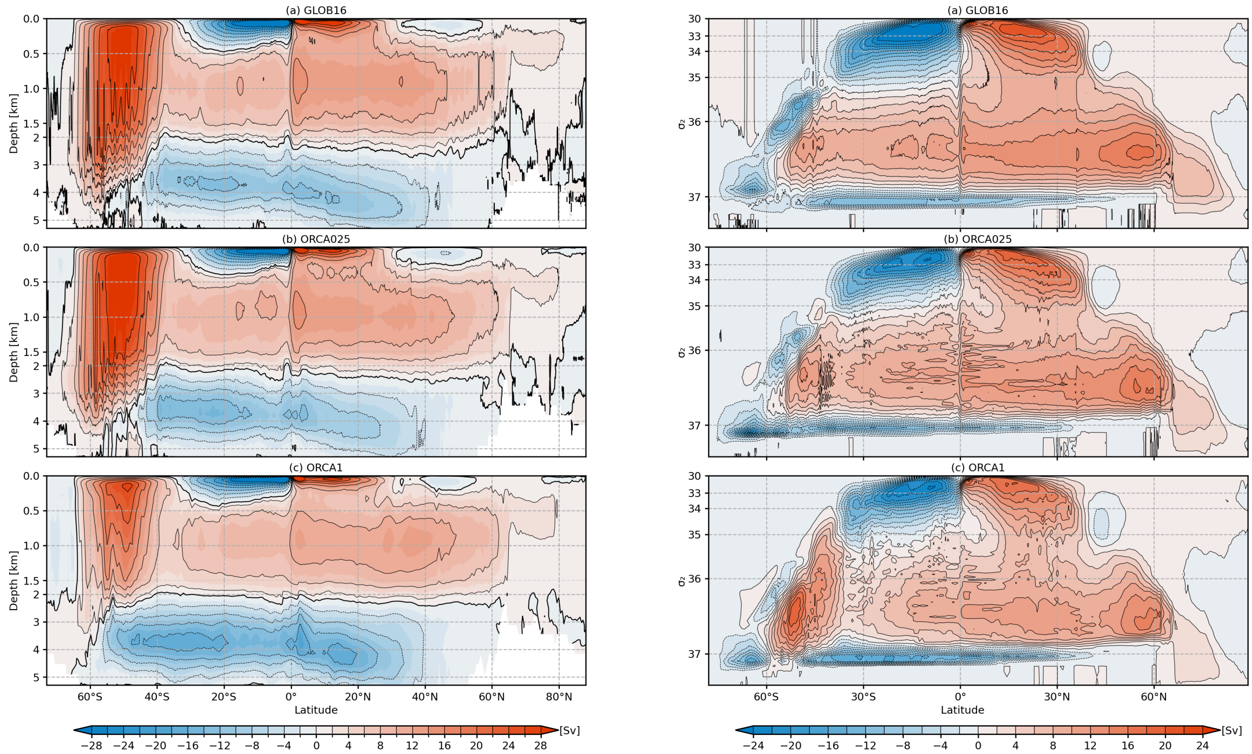


Figure RC1_3. Time-mean zonally integrated MOC computed in depth space (left) and in density space as function of σ_2 for the global ocean as reproduced by the three simulations. Overturning is averaged over the last 10 years of simulation (2009–2018).

- Farneti, R., Downes, S. M., Griffies, S. M., Marsland, S. J., et al..: An assessment of Antarctic Circumpolar Current and Southern Ocean meridional overturning circulation during 1988–2007 in a suite of interannual CORE-1125 II simulations, Ocean Model., 93, 84–120, https://doi.org/10.1016/j.ocemod.2015.07.009, 2015.

Line 466: "This suggests that longer integrations are required for GLOB16 …": It was previously stated that the analysis of all cases was for the first cycle (Line 150). Why is that only GLOB16 requires longer integration to be compared?

We followed this comment and rephrased the imprecise sentence as follows "While the upper ocean takes decades to achieve equilibrium, the deep ocean adjustment requires hundreds of years to reach a quasi-equilibrium state (e.g. Danabasoglu et al., 1996) because of the slow diffusion of active tracers. Tsujino et al. (2020) show that OMIP2 low-resolution simulations take about four cycles to spin-up, and the AMOC declines in the first cycle and slowly recovers thereafter. A longer GLOB16 integration would be necessary to reach a quasi-equilibrium behavior of the overturning in the deep ocean and analyze the long-term evolution of deep-water properties from the initial state also in the eddying ocean."

- Danabasoglu, G., J. C. McWilliams, and W. G. Large: Approach to equilibrium in accelerated global oceanic models. J. Climate, 9, 1092–1110, https://doi.org/10.1175/1520-0442(1996)009<1092:ATEIAG>2.0.CO;2, 1996.

- Tsujino, H., Urakawa, L. S., Griffies, S. M., Danabasoglu, G., et al.: Evaluation of global ocean–sea-ice model simulations based on the experimental protocols of the Ocean Model Intercomparison Project phase 2 (OMIP-2), Geosci. Model Dev., 13, 3643–3708, https://doi.org/10.5194/gmd-13-3643-2020, 2020.

Figures 13-14 and accompanying discussion: This is a long discussion of the GLOB16 results alone with no explicit comparison across resolutions. It seems unnecessary if the purpose of the paper is investigating resolution dependence rather than an assessment of GLOB16 alone.

The main purposes of this paper are to evaluate the GLOB16 model performance and to document if and how the CMCC "eddy-resolving" ocean model resolution change the representation of large-scale ocean variability with respect to observations and lower-resolution models, highlighting the relative advantages and disadvantages of running ocean–sea ice models at such resolution.

In agreement with previous studies, our results show that a number of prominent biases and model errors persist, or even worsen, despite increases in model resolution. However, the finer resolution remains one possible way in which model capabilities can be enhanced, thanks to the explicit representation of eddies.

In Figure 13 [*figure 14 in the revised manuscript*] and Figure 14b [*figure 15b in the revised manuscript*], we present the time evolution of the Atlantic MOC and MHT at fixed latitudes for GLOB16 in comparison to available estimates. We decided to not show the ORCA1 and ORCA025 results because the interannual variability of these two metrics is not largely affected by the resolution (see figures below).

Mean values of both quantities at 26.5N are included in the text for observation and all models.

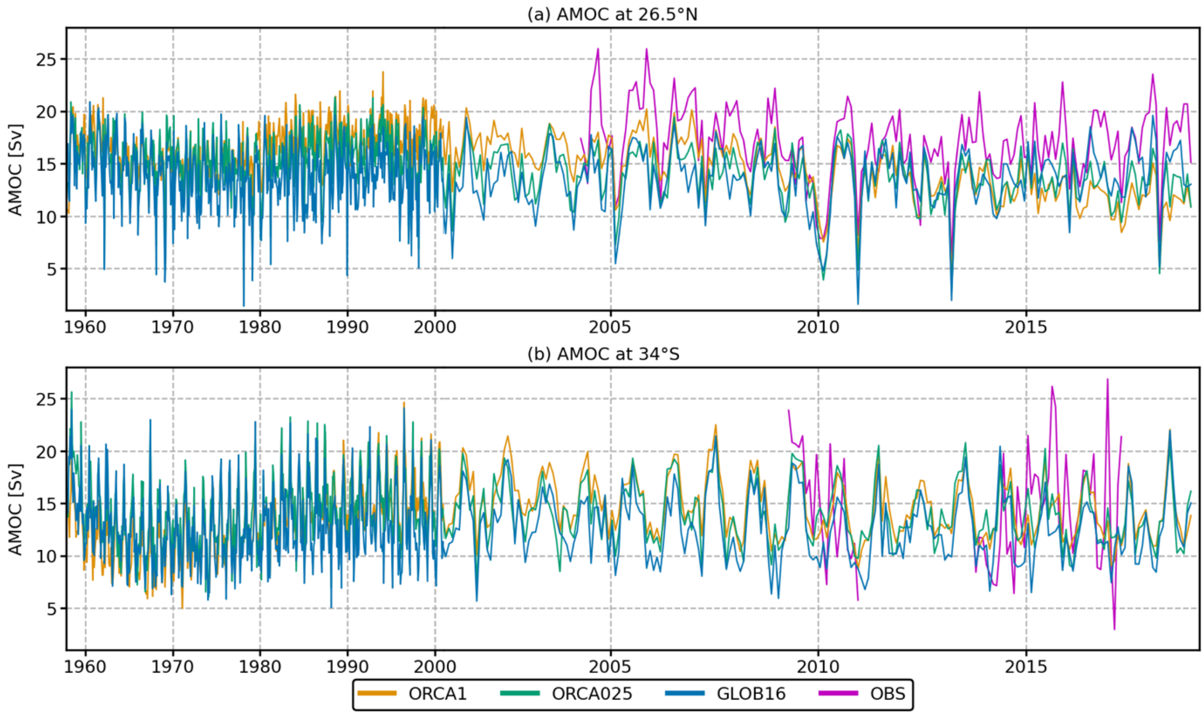


Figure RC1_4. time evolution of monthly mean AMOC transports, defined as the maximum value of the global overturning stream function in GLOB16 (blue line) ORCA025 (green line) and ORCA1 (orange line) computed (a) across 26.5°N and compared to RAPID estimates, and (b) 34°S compared to SAMBA record.

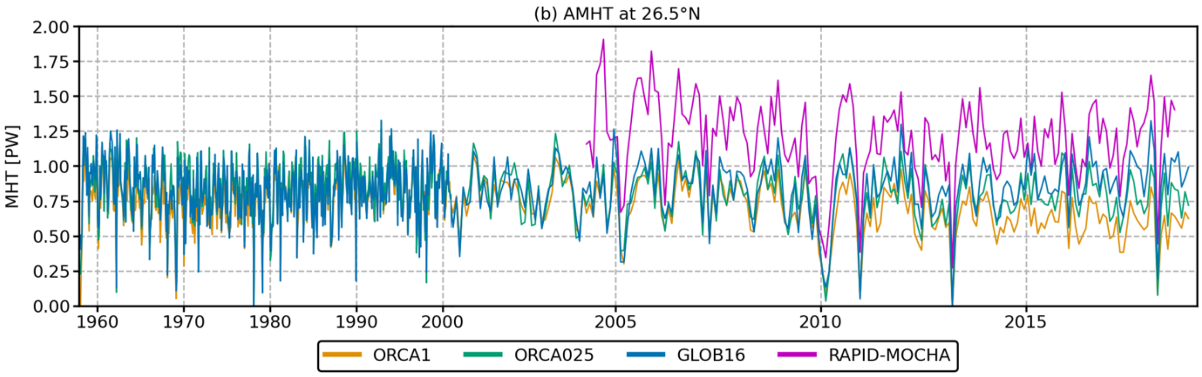


Figure RC1_5. Times series of the monthly-mean total AMHT in GLOB16 (blue line), ORCA025 (green line) and ORCA1 (orange line) across 26.5°, compared to the RAPID record (magenta).

Line 560: Treguier 2012 and Robert 2016 refs missing

References added