We thank both reviewers for their constructive and thorough comments. We focused our efforts on two main points raised by the reviewers: First, we added an analysis of the top-of-the-atmosphere (TOA) balance and show that the need for TOA tuning is smaller for the medium-resolution grids, and even smaller for the high-resolution grid. Second, to further support the analysis, we computed "performance indices" (see definition below) for the different simulations that grade the quality of the overall simulated climate (Reichler and Kim, 2008). This shows that the higher resolution grids with reduced biases in the deep ocean do not result in a degradation of the overall simulated climate, but rather come along with an improvement.

Finally, since the intent of the paper is to mainly introduce AWI's CMIP6 model configurations in GMD and to document the identified sensitivities to spatial resolution, there is, of course, a certain level of compromise in terms of what is shown here, and detailed budget analyses need to be left for future oceanographic studies in other journals.

Reviewer comments are in blue, our response is in black.

Anonymous Referee #1

Review of "Sensitivity of deep ocean biases to horizontal resolution in prototype CMIP6 simulations with AWI-CM1.0" by Rackow and Co-authors

The manuscript looks into the role of increased horizontal resolution in select regions of the ocean component of a coupled model in addressing, i.e., reducing, some of the deep ocean temperature and salinity biases in the Atlantic and Southern Ocean Basins. The authors argue that the ocean biases develop primarily from the surface, propagating along related isopycnals to the deep ocean. Higher horizontal resolution in the outcrop regions of these isopycnals appears to reduce such biases at depth. Although it is an interesting piece of work, I find the analysis rather superficial and qualitative, for example, relying on animations, rather than quantitative analysis.

We thank the reviewer for the assessment and the constructive comments. The animations were done after the analysis was performed and were only meant as additional (supplementary) information, in order to better illustrate what is going on.

I recommend major revisions along the following lines:

1. The Introduction actually introduces some physical mechanisms based on several previous studies concerning how deep temperature and salinity biases can emerge. In particular, the role of vertical mean and eddy heat transports is mentioned. Unfortunately, the manuscript does not get back to these points until the last section, and more importantly does not present a quantitative analysis exposing the role of various mechanisms. I strongly think that budget analyses should be included in the manuscript, in particular, exposing the changes in vertical eddy transports with increased horizontal resolution.

In our prototype simulations, analyzed in this manuscript, we only saved monthly mean output. Therefore, it is impossible to carry out a thorough budget analysis suggested by the reviewer. However, the final CMIP6 simulations will most likely include much more (eddy) diagnostics and output at higher frequency, so that the issue will be revisited. Moreover, work has recently started to implement more ‘online’ eddy diagnostics directly into the code of FESOM1.4’s successor “FESOM2”, which will allow for detailed budget analyses in future simulations. At this stage, we can state that the presented results (reduced drift at depth around 1000m, smaller biases in the deep ocean) are consistent with findings previously published in the literature, and our along-isopycnal analysis adds to the existing discussion.

Since we agree that further analyses would certainly make sense, we decided to add a cautionary note to the summary:
“Overall, we have shown major improvements when using medium-resolution (MR) and high-resolution (HR) meshes on representing the hydrography in the deep ocean. These grids are partly eddy-resolving and partly at most eddy-permitting, so that eddy parameterizations still need to be applied locally. This calls for dedicated in-depth analyses of eddy heat fluxes (and budgets) and their representation on multi-resolution unstructured grids in future studies.”

We also added the information about the available monthly-mean output for other readers:
“In this study, we will analyze monthly-mean output of five pre-industrial simulations over a common 100-yr period.”

2. The authors identify three regions for the source of deep biases. The first is the Strait of Gibraltar. I do not necessarily agree with the authors view that incorporation of tides will improve the representation of Mediterranean Outflow. The outflow / overflow processes require resolutions of order 10s of meters in the horizontal and meters in the vertical. Two possible solutions are an overflow parameterization and changes in the bottom / lateral topography at the outflow of the Strait of Gibraltar to minimize spurious entrainment.

We agree with the reviewer that tides are most certainly not the panacea for the representation of the Mediterranean Outflow in climate models; there is, however, a role for the later spreading of waters from the Gulf of Cadiz into the North Atlantic (Izquierdo and Mikolajewiscz, 2018). Our intention was to list one possible remedy for some of the observed differences to the simulated climate. As suggested by the reviewer, we now also added a discussion of the other approaches that are discussed in the literature to the paper, with additional references to Wu, Danabasoglu, and Large (2007; overflow parameterization) and Izquierdo and Mikolajewicz (2018):

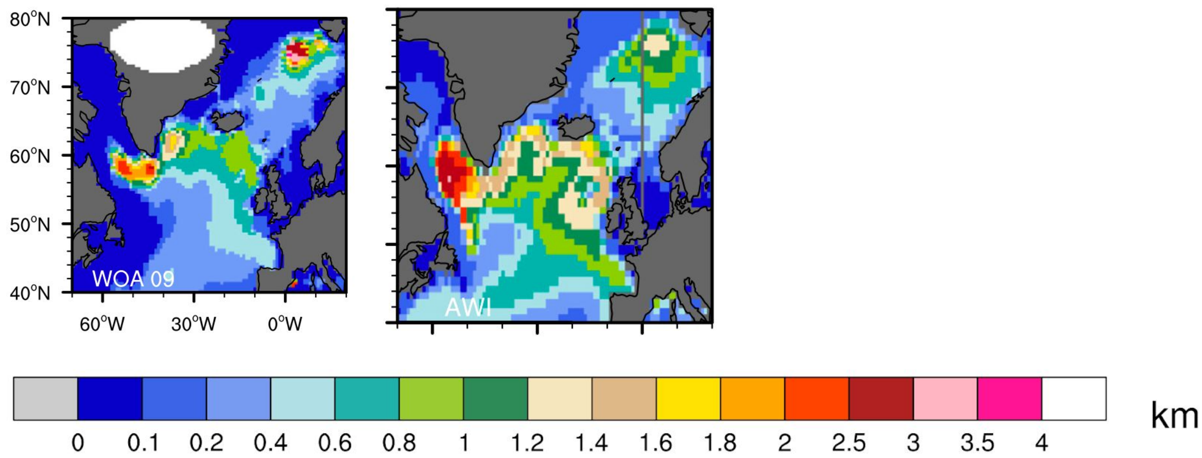
“We hypothesize that at these resolutions, smaller issues become relatively more apparent, that is other processes might need to be included for a proper simulation of the Strait of Gibraltar outflow and spreading of Mediterranean Waters into the North Atlantic. Also, resolving the overflow processes at the Strait of Gibraltar would require resolutions on the order of tens of meters in the horizontal (Izquierdo and Mikolajewicz, 2018) and meters in the vertical direction, which is still far from the resolutions applied in this study.

Two possible solutions are therefore the use of an overflow parameterization (Wu et al, 2007), which is currently not implemented in the model, or systematic changes to the bottom (and lateral) topography at the outflow of the Strait of Gibraltar to minimize spurious entrainment. In order to simulate the correct spreading of Mediterranean Waters from the Gulf of Cadiz into the North Atlantic, another approach could be to add additional physics like the effect of tides (Izquierdo et al, 2016), which are usually not included in current climate models. Without tides, ocean models often simulate erroneous south-westward spreading, leading to stronger biases when compared to climatology than in simulations with active tides (Izquierdo and Mikolajewicz, 2018).”

In the outlook, we now also mention an outflow parameterization as a possible step towards an improved representation, and that much more systematic efforts are needed to improve the representation of Gibraltar for different resolutions. In this study, the representation (width and depth) changed with increasing spatial resolution, see the plot below in an answer to reviewer #2. We did not try to keep the same geometry in Gibraltar for all the different meshes. Such future work is a necessary next step, and it is possible for us because the ocean model supports variable-resolution grids.

The second source is identified as the erroneous downwelling associated with anomalously deep mixed layers in the northeastern North Atlantic. This statement is not justified. How do you know that the downwelling is erroneous and that the mixed layer depths are anomalously deep?

This statement is based on previous findings from stand-alone ocean simulations within the coordinated ‘CORE2’ intercomparison project (Danabasoglu et al., 2014; their Figure 13), where it was shown that the mixed layer (using the “LR” grid) is anomalously deep in this area. For convenience, in the following we show the relevant part of their figure here (left: World Ocean Atlas; right: FESOM stand-alone with LR grid):



(Figure excerpt from Danabasoglu et al., 2014; https://www.sciencedirect.com/science/article/pii/S1463500313001868?via%3Dihub *)*

The MLD pattern (for MLD>500m, see green contours in supplemental animation S1 of LR) reproduces the behaviour known from the uncoupled simulation. Since the Gulf Stream in the REF and LR simulations is too zonal and reaches the northeastern North Atlantic, part of the flow has to downwell here (negative w's), which we suspect could explain part of this deficiency by entraining waters and deepening the mixed layer. We have added this information to the text:

"Since the Gulf Stream in the LR (and REF) simulations is too zonal and reaches the northeastern North Atlantic, part of the flow has to downwell here, which we suspect could explain part of this deficiency by entraining waters and deepening the mixed layer. Other factors influencing the mixed layer depth could be biased buoyancy fluxes or the restratification process via eddy activity."

The third source is presumably related to a displacement of isopycnals which are identified as too steep when eddies are parameterized. First, the analysis is not quantitative and I do not really follow the argument.

It is a classical hen-and-egg problem where you can not decide what comes first, the steepened isopycnals or the temperature biases?
The argument is that a strong warm (cold) bias dipole like in Fig.10b,left) will decrease (increase) density in the 600-1000m range. The 31.8 isopycnal (black) will therefore be lower on the northern side and higher on the southern side of this bias dipole when compared to climatology (magenta line), steepening the isopycnal. The other way round, isopycnal slope is partially controlled by eddies (or by the eddy parameterization), which tend to flatten isopycnals. If the slope diverges strongly from climatology --be it because eddies are not sufficiently resolved or because the GM coefficient is locally too small-- this should lead to similar temperature biases. Fig.10b,right) shows how a more realistic isopycnal slope in MR coincides with much smaller temperature biases. We have some control on the isopycnal slope by locally resolving eddies or, potentially, by locally tuning the GM coefficient in future simulations:
"Since we were are using a default GM coefficient for all simulations, it can be argued that a regional tuning of GM with a horizontally varying coefficient (Visbeck et al, 1997; Danabasoglu et al, 2012) could lead to a better simulation of the Southern Ocean in low-resolution AWI-CM configurations."

In MR, the isopycnal slope of the 31.8 contour between 40°S and 45°S at 10.5°E is close to climatology (about 300m per 5° of latitude), and it is more than doubled in LR (about 700m per 5° of latitude). We added this information to the text:
"Already at medium resolution (MR), the simulated isopycnal slope is about halved compared to LR and much closer to the observed slope (Fig.10b, right)".

Second, this is likely due to the issues with the details of the mesoscale eddy parameterization used. A description of the parameterization as implemented in the model should be included. Furthermore, since the REF case is much cheaper, a couple of cases with modified versions of the parameterization could be tested as alluded to in the text.

We added a basic description of the mesoscale eddy parameterization and refer for further details to Wang et al (2014):

"In order to parameterize eddies at non-eddy resolving resolutions, the Gent and McWilliams (1990) parameterization (GM) is applied with isoneutral diffusion (Redi, 1982). All prototype simulations used a reference diffusivity $K_{ref}(x,y) = 600\ m^2\ s^{-1}$, which is scaled by the local resolution (Wang et al., 2014), and a GM coefficient $K_{GM}=K_{ref}/2$. As detailed by Wang et al (2014), tapering functions following Danabasoglu and McWilliams (1995) and Large et al (1997) are also applied to K_{GM}.
Depending on the local resolution, the GM parameterization in FESOM1.4 is smoothly switched off at resolutions smaller than 25 km (red areas in Fig.2), and its effect increases linearly until 50 km, when the parameterization is fully active (Wang et al., 2014). For example, the parameterization is locally switched off when using the 'MR' and 'HR' meshes, which are locally eddy-resolving, and it is generally active in the lower-resolution 'LR' mesh (see next sections). The thresholds of 25 km and 50 km can be considered to be tuning parameters and were chosen in stand-alone simulations with FESOM1.4 using the LR grid. For the Arctic, changing the numbers can result in too diffuse boundary currents (Wang et al., 2014) and their (automatic) choice remains an important research topic for multi-resolution climate applications with AWI-CM."

There is certainly scope for an improved eddy parameterization/implementation in our model. The possible modified versions of the eddy parameterization referred to in the text (e.g. Visbeck et al) are not yet implemented in the model and will most probably be tackled in future projects.
However, as a future perspective, we strengthened the point in the discussion that by locally tuning the coefficient (e.g. by using high-res simulations as a template) we could possibly get similar answers with a low-resolution model:

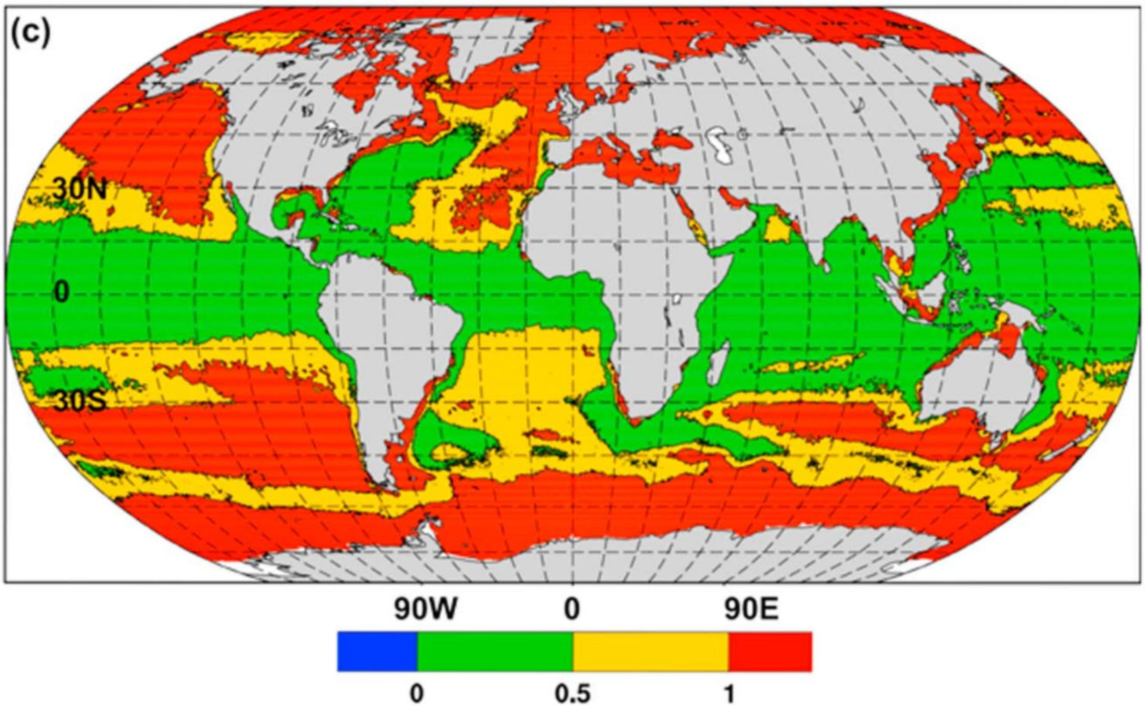
"Since we were using a default GM coefficient for all simulations, it can be argued that a regional tuning of GM with a horizontally varying coefficient (Visbeck et al. 1997, Danabasoglu et al. 2012) could lead to a better simulation of the Southern Ocean in low-resolution AWI-CM configurations. Moreover, high-resolution simulations and their effective K_{GM} could also serve as a template for the regional tuning of low-resolution simulations."

Incidentally, I am not sure what is meant by mean absolute error. Is this the root-mean-square (rms) error?

The absolute error is computed at every gridpoint as the absolute difference |T_m - T_o|, where T_o is the observed and T_m is the modeled value (e.g. potential temperature). In the end, a horizontal mean over all gridpoints is performed to get the "mean absolute error". This way, smaller values are indicative of true improvements and are not caused by compensating biases of different sign. We added an explanation of the term to the caption of Table 1.

3. The text refers to higher resolution configurations as (regionally) eddy-resolving in various places. Are they? As far as I can tell, they are still mostly eddy-permitting. A definition of what is meant by eddy-resolving and spatial maps of eddy-permitting and eddy-resolving regions for each configuration should be included. The text says “resolving the Rossby radius”, but that is not a quantitative statement.

The map for the HR grid is given in the paper by Sein et al. (2016), their Fig.4c (see figure below).



(Fig.4c in Sein et al., 2016**: Designing variable ocean model resolution based on the observed ocean variability)* ***Green areas *are eddy-resolving, e.g. over the Western Boundary Currents,* ***yellow areas*** *are eddy-permitting, e.g. in the ACC, and* ***red areas*** *need to fully parameterize the effect of eddies.*

We added this information to the text:
“For a spatial map of the eddy-permitting and eddy-resolving regions on the HR grid, please refer to Fig.4c in Sein et al. (2016).”

Following the study by Hallberg (2013), the transition between eddy-permitting and eddy-resolving grids is two grid intervals per Rossby radius, but finer resolution might still be needed to capture mesoscale eddy dynamics. We rewrote the paragraph as follows:

“Ultimately, we will target coupled configurations with a globally eddy-resolving mesh, which implies “resolving the Rossby radius” almost everywhere with at least 2 grid intervals per Rossby radius (Hallberg, 2013). Using this criterion, we have recently reported on the development of such a 'frontier' mesh (XR; see Fig.2, right

globe), with resolution capped at 4km (7km) in the Arctic (Antarctic) (Sein et al., 2017). Sein et al. (2017) note that an even finer resolution will be required locally to fully capture mesoscale eddies."

What is the physical justification for cutting of the eddy parameterization below 25 km resolution, knowing that the resolutions are mostly on the eddy-permitting side?

Please refer to the answer to reviewer #2 below. Red areas in Fig. 2 show where the GM parameterization is switched off. We added this sentence to the figure's caption.

Also, as far as I can tell, the number of vertical levels is not given in the manuscript.

The number of layers is 46 for all meshes. The levels are located at 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 115, 135, 160, 190, 230, 280, 340, 410, 490, 580, 680, 790, 910, 1040, 1180, 1330, 1500, 1700, 1920, 2150, 2400, 2650, 2900, 3150, 3400, 3650, 3900, 4150, 4400, 4650, 4900, 5150, 5400, 5650, 5900m. We added this information to the text.

4. I am unsure if all the cases represent an apples-to-apples comparison. Specifically, these are fully coupled, pre-industrial simulations. Changes in one component will undoubtedly introduce the need to retune the top-of-the-atmosphere (TOA) radiation budget. Please provide a table with the TOA values for each configuration. My point is that if the reduced bias cases show large negative TOAs in comparison to the REF case, then when the coupled model is retuned, then it is possible that the deep ocean biases will reappear.

We think it is a fair comparison since we use the same atmospheric settings for all (T127) simulations, except that we changed the ocean grid. The atmospheric parameters and settings were not tuned to a particular AWI-CM configuration; instead, they reflect typical tuned values of the sister model MPI-ESM, which uses the same atmospheric component (ECHAM6) but a different ocean model (MPIOM). We deliberately decided not to tune the TOA for the prototype simulations analyzed in our manuscript because it was our intention to isolate the sensitivity to the ocean resolution first.

As correctly pointed out by the reviewer (and stated in the last paragraph of our revised manuscript) any change to the model components will make a retuning of the TOA necessary, and this is done for the final CMIP6 control runs.

We want to emphasize again that we did not aim for a balanced TOA close to zero already after 100 years of simulation. For example, the deep ocean in a (present-day) simulation with the REF grid slowly drifts over 1500 years before it reaches a quasi-equilibrium (Sidorenko et al. (2015), Rackow et al. (2016)), and similar drifts have been reported for pre-industrial simulations. Tuning the TOA to zero at the beginning could potentially produce unwanted effects in the balanced state. The higher-resolution ocean grids in the pre-industrial configurations analyzed here appear to show much reduced drifts compared to REF (Hovmoeller plots in Fig. 4).

The global net balance at the TOA in the last 30 years of the simulations with T127 atmosphere is as follows:

LR: 1.02 W/m^2 (low-resolution)

MR: 0.89 W/m^2
MR0: 0.62 W/m^2 (medium resolutions)

HR: 0.44 W/m^2 (high resolution)

This indicates that the need for TOA tuning is smaller for the medium-resolution grids, and even smaller for the high-resolution grid.

The REF simulation (with a different atmosphere at coarser T63 resolution), which has the strongest deep ocean bias at 1000m, has a better TOA balance (0.58 W/m^2) than the LR and MR/0 simulations: it is more on the level of the medium- to high-resolution configurations with T127 atmosphere. Nevertheless, it has the strongest deep ocean bias, similar in magnitude to LR with a TOA balance of 1.02 W/m^2 (Fig.3a). This already indicates that the (global) TOA balance is not directly related the North Atlantic deep ocean biases, which are determined by localized phenomena (performance over outcropping regions, Strait of Gibraltar). Thus, a global TOA retuning is unlikely to affect the deep ocean bias significantly. We therefore added a cautious note to the text that a retuning might impact the deep ocean biases:

"Tuning could potentially affect the deep ocean simulation, although the global TOA balance in particular appears not to be directly related to the magnitude of North Atlantic deep ocean biases (not shown)."

Additionally, please include comparisons of the Atlantic meridional overturning circulation (AMOC), Labrador Sea Deep Water formation / mixed layer depth, and the northward heat and salt transports to show that the reductions in the deep biases are not occurring at the expense of degradations in several other climatically important fields.

We agree that, especially in fully coupled systems, changes or improvements in one component could potentially negatively affect other climatically relevant processes or fields. Since there is a vast number of important diagnostics (which were already discussed in the introductory papers by Sidorenko et al. (2015) and Rackow et al. (2016)) and in order to give a comprehensive answer, we extended the atmospheric "Performance Index" (PI) analysis, which was originally introduced by Reichler and Kim (2008), to ocean fields. How these indices are computed is detailed in the new Appendix B.
It basically summarizes the modelled mean climate into a score that can be quantitatively compared to other configurations and models. This modified version of the PI for important atmospheric parameters was already used in Sidorenko et al. (2015) and Rackow et al. (2016) to judge the modelled climate of AWI-CM; here we introduced the oceanic PIs for the first time.

The *atmospheric* PI for our configurations are as follows (<1 means better than the average of the considered CMIP5 models):

REF: 1.03
LR: 0.87
MR: 0.81
MR0: 0.79
HR: 0.80

The improvements, thus, roughly follow the oceanic resolution, so that bias reductions in the surface and deep ocean are not occuring at the expense of degrading atmospheric fields. Instead, the atmospheric simulation appears to improve, especially when going from T63 to T127 (REF->LR). However, the simulation of the atmosphere further improves when going from LR to the higher ocean resolutions (all with T127).

The oceanic PI are as follows:

	Global	North Atlantic	Southern Ocean
REF	0.87	0.98	0.68
LR	0.72	0.80	0.74
MR	0.64	0.62	0.62
MR0	0.61	0.57	0.62
HR	0.66	0.63	0.62

Again, while going from REF (with T63) to LR (with T127) generally improves the ocean simulation, except in the Southern Ocean-, going to even higher ocean resolutions while keeping the same atmosphere (T127) further improves the simulation.

We conclude that the bias reductions in the deep ocean do not come at the expense of degradations in the simulation of the whole system.

5. In the last paragraph of section 3.4, it is stated that "higher spatial resolution is needed . . .. to better simulate the position of the Gulf Stream." I thought that there were studies in literature showing that the high resolution is not really the silver bullet. Perhaps an expanded discussion should be included here.

We agree that we did not give a balanced discussion in the manuscript and therefore extended the discussion of the "Gulf Stream separation" topic. As was suggested by reviewer #2, we expanded the discussion as follows:

“While a strong resolution-dependence was also shown by Marzocchi et al. (2015), there are additional ways for getting a more realistic Gulf Stream separation. These include details of the numerical scheme that can affect current-topography interactions (Penduff et al. 2007) or the representation of non-local dynamics that impact the formation of a northern recirculation gyre along the North American coast, such as the Deep Western Boundary Current downstream of Cape Hatteras (Zhang and Vallis, 2007) and the cold Labrador Current northward of the Gulf Stream front (Sein et al., 2017).”

Also, I do not really follow the argument made in the last paragraph of section 3.4.1.

What we want to say here is that these biases are likely due to a northward shift of the surface isopycnals, indicating a shift of the water masses in this area; a southward shift (by flattening the isopycnal slope) could thus result in strongly reduced biases, as indeed seen in the MR simulation. We rewrote the paragraph as follows:

“Interestingly, the surface representation (SST bias) of this warm/cold interior bias to the west of Cape Agulhas and a similar dipole-like bias in the Brazil-Malvinas Confluence region are cleanly separated into their warm and cold parts by the σ_1=30.5 isopycnal surface contour (red contour in Fig.10a, left) in LR. This suggests that these biases could be caused by shifted water masses as indicated by the erroneous northward shift of the σ_1=30.5 contour, leading to a warm bias on its northern side and to a cold bias on its southern side. Flattening the slope would result in a southward shift with potentially reduced biases. Indeed, the surface biases are strongly diminished in MR (Fig.10a) with better resolved eddies and the associated flatter isopycnals, which are a close fit to the target contours from PHC (Fig.10b).”

Anonymous Referee #2

“Sensitivity of deep ocean biases to horizontal resolution in prototype CMIP6 simulations with AWI-CM1.0” by Thomas Rackow et al.

Rackow et al. are describing a hierarchy of climate model using the AWI-CM. They present the capability of the ocean model on unstructured mesh for climate application. The focus is on the benefit of using local refinement in eddy active region to decrease the deep temperature bias. In addition to that, they discuss why the high resolution decrease the bias. This leads to a discussion on the initialisation strategy of the model configuration in case no eddy parametrisation is activated. I recommend a major revision

We thank the reviewer for a thorough review and the detailed suggestions. We implemented the most pressing ones and mostly followed the suggestions for new figures, for the layout of figures and movies, and for restructuring of the text. Furthermore, the “performance index analysis” (see new Appendix B) shows that the medium- and high-resolution configurations do not degrade the overall climate (both ocean and atmosphere), but rather improve the whole simulation.

1 Major Comments:

- At the end of the paper, I am still wondering if this paper is a paper analysing possible sources of deep bias in climate model using a hierarchy of climate model with various ocean resolution or if this paper is a description on possibility open by unstructured mesh ocean model for climate application with an overview of the improvement generated by the local refinement. In the first case, the paper is maybe not adapted for GMD. In the second case, the analysis is only focused on the deep bias and nothing else. So it is not enough to convince me it is worthwhile to use this capability in a climate model for decade to century. There is no evaluation of other basic climate index as sea ice, ACC, AMOC, meridional heat transport …

To provide a comprehensive picture of the simulation quality, we added “performance indices” for both the ocean and atmosphere (see also answer to reviewer #1 and the new Appendix B). This is the first time that performance indices have been computed for an ocean model and the scores clearly highlight the improvements in the higher-resolution ocean configurations.

AWI-CM REF (with T63 atmosphere) has been extensively evaluated in two papers in Climate Dynamics (Rackow et al. 2016, Sidorenko et al. 2015), discussing all above mentioned indices and other relevant fields. While already showing that AWI-CM, in its reference configuration, is comparable to other models, the papers also highlighted the deep ocean bias as one of the most prominent issues requiring work.
In the present manuscript, which aims at improving upon this situation before our group will participate in CMIP6, we put this bias into a more general context (CMIP models show a bias like this as well) and propose a remedy via locally increased ocean resolution. We indeed focus here mainly on the deep ocean bias, but participating in CMIP6 -with model evaluation performed by the whole community- could show that this new technology is generally promising with respect to other parameters as well.

- Discussion about the contribution of Gibraltar need to be strengthen (more detailed on the geometrical issue, overflow representation and water masses properties at the Gibraltar sill)

We added a clarification on what we initially referred to as the “geometrical issue” and added a new Figure 9, which shows the spatial discretization of the Strait of Gibraltar for all the different meshes along with the local ocean depth. We also state in the revised manuscript that more systematic efforts are needed to properly tune the width and depth in low-resolution configurations. As suggested by reviewer #1, we also added references to Wu, Danabasoglu, and Large (2007; overflow parameterization) and Izquierdo and Mikolajewicz (2018; benefit of tides) as additional perspectives for further improvements.

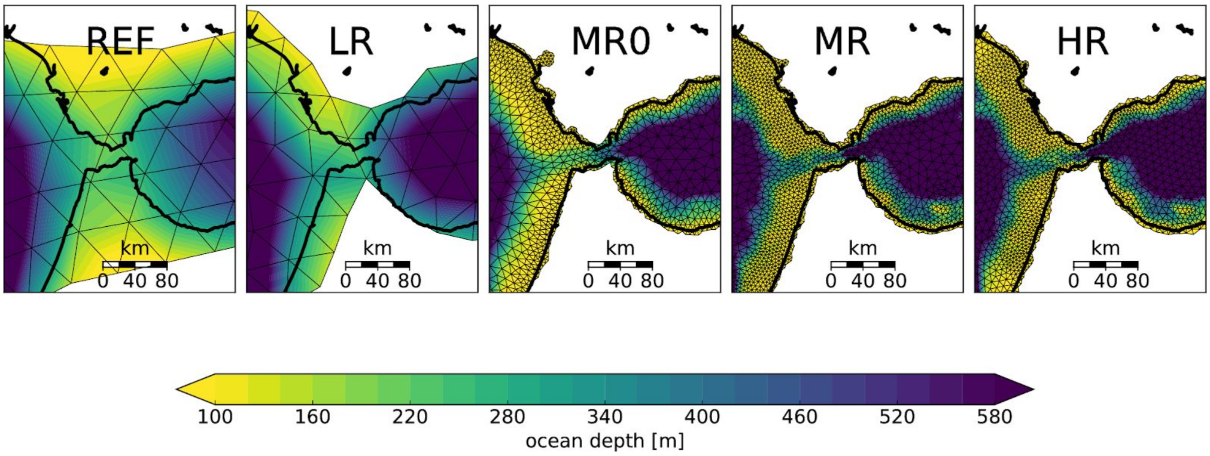


Fig.9: *Spatial discretization of the Strait of Gibraltar in the 5 different model grids. The thick black line shows the true coastline as implemented in the Basemap plotting toolbox, using data from GSHHS (http://www.soest.hawaii.edu/wessel/gshhs/gshhs.html). Triangular elements are shown with thin black lines, colors depict the local ocean depth in meters.*

Usually the bathymetry undergoes one or two steps of grid scale smoothing to make the model run stable (Wang et al., 2014). The effect of this smoothing is less evident for the higher resolution grids, since smoothing takes places on much smaller scales and therefore the resulting bathymetry is closer to the "observed" input bathymetry data. While the increased ocean resolution is expected to decrease spurious numerical mixing, the improved representation of the bathymetry to the West of the Strait of Gibraltar is expected to further add to the realism of the model and the simulated plume. However, we don't use an overflow parameterization in the current AWI-CM model yet; this point was made clear in the revised manuscript.

- In your 5 experiments, one of them do not have the same atmospheric model. The vertical profile suggest the atmospheric model resolution could also lead to strong bias reduction. For clarity, you should focus only on those having the same atmospheric model.

The reviewer is certainly right that the step from "REF" to "LR" includes simultaneous changes of both the ocean grid as well as the change from a T63 to T127 atmosphere, which was not made sufficiently clear before. We clarified this point at several places in the revised manuscript and modified the discussion accordingly. To facilitate better comparability with previous studies, where REF/T63 was introduced as a benchmark configuration of AWI-CM against which future configurations should be compared, we still decided to keep the REF configuration in the revised manuscript.

- As you mentioned a link between the deep layer and the surface via the mixed layer, you should discuss in more details what could affect the mixed layer depth intensity and location (path of the North Atlantic current, surface fresh water flux, heat flux, restratification process via eddy activity ...).

We added a discussion of these points to the text and refer the reader to the discussion and Fig.13 of Danabasoglu et al (2014) (see the relevant part of their figure above in an answer to reviewer #1):

"This is a feature that has already been identified in uncoupled FESOM simulations using the LR grid as part of the CORE-II intercomparison project (Danabasoglu et al, 2014; their Fig.13). Since the Gulf Stream in the LR (and REF) simulations is too zonal and reaches the northeastern North Atlantic, part of the flow has to downwell here, which we suspect could explain part of this deficiency by entraining waters and deepening the mixed layer. Other factors influencing the mixed layer depth could be biased buoyancy fluxes or the restratification process via eddy activity."

About the overall idea of the initialisation strategy, I found it interesting. As it is included in the result section, I think you have to try it and show result on the initial bias in the HR case. You mentioned the GM eddy parametrisation is the key in LR to avoid 'overshoot' of the bias because it is fully active from the start in LR. Why not run in parallel to your idea on 3d T/S restoring during the spin up of the eddy fields (something like GM fully active from the start with a decreasing intensity over a specific time scale).

We thank the reviewer for the positive feedback. This is an immediate idea based on our results, and that's why we decided to mention this idea directly at the end of the results section in the submitted manuscript. This is beyond the scope of this paper, but we are glad that the reviewer agrees with the relevance of testing different initialisation techniques and their potential for improving climate simulations. Since the 3D T/S restoring and other techniques still need to be implemented and tested carefully, they need to be left for future studies.

The caveat with using GM fully active from the start in high-resolution simulations, but with a decreasing intensity over the timescale when the explicitly resolved eddy field is developing, is that - in our experience - GM tends to damp a significant part of the *resolved* eddies because the parameterization decreases baroclinicity. However, we agree that this is a plausible conjecture that would be worth testing in the future.

In Minor comments, I went through the manuscript from the beginning. Some comments are related to the one mentioned above.

2 Minor Comments:

Abstract

- P1L6: 'we find that two major sources at the surface are responsible for the deep bias in the deep Atlantic': Please briefly mention these 2 mechanisms.

We extended the sentence in the abstract accordingly:

“We find that two key regions at the surface are responsible for the development of the deep bias in the Atlantic Ocean, the north-eastern North Atlantic and the region adjacent to the Strait of Gibraltar.”

Introduction

- P1L21: You mentioned a major bias is present in CMIP5. Could you add references to it in addition to your illustrations?

Although being substantial in magnitude, to our surprise the deep ocean biases in the CMIP5 models did not receive a lot of attention yet. There is a paper for one specific model (EC-Earth) by Sterl et al. (2012), where maps of temperature and salinity biases in 1000m are shown. We are not aware of other papers discussing the CMIP5 bias specifically, so we only added the Sterl et al. (2012) paper:
“A major bias present in CMIP5 models is reflected by a too warm and saline deep ocean compared to observations (e.g. in the EC-Earth model; Sterl et al. 2012).”

- P2L2: You should reformulate “..., as well as climate change (...) that is, errors are larger ...” It is not easy to understand.

We rephrased the sentence as follows:
“Importantly, the mean absolute error in deeper ocean layers is larger than the interannual variability (the standard deviation of annual means). It is also larger than the climate change signal as determined from RCP8.5 and RCP4.5 emission scenarios. Or formulated differently, deep ocean biases are larger than the signals we aim to predict, which may be cause for concern in non-linear systems.“

- P3L3: This is the first time in the main text you are using AWI-CM acronym, I think you should defined it here.

Done. We still mention in the model configuration section that AWI-CM was previously named “ECHAM6-FESOM”, with citations of Sidorenko et al. (2015) and Rackow et al. (2016).

- P3L11 and elsewhere: Be careful when using ‘eddy resolving’ term. I am not convince you are, even in the location reddish in your figure 2. You should precise where you are eddy resolving or permitting. In introduction, I can suggest something like ‘... a strong case to aim for a high resolution (X km or higher) in eddy active region ...’

For the HR grid (10--60 km resolution), we now refer the reader to a map in Sein et al. (2016) where the eddy-resolving and eddy-permitting regions are shown. The MR grid, used for the standard CMIP6 cases, uses even higher resolution locally over the Gulf Stream area (Fig.2). As suggested by the reviewer, we therefore changed the text as follows:
“[...] a strong case to aim for a high resolution (10 km and higher) in eddy-active regions not only in HighResMIP (Haarsma et al., 2016), but already for AWI's CMIP6 standard configuration.”

Model configuration

- P3L18: just mention the acronym here as you explain it before (see comments above)

Done. We now mention the acronym here (and the previous name ECHAM6-FESOM; see comment above).

- About GM details, I am sure that how to define the location where you activate GM and how to make the transition from 'off' to 'fully active' trigger a lot of discussion in your group. My question is: should it be dependant of the Rossby Radius instead of prescribed resolution threshold (25km and 50km)? At 25 km a lot of eddy active region are still not eddy resolving. Could you explain more why you choose these numbers (25 and 50), what are the sensitivity of your ocean model to these numbers?

At mid-latitudes, the Rossby radius is between 25 and 50km, which is why this initial simple choice was made. Still, 25km and 50km can be considered to be tuning parameters and where chosen in stand-alone simulations with FESOM1.4 (Wang et al., 2014), and those numbers worked well in practice and gave the best results (using the LR grid). For the Arctic, regarding sensitivities, changing the numbers can result in too diffuse boundary currents (Wang et al., 2014).

We agree that the criterion should be Rossby radius-dependent, and it already is in FESOM1.4's successor FESOM2. However, even in the formulation based on the Rossby radius, tuning is required because the scales of baroclinic instability do not only depend on the Rossby radius, but also on details of the potential vorticity gradient profile and the surface buoyancy gradient.

In summary, these numbers are currently considered to be tuning parameters and how to switch between "on" and "off" GM regions will remain an important research topic for multi-resolution applications with FESOM (Wang et al., 2014).

We added the following sentences to the text:
"At mid-latitudes, the Rossby radius is between 25 and 50 km, which is why this simple choice was made. Still, the thresholds of 25km and 50km can be considered to be tuning parameters and were chosen in stand-alone simulations with FESOM1.4 using the LR grid. For the Arctic, changing the numbers can result in too diffuse boundary currents (Wang et al, 2014) and their (automatic) choice remains an important research topic for multi-resolution climate applications."

"Depending on the local resolution, the parameterization is smoothly switched off at resolutions smaller than 25 km (red areas in Fig.2)"

We added "The GM parameterization is switched off within the red areas." to Fig. 2's caption.

- You should specify also in your model configuration

o Your input data for the bathymetry

The bathymetry in the model is based on a blend of several bottom topography data sets, as detailed in Wang et al (2014): North of 69°N, the International Bathymetric Chart of the Arctic Oceans (IBCAO version 2, Jakobsson et al., 2008) is used (2km resolution). South of 64°N, the General Bathymetric Chart of the Oceans (GEBCO) is used (1min resolution). The topography is computed as a linear combination of the two charts between 64°N and 69°N.

We added:
"The bathymetry in the different grids is based on a blend of the IBCAO (Jacobsson et al, 2008) and GEBCO (...) bottom topography data sets, as detailed by Wang et al. (2014)."

o Your vertical coordinate system and number of vertical level and resolution range

Done. We added information on the vertical levels to the text, both the number (46) and the exact levels.

o If you are using some icebergs representation, how do you represent iceberg (iceberg model or prescribed pattern, melt set in surface or spread between surface and iceberg draft depth) and how you compute its calving rate.

The model did neither use interactive icebergs nor spatio-temporal templates representative of iceberg drift for the distribution of land ice over the ocean. The runoff and calving from land was put into the ocean in a narrow band around Greenland and Antarctica, as is the default practice in the model.

- P4L1 : try to avoid pages with figures, tables and with only a few lines of text at the bottom. It is really easy to miss these lines.

Done. We removed single lines of text below figures 4 and 5 and below Table 2.

- About the XR resolution, you should just mention it in the conclusion as perspective and remove reference to it. In the main text, I found it not useful, as you do not show and discuss any result from this configuration.

The motivation for the medium-resolution grids is that we can run simulations with spatial resolution smaller than 10km in eddy-active regions (MR0/MR/HR) already today, even though we cannot easily run the global XR resolution over climate-relevant timescales yet. As XR is the future goal and constitutes the motivation for the design of our current medium-resolution meshes, we'd like to keep it in the main text.

Results
- P6L30: The figure 3 do not represent a drift. So please reformulate.

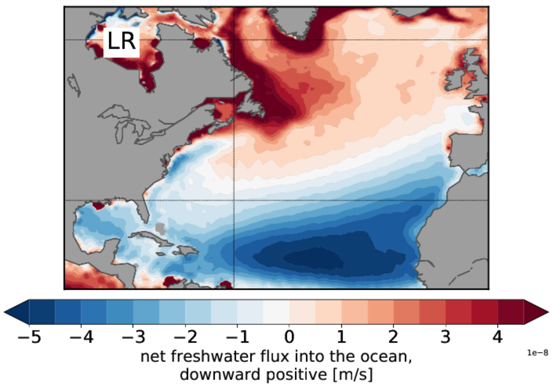
Done. We reformulated the sentence as follows:

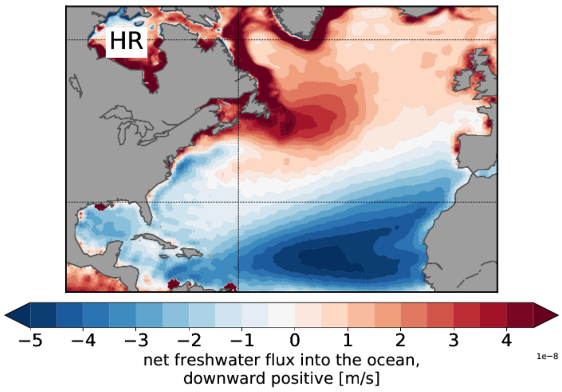
“Temperature and salinity show major improvements for medium- and high-resolution configurations, as seen from horizontally averaged temperature and salinity profiles […]”

- *About the S profile there is some differences which seems not related to resolution:*

o Surface salinity error are from- 0.2 to 0.2 without clear resolution dependence.So as it is a couple run, if you change your atmospheric resolution (REF vs LR) or you oceanic resolution (LR, MR, MR0 and HR), your surface fresh water forcing can change. So, I am wondering if your surface fresh water forcing in all your run is similar. As you discuss impact of mixed layer depth on error in depth, I think it is quite important for the discussion in section 3.4.

We have checked the net freshwater flux into the North Atlantic ocean (years 71-100) and we could not identify obvious strong differences. As an example, here is the freshwater flux for LR (approx. -0.2 surface salinity error) and HR (approx. +0.2 surface salinity error):





o In depth (deeper than 1500m) the resolution of the atmospheric model seems to play a big role in it. All the model using T127 atmospheric model have the same error. It is less clear in temperature but it still looks significant deeper than 2000m. You should add discussion on it or maybe remove REF simulation from the paper.

We added the following paragraph to the manuscript:

“The simultaneous change of the ocean and atmospheric resolution from REF/T63 to LR/T127 leads to a clear improvement of the salinity profiles below 1500m, and all configurations with T127 atmosphere (LR, MR0, MR, and HR) share a very similar salinity bias in this range. While it is difficult to say what the relative influence between the atmospheric resolution change (T63 vs T127) and the switch of the ocean grid (REF vs LR) is, it appears that surface conditions can significantly impact deep ocean biases.“

- *P7L5: By stronger deep cell, what do you mean? do you mean deep overturning cell?*

Yes, “deep cell” was short for “deep overturning cell”. We corrected this in the revised document.

- *All the discussion about Gibraltar:*

o Could you add precision about the geometric error in your configuration (i.e. model strait width compare to reality)?

We added a scale bar to Figure 9 to see the model strait width (in reality, it is about 15km).

o In Figure 6, we clearly see that the salinity in depth is much more saline than the observations. What is the quality of the water masses going out of the Med. Sea at Gibraltar? Does it impact your analysis?
o Gibraltar is a shallow sill and the connection with the deep layer of the ocean is made via cascading of the dense water (Gibraltar overflow). However, the modelisation of this process is quite challenging in ocean model. So, is the Gibraltar overflow well represented in yours simulations? If no, what are the impact of it on your simulations and sensitivity. You should mention the Med. overflow in your discussion, its representation in FESOM and its importance compare to the geometrical factor you mentioned.

For further discussion about the Strait of Gibraltar, please see our answer to the reviewer's major comment above (and answer to reviewer #1). We do not use an overflow parameterisation yet, this information was also added to the text.

- About the discussion in surface conditions:
o See comments earlier on fwf
o P8L18: 'no heat sources': Could you precise if you are using a geothermal heating. If yes, maybe reformulate the first sentence.

We are not using geothermal heating in our simulations. We added the sentence "*The ocean model also does not apply geothermal heating as lower boundary condition (e.g., Adcroft et al. 2001; Downes et al. 2016).*"

o Could you mention the effect of the contribution of the advection from the other basin into your analysis domain. Gulf Stream and NWC: There is many modelling paper reporting issue in modelling these area, discussing the possible reason for it and the impact on the large scale. You should not only mention resolution as possible reason. You can mention for example the numerical scheme used (penduff et al., 2007: https://www.ocean-sci.net/3/509/2007/os-3-509-2007.pdf), or the representation of the DWBC (Zhang and Vallis, 2007: https://journals.ametsoc.org/doi/10.1175/JPO3102.1). Resolution dependence is also visible in Marzocchi et al., 2015:https://www.sciencedirect.com/science/article/pii/S0924796314002437#f0010)

We thank the reviewer for pointing us to these papers. We added all the suggested references to the text. We now also explicitly mention the representation of important non-local dynamics (DWBC downstream of Cape Hatteras, impacting the formation of a cyclonic northern recirculation gyre, Zhang and Vallis 2007; cold Labrador current that meets the warm Gulf Stream, Sein et al. 2017):

"While a strong resolution-dependence was also shown by Marzocchi et al. (2015), there are additional ways for getting a more realistic Gulf Stream separation. These include details of

the numerical scheme that can affect current-topography interactions (Penduff et al., 2007) or the representation of non-local dynamics that impact the formation of a northern recirculation gyre along the North American coast, such as the Deep Western Boundary Current downstream of Cape Hatteras (Zhang and Vallis, 2007) and the cold Labrador Current northward of the Gulf Stream front (Sein et al., 2017).”

o P10L2 please precise ‘This region (hatched in Fig. 7). Do you mean the difficulty to simulate a correct NWC and GS ?

We replaced “This region...” with “The cold temperature spot…”. Yes, the cold spot is impacted by the difficulties to simulate a correct North West Corner and Gulf Stream.

o You mention that the issue with the Gulf Stream and NWC is not in direct contact with the outcropping isopycnals you are interested in but the representation of the GS and NWC strongly impact the North Atlantic Current which reach the latitude you are interested in. So it could be the location of the outcropping region is determined by the path of the NAC. Could you add discussion about this.

We extended the paragraph as follows:
“Although the Gulf Stream and its extension could impact the location of the outcropping regions, the strong cold temperature spot (hatched in Fig.7b--d) is, however, not in direct contact with the deep ocean around 600--1000m depth via outcropping isopycnals and thus does not limit the analysis of the present manuscript, which is focused on the deep ocean.”

- About the along-isopycnal bias propagation:
o See comments about Gibraltar above
o For the mixed layer source, see comments about surface fwf above. About the realism of the >500m convection, could you show comparison with observation or at least reference showing what the mixed layer depth should be.

Please see the above plot in an answer to reviewer #1, where we have shown the relevant part of Fig.13 in Danabasoglu et al. (2014), with a reference and an ocean stand-alone simulation using the LR grid. We added the following text:

*“This is a feature that has already been identified in uncoupled FESOM simulations using the LR grid as part of the CORE-II intercomparison project (Danabasoglu et al, 2014; their Fig.13). Since the Gulf Stream in the LR (and REF) simulations is too zonal and reaches the northeastern North Atlantic, part of the flow has to downwell here, which we suspect could explain part of this deficiency by entraining waters and deepening the mixed layer.
Other factors influencing the mixed layer depth could be biased buoyancy fluxes or the restratification process via eddy activity.”*

o As you are talking about deep bias, I think is is worth adding discussion about the Nordic Sill overflow. Is the representation of the Nordic sill in your various configuration affect your conclusions?

Although it could impact deep ocean biases as well, we have not analyzed the Nordic Sill overflow in this study. Similar to what we have written about the Mediterranean overflow, it would require much more systematic efforts. In this study, we used available ocean grids that were based on different mesh design principles (see section 2.2) and the representation of the Nordic Sill overflow was only one of many target quantities.

o 4 supplementary documents in half a page of discussion I found it too much. Could you find a way to represent the point you want to make in a figure? Often reader like me do not take the time to get back on their browser, find the link, click on it and watch 4 movies.

In order to keep a reasonable number of figures, we only added one new figure for illustrating the story about the Southern Ocean bias development, which was difficult to see in the movies (see Fig.11 below).

However, we agree that 4 supplementary documents can be rather overwhelming and we reduced the number of referenced supplementary movies to two in this subsection. Since the propagation of the bias along relevant isopycnals is already shown is Figure 8 (second and last row for LR and HR, respectively), the revised subsection now only refers to Fig. 8 (rather than to movies S3 and S4 as well). In the caption to Fig 8, the interested reader will still find the link to the animated versions.

o You focus on the large improvement between LR and HR, but I found that there is also a large improvement between REF and LR (it let suggest also that the atmospheric model resolution is also important in decreasing the bias in depth.). See comments above on maybe removing REF from the document as LR and REF has roughly the same resolution.
o Please reformulate the conclusion of this section based on the comments above.

In contrast to the improvements seen in the salinity profiles between REF and LR below 1500m, which we state now in the revised manuscript, we cannot identify a large improvement in potential temperature sections for years 91-100 between the LR and REF simulations (Figure 8). However, we added a sentence why we focus on LR and HR in this subsection:
"To isolate the influence of the chosen ocean grid using the same atmospheric T127 configuration, we will focus on the LR and HR configurations here as examples."

o About the SST bias you mention at the end, please mention a reference to a figure.

Done, we now refer to Fig.7b--d.

- Displacement and tilt of isopycnal:
o You explain why the slope of the isopycnal is different but I think you should add clearly, why this leads to temperature bias along the isopycnals?

Please refer to the answer to reviewer #1 above (hen-and-egg problem).

o All your paper is focussing onto the depth 1000m. So I suggest for clarity to remove the discussion on the 200-300m depth range P15L7 to L14.

Please see the answer to reviewer #1, where we rephrased this paragraph.

o In your supplementary materials we clearly see in the LR case an error propagating from the Good Hope cap toward south America. Do you know why this propagation and not a bias intensifying all along the Atlantic Southern ocean?

We don't think that the behaviour in the Southern Ocean is similar to what we wrote about the error propagation in the North Atlantic. In the Southern Ocean, although it looks like an error propagation in the supplementary material, it could rather be a bias that develops uniformly along the whole Atlantic Southern Ocean, depending on how eddies are treated and how the GM parameterisation is applied. We checked that the term "propagation" is just used for the North Atlantic case in the manuscript.

o You mentioned that this strong bias in the Atlantic is due to difficulty of GM to balance the Ekman transport. So, why the error is so large in the Atlantic sector only? The other sector are quite good in LR and REF compare to HR.

The Atlantic sector of the Southern Ocean is the place where the waters of North Atlantic origin upwell (governed by a balance between westerly winds and by eddies, which try to flatten the slope of isopycnals), and part of it is returned back to form the AMOC. One could say that this crucial process distinguishes the Atlantic sector from the Pacific sector, since the Pacific MOC does not show the same behaviour as known from the AMOC streamfunction. Biases in isopycnal slope in the Atlantic sector will be seen as temperature biases, as explained in an answer to reviewer #1 above.

- Initialisation method:
o P15L17:I found the mention of 'usually based on a smoothed climatology as done in this study' confusing. I suggest to remove it. If you effectively smoothed the climatology, mention it in the previous sentenced and in the model configuration section.

We did not smooth the climatology; we wanted to say that climatologies used for the initialization of ocean models are typically rather smooth. We changed "smoothed" to "rather smooth".

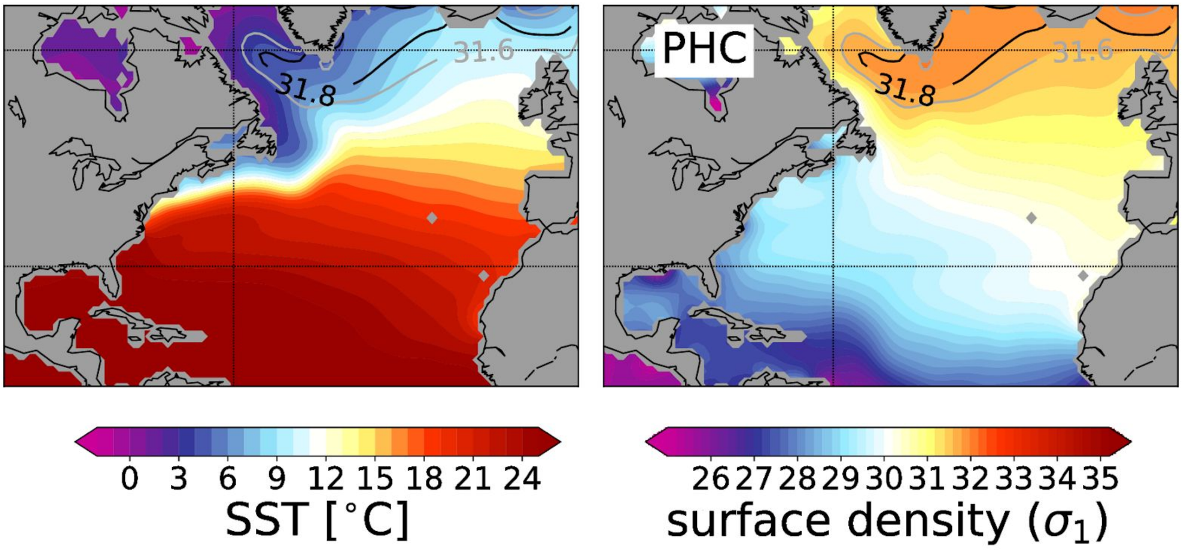
o P15L19: what is the time scale you imply exactly by 'fast' adjustment? days? months? years?

We imply a time scale of a couple of months, up to a year. We added:
"After this first phase of fast adjustment, which takes months to one year …"

o About the example you mention (bias in the east North America), I will be more cautious. I agree that if the Gulf Stream is to north, you will have a warm bias in the Northern

Recirculation Gyre but based on the information you show, we don't know if PHC is representing this coastal area with strong boundary current correctly (you have strong temperature front in this area). You should at least put PHC sst in Figure 7.

The strong boundary current and its temperature front is clearly seen in the PHC data set as plotted below. As suggested by the reviewer, we added this panel to Figure 7 and rewrote the caption accordingly.



o Could you add precision about the time needed for the eddy fields to develop in your configuration?

The time needed for the development is about 20-30 years, as we have seen in previous simulations with higher-frequency output. This fits the timescale cited in the manuscript (Allison et al., 2010).

o P15L28: I think you should add a specific plot to show this instead of claiming it 'evident' on a supplementary material video. I had to watch back and forth frame by frame to be convince.

As suggested by the reviewer, we added a new Figure 11 which shows the relevant frames of the supplementary movies.

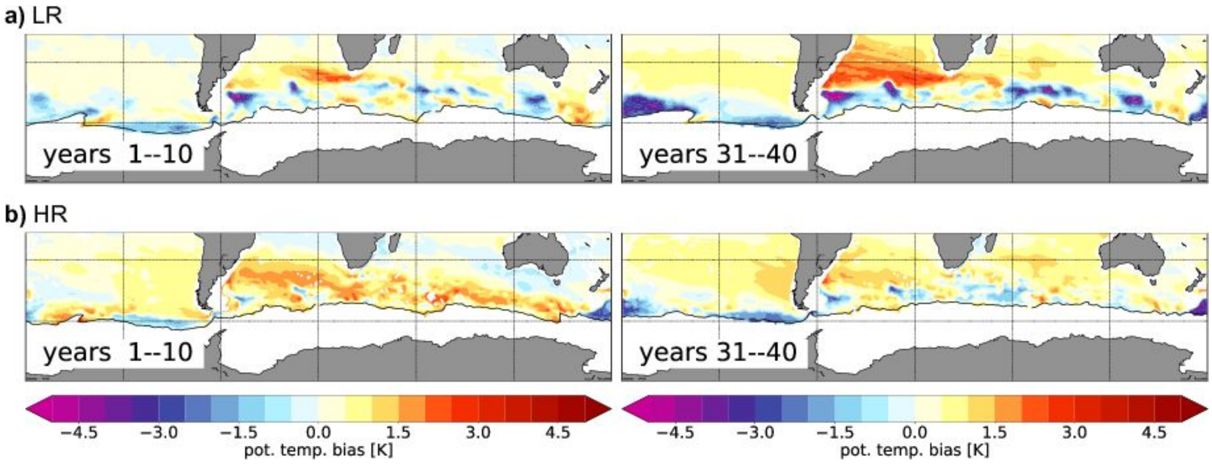


Fig.11: *Southern Ocean maps of along-isopycnal potential temperature biases in* ***a)*** *LR and* ***b)*** *HR for years 1--10 and 31--40. Black contours show the outcropping location of the* σ_1*=31.8 isopycnal. For animations of the bias development with a 10-yr running window, see supplementary animations S1 and S2 (Rackow et al., 2018b).*

We now refer to the new figure in the main text (rather than to the movies) and the interested reader can still have a look at the animations (mentioned in the caption of Figure 11): *"Interestingly, in HR, an initial movement of the 31.8 isopycnal surface contour in the Southern Ocean towards the equator apparently leads to larger initial biases than in LR (Fig.11, left), and then it returns back to the south after 20 or 30 years. In years 31--40, the biases seem to recover and are again smaller than in LR (Fig.11, right)."*

- Conclusions:
o In the model configuration section and introduction, you insist a lot on the local resolution, its benefice to run climate model. I was expecting it to be mentioned at the beginning and in a stronger way than you did.

We agree and now mention this capability in the first paragraph of the Conclusions: *"We found that the deep bias seen in AWI-CM-LR and REF is systematically reduced when moving to successively higher resolutions (10 km and higher) in eddy-active regions, a capability supported by FESOM1.4's use of multi-resolution ocean grids. Although there is certainly scope for improved eddy parameterizations, our results thus highlight the benefit of using high-resolution ocean components in climate modelling."*

o I found the word 'the three worst performing CMIP5 model' not well chosen here without mentioning the criterion used for the assessment.

This was certainly an unfortunate formulation without mentioning the specific criterion. The formulation has been changed to *"However, we could not identify a clear dependence of deep ocean biases on the vertical mixing schemes used in CMIP5 models: the three models with the strongest absolute error at a depth of 1000 m (GISS-E2-R, MPI-ESM-LR, GFDL-CM3; ...) use either KPP or PP mixing".*

o Rewrite the discussion on Gibraltar based on the comments on the overflow and

Med. Sea water property.

We extended the discussion stating that much more systematic efforts are still required to study the Mediterranean outflow, and that --besides the shown sensitivity to resolution-- an overflow parameterization might be a necessary next step to improve the model performance:

"We identified two major sources for the deep ocean biases in the Atlantic ocean. The first source is the Strait of Gibraltar, which is likely to be a geometric issue related to the spatial discretization of this narrow strait (15 km) at relatively coarse resolution that is typical for CMIP5 models (about 100 km), and that often leads to increased Mediterranean outflow (e.g. Sterl et al., 2012). Much more systematic efforts are required to tune the horizontal and at the same time the vertical discretization of the Strait of Gibraltar. [...] We suspect that [...] the addition of an overflow parameterization might be necessary steps to further improve the model performance."

o P17L33: By 'outcropping often happens too far to the north compared to observations', please clearly specify what you imply? Do you imply that isopycnals outcrop in a region with stronger heat fluxes, warmer atmosphere ...?

We want to point out that by changing the outcropping position to the North, denser water masses will be in contact with atmospheric conditions (fluxes) that are usually in contact with lighter waters. This shift will typically be accompanied by biases in (surface) temperature, for example, as mentioned earlier. We added:

"Thus, outcropping often happens too far to the north compared to observations, so that denser water masses will be in contact with atmospheric conditions (fluxes) that are usually in contact with lighter waters, which can impact water mass transformation."

o Most of your paper is on the deep bias and you mentioned an example of bias developing at 200m depth. As I mentioned earlier, to keep your paper focus you should maybe get rid of the paragraph discussing this.

Please see our earlier reply.

o P18L5 to L13: You should move this paragraph earlier in the conclusion, maybe at the beginning.

We thank the reviewer for this suggestion. Indeed, the conclusion reads more natural now that this is the second paragraph, right after we highlight the multi-resolution capability of AWI-CM (as also suggested by the reviewer).

o P18L29: 'we have shown major improvement'. You need to add limitation to this statement. You only show major improvement on the T/S bias at 1000m. We don't know at all if it improve the MOC, MHT, bottom water formation.

We changed the statement to:

“Overall, we have shown major improvements when using medium-resolution (MR) and high-resolution (HR) meshes on representing the hydrography in the deep ocean around 1000m. These improvements at depth do not come at the expense of degradations in other climatically relevant fields, as shown by a performance index analysis (Appendix B).”

3 *Figure and table comments*

*- **Fig. 1**:*
o replace left/right by ‘a)’ and ‘b)’ and add it on the figure
o Comments on what you should see ‘In the first hundreds meter ...’ should go into the text not in the caption.
o Mean abs. error in the top 300m is hard to see (overwritten by blue and red line), maybe consider using transparency and envelope.

We added labels to the figure and moved part of the comment on what can be seen to the text. By changing the order of plotting, we were able to increase the visibility of the mean absolute errors in the top 300m (black, plotted last).

***-Fig. 2**: remove XR if you follow my comments on removing XR from the text.*

See comments above.

-Fig. 3:
o As for Fig. 1, replace left/right by ‘a)’ and ‘b)’ and add it on the figure
o Comments on what you should see ‘With the medium- and ...’ should go into the text not in the caption.

We added the labels to the figure and changed the caption accordingly. Although the comment is already given in the text, we’d like to repeat this statement here for readers who quickly scan papers, because it is the key message of the figure and a key message of the paper.

***-Fig. 6**: split left column from the right column and put a label for each figure and use it in the caption.*

As suggested by the reviewer, we split the figure into a left and a right column and added the labels a) and b) to them.

***-Fig. 7**: add a label for each figure and add PHC sst figure and maybe use the same colorbar as in figure 6.*

Done. We also added the SST plot for SST in Fig.7a. While a) and e) now show the SST and surface density structure in PHC, b)-d) still show the SST bias in the 3 simulations with respect to PHC and f)-h) show the surface density structures in the simulations. We kept the colorbar as it was because the surface bias plots did not improve with the smaller range typical for the biases at depth in Figure 6.

*- **Fig. 8**: remove red line, as they are not commented on this figure.*

We still decided to keep the red lines in the figure for better orientation, and for consistency between figures 8 and 10.

*-**Fig. 9**: If you remove discussion on 30.8 and 30.5 isopycnal line, do not forgot to remove it here.*

We kept the lighter isopycnals (colored red) as mentioned in an answer above.

You are not commented the green line in this figure, so please remove it.

We refer to the green line in the manuscript: *"The mixed layer (green line and shading in Fig.8) is deep enough so that surface biases can reach the 31.8 and neighboring isopycnals, from where the signal is further advected towards the south. "*

*- **Supplementary movies** : please add a date on each frame, so we know where we are when we look at it (discussion on initial condition)*

Done. We added a time axis to Version 2 of the movies that are archived at Zenodo (https://doi.org/10.5281/zenodo.1323333). The DOI resolves to the latest version of the movies.

*- **Table 1**: Add interannual std and climate change signal in the top or bottom cells.*

We added the interannual std. dev. as well as climate change signals (RCP4.5 and 8.5) for the CMIP5 models in three new columns. We thank the reviewer for this suggestion because it revealed an interesting connection. The strength of the climate change signal appears to be somewhat linked to the quality of the ocean mean state (the magnitude of the error): models with smaller errors tend to simulate less pronounced climate change in the global ocean while models with larger errors in the mean state tend to simulate a stronger climate change in the ocean.

We added this preliminary finding to the introduction: *"[...] the magnitude of the projected climate change in the ocean appears to be ordered according to the models' mean absolute errors in the ocean (Table 1)."*

*- **Table 2**: Remove XR line and remove the internal name (not used in the manuscript).*

We'd like to keep the name because it guarantees better comparability with previous papers from our team, where names like "CORE" or "GLOB" were used. The renaming to "LR", "MR", "HR" and so on is only a relatively recent development in the course of our institute's participation in CMIP6. We therefore changed “internal name” to “previous name” in the table. Why we wish to keep XR in the manuscript has been discussed in the above answers.

Sensitivity of deep ocean biases to horizontal resolution in prototype CMIP6 simulations with AWI-CM1.0

Thomas Rackow[1], Dmitry Sein[1], Tido Semmler[1], Sergey Danilov[1], Nikolay Koldunov[1,2], Dmitry Sidorenko[1], Qiang Wang[1], and Thomas Jung[1,3]

[1]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany
[2]MARUM, University of Bremen, Bremen, Germany
[3]Institute of Environmental Physics, University of Bremen, Bremen, Germany

Correspondence to: Thomas Rackow (thomas.rackow@awi.de)

Abstract. CMIP5 models show substantial biases in the deep ocean that are larger than the level of natural variability and the response to enhanced greenhouse gas concentrations. Here we analyse the influence of horizontal resolution in a hierarchy of five multi-resolution simulations with the AWI Climate Model (AWI-CM), which employs a sea ice-ocean model component formulated on unstructured meshes. The ocean grid sizes considered range from a nominal resolution of $\sim 1^\circ$ (CMIP5-type) up to locally eddy-resolving. We show that increasing ocean resolution locally to resolve ocean eddies leads to a major reduction in deep ocean biases. A detailed diagnosis of the simulations allows to identify the origins of the biases. We find that two ~~major sources~~ key regions at the surface are responsible for the development of the deep bias in the Atlantic Ocean, the north-eastern North Atlantic and the region adjacent to the Strait of Gibraltar. Furthermore, the Southern Ocean density structure is equally improved with locally explicitly resolved eddies compared to parameterized eddies. Part of the bias reduction can be traced back towards improved surface biases over outcropping regions, which are in contact with deeper ocean layers along isopycnal surfaces. Our prototype simulations provide guidance for the optimal choice of ocean grids for AWI-CM to be used in the final runs for phase 6 of the 'Coupled Model Intercomparison Project' (CMIP6) and for the related flagship simulations in the 'High Resolution Model Intercomparison Project' (HighResMIP). Quite remarkably, retaining resolution only in areas of high eddy activity along with excellent scalability characteristics of the unstructured-mesh sea ice-ocean model enables us to perform the multi-centennial climate simulations needed in a CMIP context at (locally) eddy-resolving resolution with a throughput of 5–6 simulated years per day.

1 Introduction

Biases at the ocean surface are relatively well studied (e.g. Wang et al., 2014a). However, climate models also suffer from less known biases in the deep ocean that have the potential to impact the storage of heat by the ocean. This issue may be of relevance for projections of the future climate performed in the framework of the Coupled Model Intercomparison Project (CMIP; Taylor et al., 2012).

A major bias present in CMIP5 models is reflected by a too warm and saline deep ocean compared to observations ~~,~~ (e.g. in the EC-Earth model; Sterl et al., 2012). This systematic error (Table 1) is illustrated by comparing temperature pro-

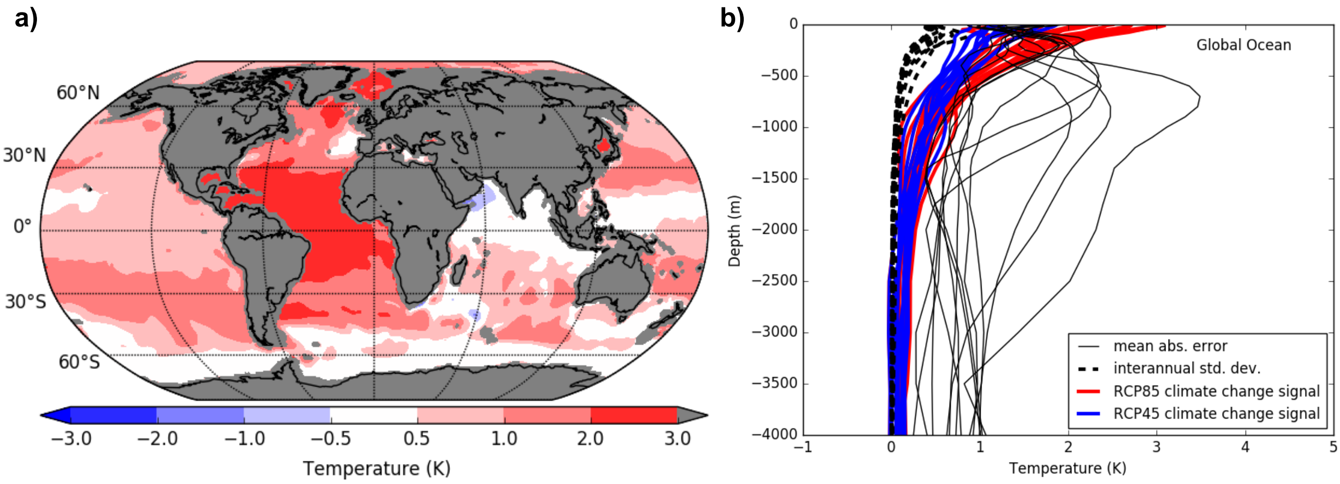


Figure 1. Biases in CMIP5 models with respect to the PHC climatology (PHC 3.0, updated from: Steele et al., 2001). ~~**(left)**~~ **a)** Ensemble mean DJF potential temperature bias [K] at a depth of 1000 m in CMIP5 historical simulations for the period 1971—2000. ~~**(right)**~~ **b)** Individual depth-profiles of the mean absolute potential temperature error in the CMIP5 models (black lines). ~~In the first hundreds of meters, the individual mean absolute errors exceed the~~ The interannual standard deviation [K] (black dashed lines) ~~; in deeper ocean layers,~~ and the ~~error is larger than the~~ RCP4.5 and RCP8.5 climate change signal [K] (2071–2100 minus 1971–2000) ~~(~~; blue and red lines) are given for comparison.

files from CMIP5 historical runs (Fig. 1~~, right~~b) with the PHC3 climatology (PHC 3.0, updated from: Steele et al., 2001). Importantly, the mean absolute error in deeper ocean layers is larger than the interannual variability (the standard deviation of annual means)~~as well as~~. It is also larger than the climate change signal ~~(~~as determined from RCP8.5 and RCP4.5 emission scenarios~~) — that is, errors~~. Or formulated differently, deep ocean biases are larger than the signals we aim to predict, which may be cause for concern in non-linear systems. When considering horizontal maps of the multi model-mean potential temperature bias in 1000 m depth (Fig. 1~~, left~~a), one can clearly see that the largest bias is located in the Atlantic sector. Although one could argue that this error is "well-hidden" from the atmosphere, thus having little impact on atmospheric parameters, it has the potential to change the outcropping region and position of isopycnals. This could lead to a wrong mapping of the deep ocean to the surface, which may result in erroneous water mass formation. In turn, this can potentially have significant effects on the heat uptake of the deep ocean, thus impacting climate change projections. As an example, the magnitude of the projected climate change in the ocean appears to be ordered according to the models' mean absolute errors (Table 1).

Previous work has identified an important role for mesoscale eddies, showing that they act "as a barrier or gatekeeper to heat penetration from the surface into the ocean interior" (Hewitt et al., 2017) by counter-acting the downward heat transport from the mean ocean circulation (Griffies et al., 2015; von Storch et al., 2016). With the increase in simulated eddy activity when increasing resolution towards 0.1°, the magnitude of vertical eddy heat transport also increases, which in turn reduces temperature drifts in the simulated deep ocean when compared to coarse-resolution ocean models of about 1° (Hewitt et al., 2017; Griffies et al., 2015). The physical mechanism behind this upward eddy heat transport is the mixing of heat along inclined surfaces of constant density (isopycnals) by eddies and eddy-induced transport. However, the position and tilt of the isopycnals themselves is also strongly impacted by mesoscale eddies, which can influence the 'mapping' from the surface ocean layers to the deeper ocean. Since globally eddy-resolving climate simulations are still very expensive in a CMIP-context, and since current eddy parameterizations do not seem to capture vertical eddy fluxes to full degree (Hewitt et al., 2017), local refinement to explicitly resolve regions of high eddy activity is thus a promising approach to tackle deep-ocean biases (Zadra et al., 2017).

To study the impact of horizontal resolution on the biases in the deep ocean, the ~~AWI-CM~~ AWI Climate Model (AWI-CM; Sidorenko et al
employed in this work. The 'deep bias' can be reproduced in the AWI-CM 'benchmark' configuration that has a rather coarse nominal ocean resolution of $\sim 1°$ typically employed in CMIP5 (not shown). Therefore, the model is well-suited to study the impact that locally enhanced resolution can have on deep ocean biases in CMIP5 models. In order to test the hypothesis that locally too coarse spatial resolution is responsible for the development of the deep ocean biases, we gradually increase the number of ocean grid points in four additional AWI-CM configurations with otherwise identical settings and parameter choices. It is shown that the strong 'deep bias' in the North Atlantic reduces with higher resolution to rather small values that are comparable to those found in other ocean basins. Together with a competitive throughput of 5–6 simulated years per day for the highest analyzed resolutions, this gives a strong case to aim for a ~~(locally) eddy-resolving configuration~~ high resolution (10 km and higher) in eddy-active regions not only in HighResMIP (Haarsma et al., 2016), but already for AWI's CMIP6 standard configuration.

The paper is structured as follows. Section 2 introduces the model configurations and the hierarchy of ocean meshes with systematically increasing spatial resolution in the North Atlantic. The sensitivity of vertical profiles and horizontal maps of surface and interior biases to increasing spatial resolution is studied in section 3, as well as the development of deep ocean biases along relevant surfaces of constant density. The paper closes with a conclusion and further discussions in section 4.

2 Model configuration

The ~~AWI Climate Model (AWI-CM, formerly ECHAM6-FESOM; Sidorenko et al., 2015; Rackow et al., 2016)~~ AWI-CM (formerly ECHA
a coupled configuration in which ECHAM 6.3.01 (Stevens et al., 2013) is coupled to the Finite Element Sea Ice-Ocean Model (FESOM1.4; Wang et al., 2008; Timmermann et al., 2009; Sidorenko et al., 2011; Wang et al., 2014b). It supports unstructured multi-resolution grids for the ocean and sea ice and has shown good performance in simulating present-day climate when compared to more traditional regular-grid climate models participating in CMIP5 in terms of both the mean climate state (Sidorenko et al., 2015) and ~~the~~ climate variability (Rackow et al., 2016). Compared to the coupling procedure detailed in the above mentioned studies, the model now uses a bicubic mapping for the interpolation of the wind-stress components to the ocean grid in order to better conserve higher-order properties like the curl (Valcke, 2013; Valcke et al., 2013). Furthermore, we removed the intermediate exchange grid from the coupling procedure and instead introduced a direct coupling of the unstructured ocean and quasi-regular atmospheric grids. This effectively reduces the number of necessary interpolations. In this study, we will analyze monthly-mean output of five pre-industrial simulations over a common 100-yr period. The simulations are initialized from the PHC climatology (PHC 3.0, updated from: Steele et al., 2001) and zero velocities. In order

Table 1. CMIP5 models considered in the illustration of the deep ocean bias in Fig. 1, in decreasing order according to their mean absolute potential temperature error in 1000 m depth. The absolute error is computed at every gridpoint as the absolute difference $|T_m - T_o|$, where T_o is the observed and T_m is the modeled potential temperature.

CMIP5 model	mean absolute error for global ocean [K]	interannual std. dev. for global ocean [K]	climate change signal RCP4.5 for global ocean [K]	climate change signal RCP8.5 for global ocean [K]
GISS-E2-R	3.11	0.08	0.61	0.73
MPI-ESM-LR	2.43	0.12	0.38	0.47
GFDL-CM3	2.02	0.06	0.44	0.51
ACCESS1-3	1.94	0.09	0.48	0.59
IPSL-CM5B-LR	1.35	0.05	0.40	0.43
GISS-E2-H	1.03	0.06	0.41	0.52
CCSM4	0.87	0.05	0.40	0.48
HadGEM2-ES	0.87	0.07	-	0.31
NorESM1-ME	0.74	0.05	0.39	0.51
CMCC-CM	0.68	0.04	0.26	0.33
CanESM2	0.66	0.04	0.37	0.46
MRI-ESM1	0.55	0.04	-	0.26
MRI-CGCM3	0.54	0.04	0.16	0.22

This is based on the DJF season and historical runs for the period 1971–2000; for RCP4.5 and RCP8.5 the climate change signal is based on the period 2071-2100 compared to the historical period 1971-2000

to parameterize eddies at non-eddy resolving resolutions, the Gent and McWilliams (1990) parameterization (GM) is applied with isoneutral diffusion (Redi, 1982). All prototype simulations used a reference diffusivity $K_{\text{ref}}(x,y) = 600\,\text{m}^2\,\text{s}^{-1}$, which is scaled by the local resolution (Wang et al., 2014b), and a GM coefficient $K_{\text{GM}} = K_{\text{ref}}/2$. As detailed by Wang et al. (2014b), tapering functions following Danabasoglu and Mc Williams (1995) and Large et al. (1997) are also applied to K_{GM}. Depending on the local resolution, the ~~parameterization~~ GM parameterization in FESOM1.4 is smoothly switched off at resolutions smaller than 25 km (red areas in Fig.2), and its effect increases linearly until 50 km, when the parameterization is fully active ~~.~~ ~~Thus, to give an~~ (Wang et al., 2014b). For example, the parameterization is locally switched off when using the 'MR' and 'HR' meshes, which are locally eddy-resolving, and it is generally active in the lower-resolution 'LR' mesh (see next sections). At mid-latitudes, the Rossby radius is between 25 and 50 km, which is why this simple choice was made. Still, the thresholds of 25 km and 50 km can be considered to be tuning parameters and were chosen in stand-alone simulations with FESOM1.4 using the LR grid. For the Arctic, changing the numbers can result in too diffuse boundary currents (Wang et al., 2014b) and their (automatic) choice remains an important research topic for multi-resolution climate applications.

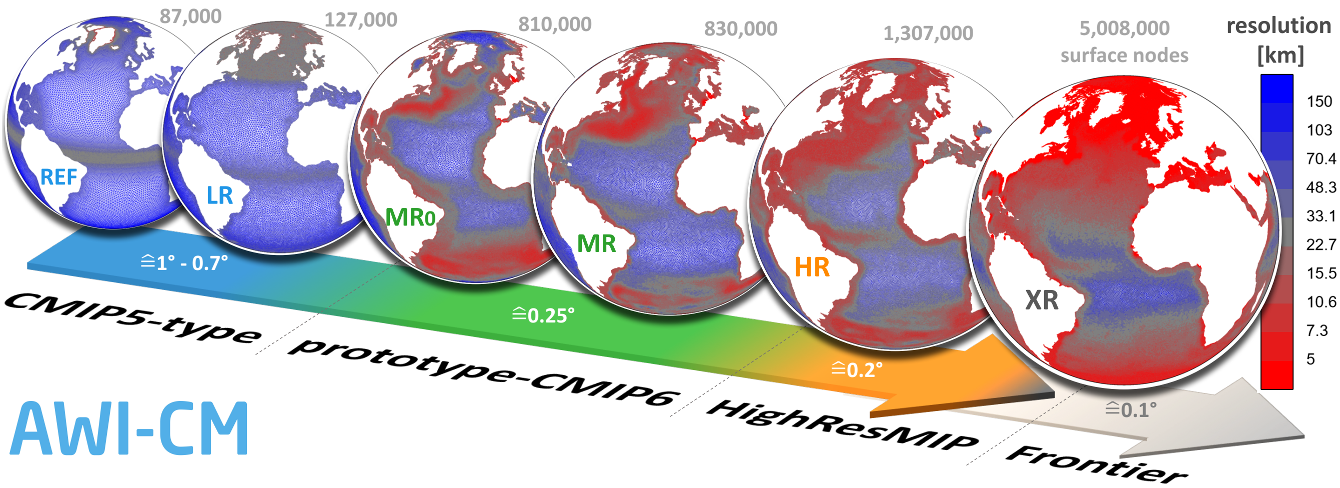


Figure 2. Hierarchy of a set of different ocean grid resolutions that are used in this study. The number of surface grid points increases from left to right and, specifically, the spatial resolution in the North Atlantic Ocean increases from REF up to XR. REF and LR use CMIP5-type spatial resolution, with moderate refinement to about 25km in the tropics and in the Arctic (Rackow et al., 2016; Sidorenko et al., 2011). MR0, MR, and HR are medium- and high-resolution meshes, following a different mesh design strategy (Sein et al., 2016), and focus on the Agulhas and North Atlantic current region. The resolution of the frontier mesh (XR) additionally follows the local Rossby radius of deformation, and is capped at 4 km (7 km) in the Arctic (Antarctic) (Sein et al., 2017). White numbers indicate the approximate spatial resolution of corresponding quasi-Mercator grids with the same number of (wet) surface nodes. The GM parameterization is switched off within the red areas ($\lesssim$25 km).

2.1 Target resolution

In order to find an optimal mesh for the CMIP6 configuration and the associated endorsed Model Intercomparison Projects (MIPs), we performed a hierarchy of prototype pre-industrial CMIP6 simulations with AWI-CM, run at different ocean resolutions (Table 2). Ultimately, we will target coupled configurations with a globally eddy-resolving mesh, which implies "resolving the Rossby radius~~almost everywhere (Hallberg, 2013). We~~ " almost everywhere with at least 2 grid intervals per Rossby radius (Hallberg, 2013). Using this criterion, we have recently reported on the development of such a 'frontier' mesh (XR; see Fig. 2, right globe), with resolution capped at 4 km (7 km) in the Arctic (Antarctic) (Sein et al., 2017). Sein et al. (2017) note that an even finer resolution will be required locally to fully capture mesoscale eddies. A lot of engineering goes into the creation

Table 2. Model settings for the different AWI-CM configurations

AWI-CM configuration	(~~internal~~previous) ocean mesh name	2D ocean grid points	atm. resolution	FESOM	time step ECHAM6	coupling	CPU cores (FESOM+ECHAM)	sim. years per day (SYPD
REF	ref87k	86,803	T63	30 min	450 s	1h	384+192	21.8
LR*	core2	126,859	T127	15 min	200 s	1h	192+576	5.6
MR0	aguv	810,471	T127	7.5 min	200 s	1h	2304+1152	6.2
MR*	glob	830,305	T127	10 min	200 s	1h	1920+1152	6.4
HR*	bold	1,306,775	T127	10 min	200 s	1h	2400+1200	5.5
XR (ocean-only)	fron	5,007,727	-	4 min	-	-	7200	1.5–2

*more details on the AWI-CM CMIP6 configurations at https://github.com/WCRP-CMIP/CMIP6_CVs/blob/master/CMIP6_source_id.json (as of March 2018)

of such meshes, balancing computational resources and simulation quality, since the multi-resolution approach allows for a flexible distribution of the grid points. It is not clear a priori how best to distribute a fixed number of degrees of freedom over the globe, and Sein et al. (2016) have coined the term "mesh design" for this non-trivial task. The XR mesh has about 5 million surface nodes, which is roughly comparable to a 1/10° quasi-Mercator mesh with about 5–6 million (wet) nodes. However, as of today, the XR mesh is still too ~~demanding computationally~~ computationally demanding for the multi-centennial simulations needed in a CMIP context. Therefore, here the idea is to retain some of the beneficial properties of the XR ocean-only simulation analyzed by Sein et al. (2017) by keeping higher resolution only in hotspots of high eddy activity. This reduces the computational cost to a level that is suitable for multi-centennial coupled climate simulations and ensemble simulations.

2.2 Hierarchy of ocean meshes

The hierarchy of different ocean grid resolutions that are used in this study is shown in Fig. 2. The number of surface grid points increases from left to right and, specifically, the spatial resolution in the North Atlantic Ocean systematically increases from 'REF' (reference or 'benchmark' mesh) up to 'HR' (high-resolution). In order to isolate the impact of horizontal resolution, the vertical levels were left unchanged: there are in total 46 levels (at 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 115, 135, 160, 190, 230, 280, 340, 410, 490, 580, 680, 790, 910, 1040, 1180, 1330, 1500, 1700, 1920, 2150, 2400, 2650, 2900, 3150, 3400, 3650, 3900, 4150, 4400, 4650, 4900, 5150, 5400, 5650, and 5900 m). Certainly, going to even higher resolutions beyond the XR mesh, a higher number with different placement of levels might need to be considered, but we kept the standard levels in all meshes for consistency. The bathymetry in the different grids is based on a blend of the IBCAO (Jakobsson et al., 2008) and GEBCO (https://www.gebco.net/data_and_products/gridded_bathymetry_data/gebco_one_minute_grid/) bottom topography data sets, as detailed by Wang et al. (2014b).

'REF' and 'LR' (low-resolution) use CMIP5-type spatial resolution ($\sim 1°$–$0.7°$) with moderate isotropic refinement to about 25 km in the tropics and in the Arctic. The LR mesh was used for ocean-only simulations within the CORE-II intercomparison project (Danabasoglu et al., 2014; Wang et al., 2016a, b) while REF was used as a 'benchmark' mesh for the coupled AWI-

CM (Rackow et al., 2016). Although all prototype simulations except REF use a T127 atmosphere (Table 2), we still include the REF/T63 benchmark configuration here for better comparability with previous studies (Sidorenko et al., 2015; Rackow et al., 2016).

The medium-resolution 'MR0' and 'MR' meshes as well as the high-resolution 'HR' mesh follow the new mesh design strategy introduced by Sein et al. (2016). The main approach is to increase resolution locally over areas of high *observed* eddy variability. While the number of grid points for MR0 and MR is kept at a similar level, MR0 focuses more grid points in the Agulhas region than MR, which in turn focuses them in the North Atlantic Current region (Fig. 2). The HR grid is more balanced in this respect and further increases the size of the areas that use locally increased resolution, resulting in an increase of the number of surface grid points by more than 60%.

It is worth mentioning that HR uses 1.3 million surface grid points (Table 2), similar to traditional 1/4° quasi-Mercator grids (about 1.5 million nodes, of which about 1 million are wet). However, the degrees of freedom on HR are differently distributed, focusing resolution on hotspots of high eddy activity such as the western boundary currents and the Southern Ocean. In fact, this configuration reaches ocean resolutions as high as 1 km locally, e.g. in the Bosporus or over the Danish straits—but still runs at a competitive throughput of 5–6 simulated years per day due to the excellent nearly linear scalability of the FESOM model (e.g. Biastoch et al., 2018). For a spatial map of the eddy-permitting and eddy-resolving regions on the HR grid, please refer to Fig. 4c in Sein et al. (2016).

3 Results

3.1 Vertical profiles of temperature and salinity

~~The drift of temperature and salinity shows~~ Temperature and salinity show major improvements for medium- and high-resolution configurations, as seen from horizontally averaged temperature and salinity profiles for years 71–100 of the pre-industrial simulations (Fig. 3). Differences in the simulated potential temperature and salinity compared to the PHC climatology peak at a depth of around 1000 m. The North Atlantic deep biases, identified both in CMIP5 models (Fig. 1) and in the benchmark ~~version~~ REF/T63 and LR/T127 versions of AWI-CM, successively decrease with increasing ~~spatial~~ ocean resolution, both for potential temperature and for salinity (Fig. 3), thus highlighting the benefit of enhanced spatial resolution. The simultaneous change of the ocean and atmospheric resolution from REF/T63 to LR/T127 leads to a clear improvement of the salinity profiles below 1500 m, and all configurations with T127 atmosphere (LR, MR0, MR, and HR) share a very similar salinity bias in this range. While it is difficult to say what the relative influence is between the atmospheric resolution change (T63 vs T127) and the switch of the ocean grid (REF vs LR), it appears that surface conditions can significantly impact deep ocean biases. Note that the slight drift in HR towards colder temperatures in the 3000–5000 m range is due to a production of denser waters around Antarctica, coinciding with a stronger deep overturning cell in this model configuration.

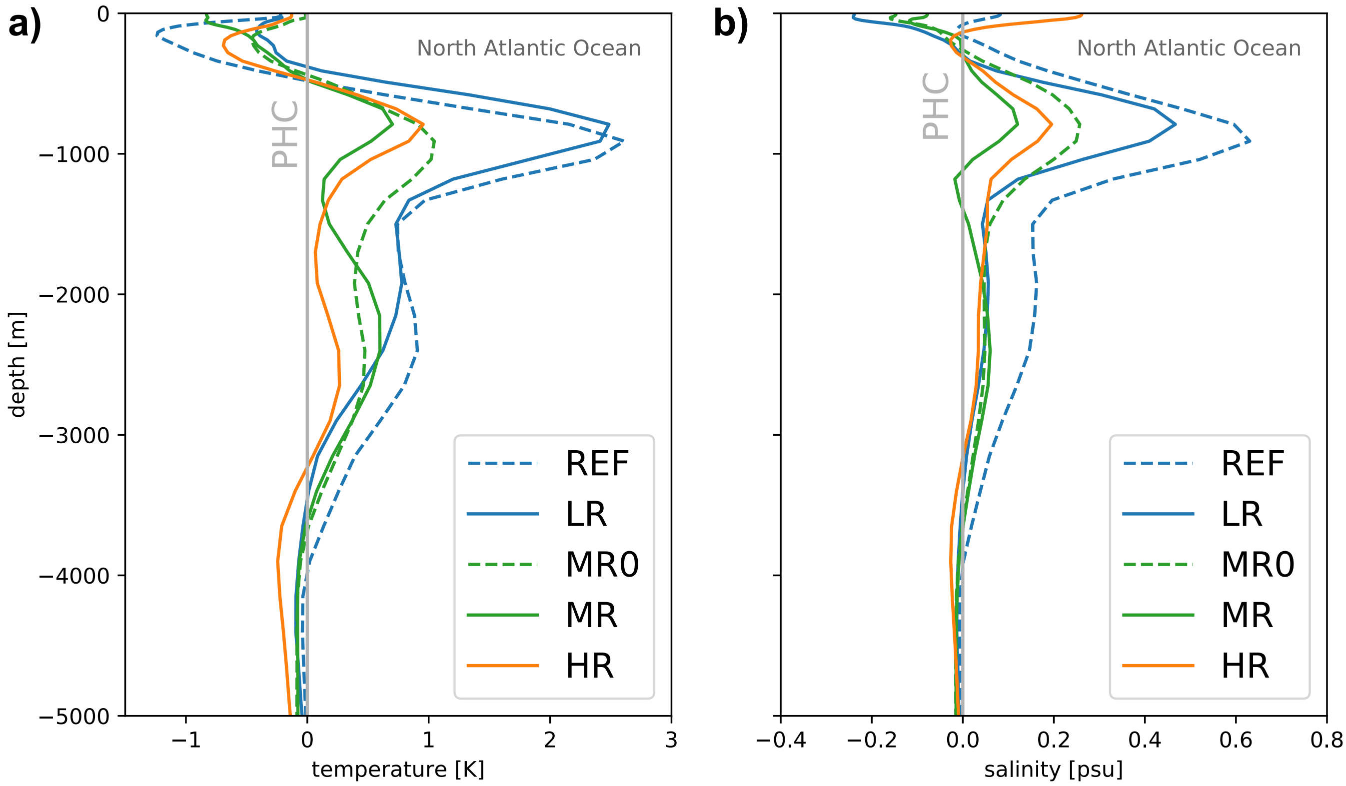


Figure 3. Profiles of potential temperature ~~(left)~~ a) and salinity ~~(right)~~ b) in the North Atlantic Ocean for years 71–100 of the pre-industrial simulations. Shown is the mean difference to the PHC climatology (PHC 3.0, updated from: Steele et al., 2001). With the medium- and high-resolution meshes, the biases around 1000 m depth decrease strongly for both temperature and salinity.

3.2 Hovmoeller diagrams for temperature and salinity drift

In addition to considering biases at the end of the 100yr-simulations discussed above, it is instructive to study the transient development of the biases over time. To this end, time-depth Hovmoeller diagrams (Griffies et al., 2015; von Storch et al., 2016; Hewitt et al., 2017) have been computed for both potential temperature and salinity. The REF and LR configurations show a strong erroneous initial warming at a depth of around 1000 m together with a cooling in the upper ocean above about 400 m (Fig. 4). In the medium- and high-resolution configurations, both the erroneous deep ocean warming and upper-ocean cooling are reduced. Consistent with the study by von Storch et al. (2016), a similar pattern holds for salinity, with freshening in the upper ocean and salinization in the deep ocean. The improvement of the salinity field with increased spatial resolution is similar to the potential temperature case (Fig. 5).

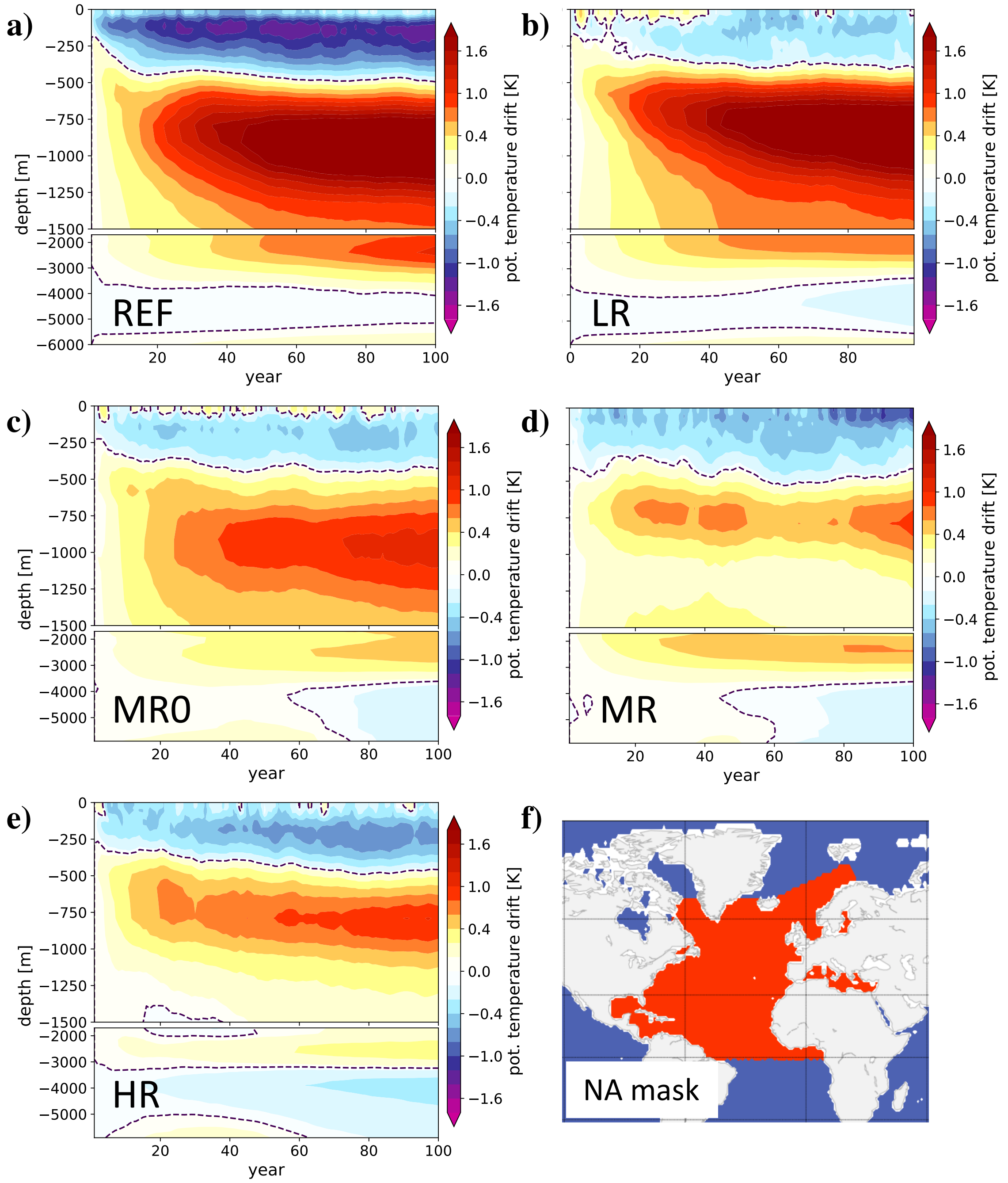


Figure 4. Time-depth Hovmoeller diagram of the potential temperature drift [K] in the North Atlantic for **a)–e)** the five pre-industrial simulations. **f)** Definition of the North Atlantic mask that was used in the Hovmoeller analysis.

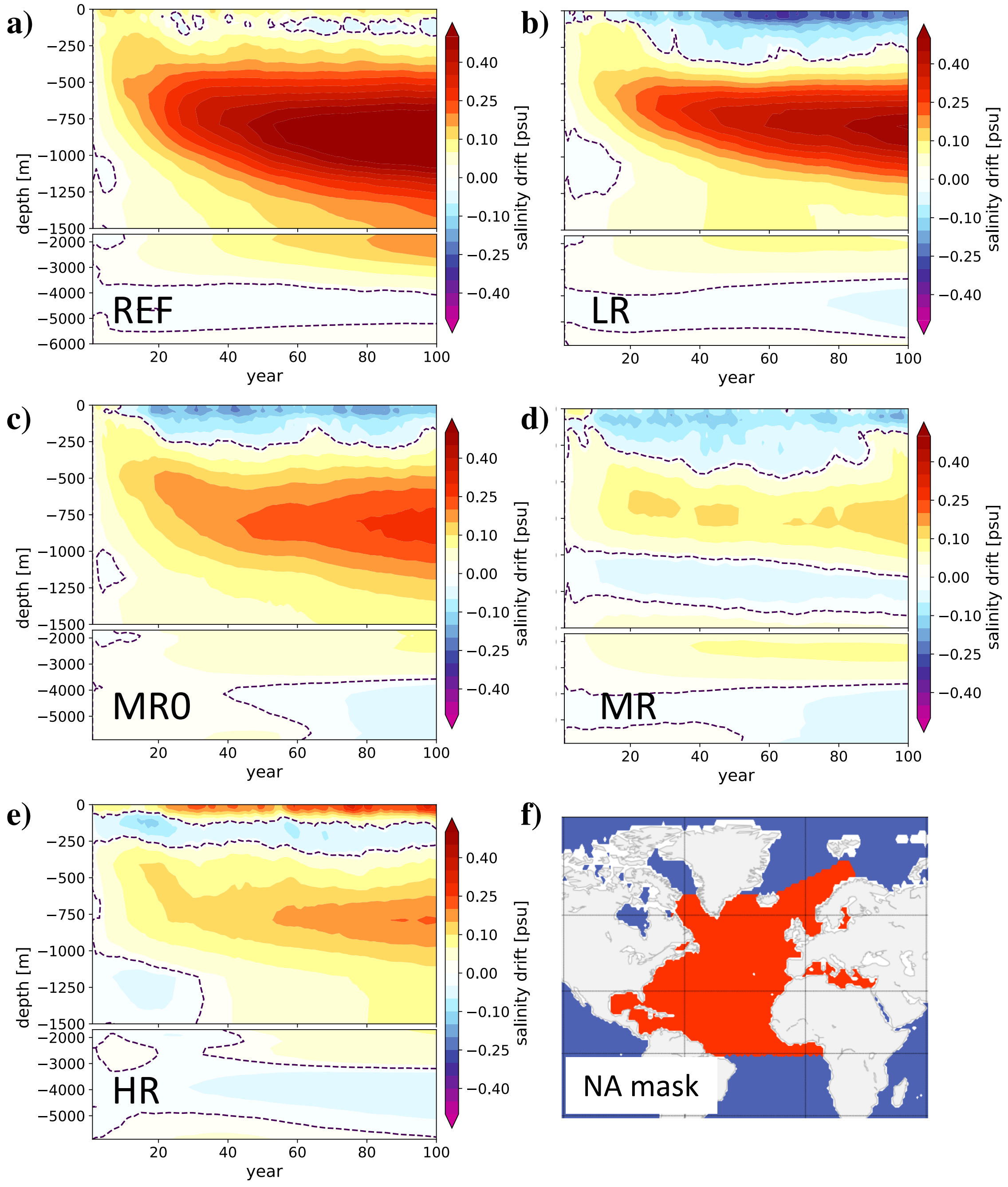


Figure 5. Time-depth Hovmoeller diagram of the salinity drift [psu] in the North Atlantic for **a)–e)** the five pre-industrial simulations. The definition of the North Atlantic mask is identical to the one used in Fig.4.

3.3 Spatial patterns of temperature and salinity biases

3.3.1 Deep ocean (1000 m)

When considering horizontal maps of potential temperature and salinity biases in the deep ocean, the REF and LR configurations show an erroneous warming in the deep Atlantic ocean (Fig. 6a), similar to the pattern identified for the CMIP5 models (Fig. 1a). With increasing resolution in the North Atlantic, there is a very consistent improvement in deep ocean hydrography (Fig. 6), making the remaining biases in MR and HR comparable in magnitude to smaller biases in the other ocean basins. Compared to the changes in the Atlantic, the other basins remain largely unchanged, suggesting that resolution changes in distant regions play a minor role for the bias reduction in the Atlantic.

It appears as if the resolution increase of MR and HR leads to overshooting close to the Strait of Gibraltar since both MR and HR change the sign of the potential temperature and salinity biases at 1000 m. We hypothesize that at these resolutions, smaller issues become relatively more apparent, that is other processes might need to be included for a proper simulation of the Strait of Gibraltar outflow ~~such as the~~ and spreading of Mediterranean Waters into the North Atlantic. Also, resolving the overflow processes at the Strait of Gibraltar would require resolutions on the order of tens of meters in the horizontal (Izquierdo and Mikolajewicz, 2018) and meters in the vertical direction, which is still far from the resolutions applied in this study. Two possible solutions are therefore the use of an overflow parameterization (e.g., Wu et al., 2007), which is currently not implemented in the model, or systematic changes to the bottom (and lateral) topography at the outflow of the Strait of Gibraltar to minimize spurious entrainment. In order to simulate the correct spreading of Mediterranean Waters from the Gulf of Cádiz into the North Atlantic, another approach could be to add additional physics like the effect of tides (Izquierdo et al., 2016), which are usually not included in current climate models. Without tides, ocean models often simulate erroneous south-westward spreading, leading to stronger biases when compared to climatology than in simulations with active tides (Izquierdo and Mikolajewicz, 2018).

3.3.2 Surface conditions

Since there are no heat sources or sinks in the interior ocean, the observed deep bias cannot develop in-situ. The ocean model also does not apply geothermal heating as lower boundary condition (e.g., Adcroft et al., 2001; Downes et al., 2016). Furthermore, since there is no sizable cold (fresh) bias above 1000 m, it cannot be entirely explained by a vertical redistribution of heat (salt). Instead, the surface has to be a major origin of the simulated deep ocean warming, and improvements in the deep ocean hydrography with higher resolution should be caused by improved surface fields.

Focusing on the SST bias in the last 30 years of the REF, MR0, and HR preindustrial simulations (years 71–100) in detail (Fig. 7), systematic differences between the simulations are evident (for the discussion of LR and MR, see Appendix A). The surface is consistently colder than PHC in all simulations, which is expected, since pre-industrial (PI) runs are compared with a climatology representing present-day conditions. However, in the whole Labrador Sea, REF, MR0, and HR are on the warmer side for years 71-100. When overlaying their SST bias with simulated surface isopycnals (gray and black contours in Fig. 7b–d), which represent the mapping to the deep ocean in 600–1000 m depth (see details in the sections below), it is

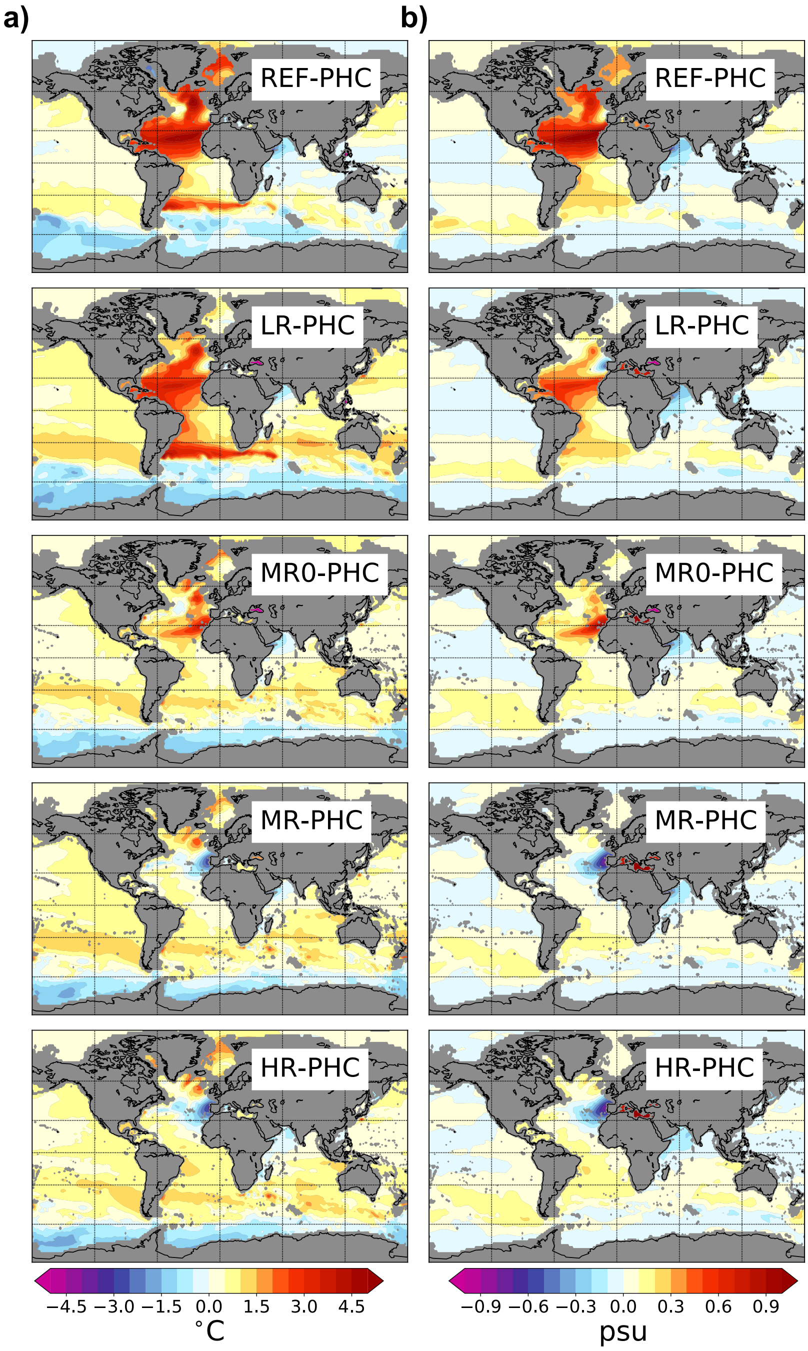


Figure 6. a) Potential temperature [°C] ~~(left column)~~ and b) salinity [psu] ~~(right column)~~ biases with respect to the PHC climatology (PHC 3.0, updated from: Steele et al., 2001) at 1000 m depth, plotted on the observational grid. A systematic decrease of the temperature and salinity biases in the North Atlantic with increasing resolution (top to bottom) is evident.

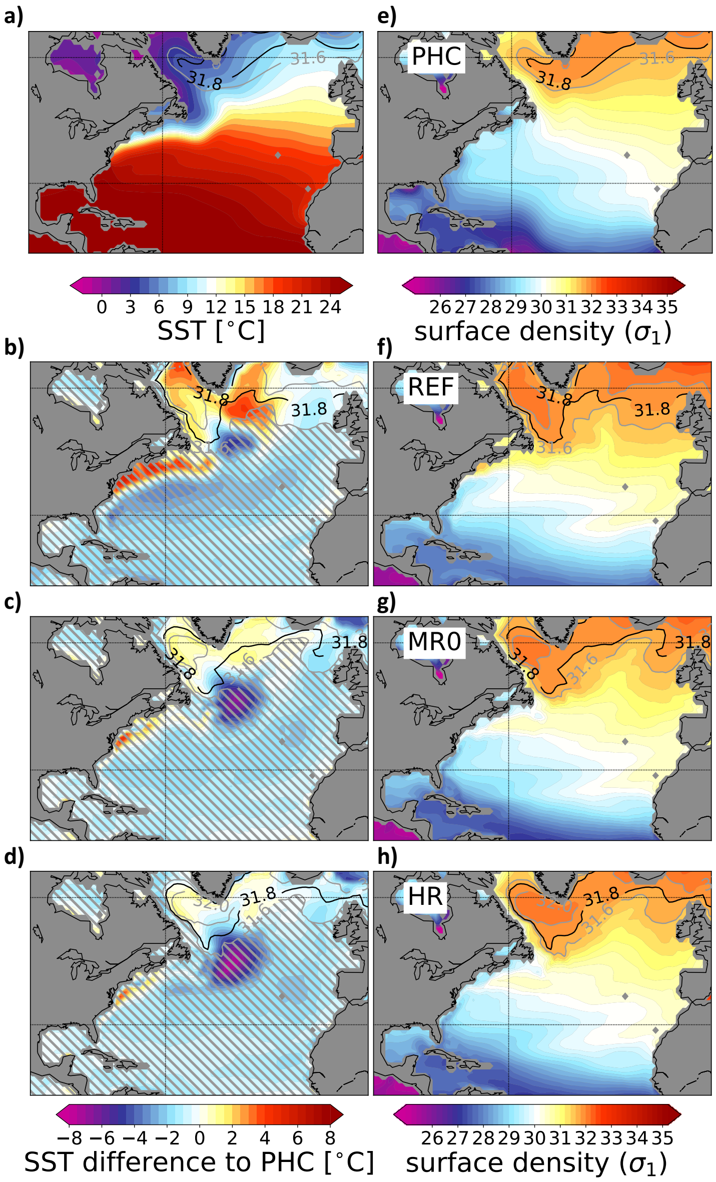


Figure 7. ~~**(top right)**~~ **a)** The North Atlantic sea surface temperature and **e)** σ_1-density structure at the ocean surface, as determined from the PHC climatology (PHC 3.0, updated from: Steele et al., 2001). Black and gray contours indicate outcropping areas for typical isopycnal surfaces found in the deep ocean around 1000 m (e.g. $\sigma_1 = 31.8$). ~~**(right column)**~~ **f)–h)** Same for the simulated density structure in REF, MR0, and HR (years 71–100). ~~**(left column)**~~ **b)–d)** Sea surface temperature (SST) biases in the 3 simulations (years 71–100) with respect to PHC in Fig. 7a. Three simulated σ_1-density contours that represent the 'mapping' to deeper ocean layers are overlaid with black and gray contours (identical to the contour levels in ~~the **right column**~~Fig. 7f–h). To highlight the SST improvements in the areas encircled by these contours, hatching grays out regions that are not in contact with the deep ocean around 600–1000m.

evident that warm SSTs over these critical regions are systematically reduced when going to the higher resolutions (left panels in Fig. 7b–d). Consistent with uncoupled ocean-only results for LR and HR (Sein et al., 2016, their Fig. 7)), which show a much better simulation of the position and separation of the Gulf Stream further south at higher resolutions, the coupled simulations analyzed here also show a successively reduced meridional warm/cold bias pattern along the East Coast of North America.

Despite these clear improvements over the deep convection sites and over the Gulf Stream region, the cold temperature spot in the North-West corner is a persistent bias and is even better visible in the medium- and high-resolution coupled simulations, since the surrounding warm biases are much reduced. Note that also uncoupled ocean-only models still struggle to properly simulate the North-West corner of the North Atlantic (Sein et al., 2017), and presumably much higher resolution along with a more detailed representation of the bathymetry is needed for the Gulf Stream to reach this area. This region Although the Gulf Stream and its extension could impact the location of the outcropping regions, the strong cold temperature spot (hatched in Fig. 7b–d) is, however, not in direct contact with the deep ocean around 600–1000 m depth via outcropping isopycnals and thus does not limit the analysis of the present manuscript, which is focused on the deep ocean.

We conclude that the deeper ocean is connected to less warm surface conditions (non-hatched regions in Fig. 7b–d) in the higher resolution model versions, and in the next section we will study how this translates to the improvements seen in the deep ocean.

3.4 Along-isopycnal bias propagation in the Atlantic

By focusing on surfaces of constant potential density (isopycnals), it is possible to trace the development of the biases from the surface to the deep ocean around 1000 m depth, where our lower-resolution simulations and the CMIP5 models show the strong anomalous warming (Fig. 1). We compute running 10-yr means for the temperature bias along the $\sigma_1 = 31.8$ isopycnal (σ_1 denotes potential density, referenced to 1000 m depth). We chose this specific isopycnal, because it coincides with a depth of 800–1000 m in the North Atlantic area (Fig. 8). It also lies in the middle of the envelope formed by the 31.6 and 32.0 contours that were already shown in Fig. 7 (gray contours). As an example, when

To isolate the influence of the chosen ocean grid using the same atmospheric T127 configuration, we will focus on the LR and HR configurations here as examples. When looking at the bias development in LR (see animation S1 in the video supplement (Rackow et al., 2018b)), there are two major surface source regions for the deep bias in the Atlantic—the Strait of Gibraltar and the north-eastern North Atlantic. The first source of the bias is likely to be of geometric nature, since the very narrow Strait of Gibraltar cannot be properly discretized at coarse resolutions(without . A systematic geometric tuning of the ocean bathymetry in this area); was not attempted (Fig. 9), and higher resolutions are needed to get a more realistic outflow. Furthermore, an overflow parameterization or additional physics like tides could further improve the spreading of Mediterranean Waters into the North Atlantic.

The latter source is related to enhanced downwelling and an erroneously deep mixed-layer (≥ 500 m; green contours in the supplemental animation) in the north-eastern North Atlantic, which is linked to the erroneous pathway of the . This is a feature that has already been identified in uncoupled FESOM simulations using the LR grid as part of the CORE-II intercomparison project (Danabasoglu et al., 2014, their Fig. 13). Since the Gulf Stream in this model version. the LR (and REF) simulations is

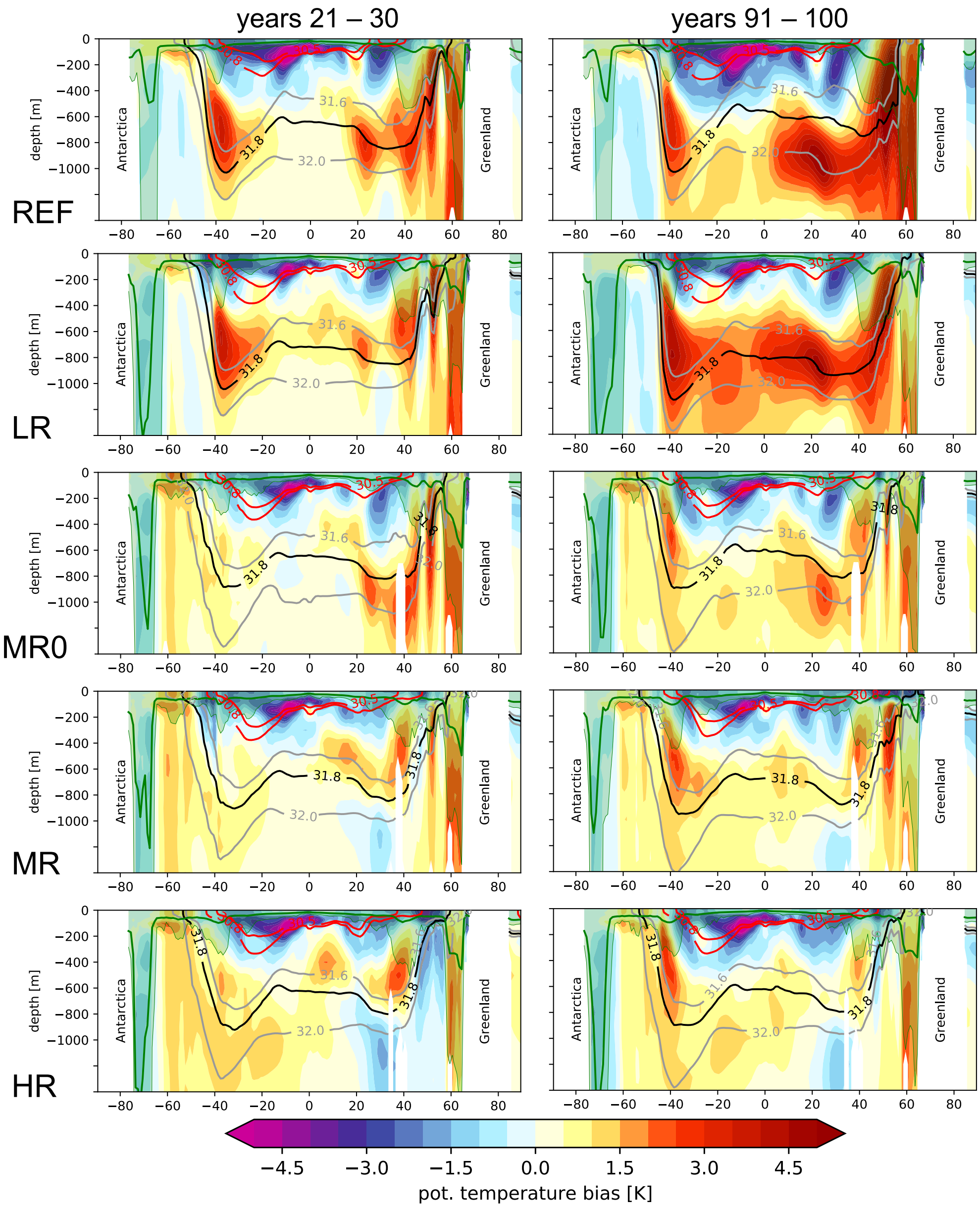


Figure 8. Meridional section at 30.5 °W through the Atlantic Ocean for the potential temperature bias in the five simulations. The difference compared to the PHC climatology is shown with colours for years 21–30 **(left column)** and years 91–100 **(right column)**, illustrating the North Atlantic bias development along isopycnal layers (see animations S3 and S4 for LR and HR with a 10yr running window in the video supplement (Rackow et al., 2018b)). The contours show σ_1 density contours that are representative for the deep ocean between 600 and 1000 m (gray and black; σ_1 =31.6, 31.8, 32.0) and for the surface ocean until a maximum depth of about 300 m (red; σ_1 =30.5, 30.8). The average (maximum) mixed layer depth in the 10-yr windows is overlaid with a green line (green shading).

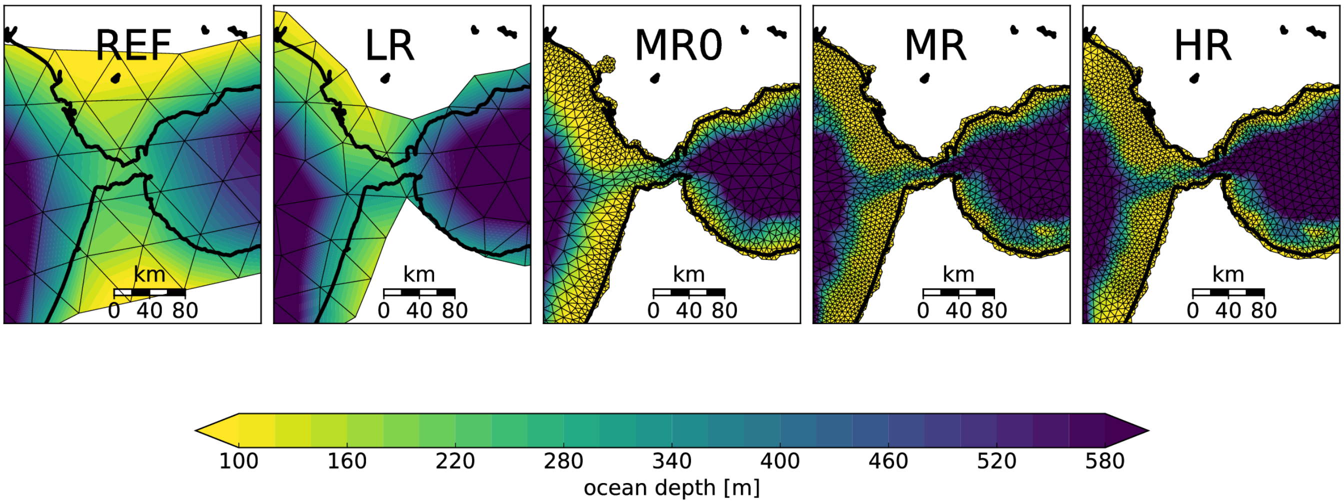


Figure 9. Spatial discretization of the Strait of Gibraltar and the Gulf of Cádiz in the five different model grids. The thick black line shows the true coastline as implemented in the Basemap plotting toolbox, using data from GSHHS (URL not shown in diff document). Triangular elements are shown with thin black lines, colors depict the local ocean depth in meters.

too zonal and reaches the northeastern North Atlantic, part of the flow has to downwell here, which we suspect could explain part of this deficiency by entraining waters and deepening the mixed layer. Other factors influencing the mixed layer depth could be biased buoyancy fluxes or the restratification process via eddy activity. By comparing the years 21–30 to the years 91–100 of LR (second row in Fig. 8~~and animation S3) ,~~) the advective nature of the bias signal propagation from the surface in high latitudes to the deep ocean at lower latitudes is evident, which coincides with the mean currents of the subtropical gyre that go into the same direction.

The mixed layer (green line and shading in Fig. 8~~and in video supplement S3)~~) is deep enough so that surface biases can reach the 31.8 and neighboring isopycnals, from where the signal is further advected towards the south. ~~The animation (video supplement S1) further shows how~~ Eventually the signal is ~~eventually~~ advected towards the equator, from where it propagates to the East as a Kelvin wave (video supplement S1).

In contrast, all above mentioned issues are almost absent in the HR configuration (see last row in Fig. 8 ~~and the animations~~ or the animation S2~~and S4~~), which is a major improvement compared to the previous AWI-CM-LR configuration. This strongly suggests that also in the CMIP5 models ~~it is~~ the lack of spatial resolution ~~that is causing~~ is favouring biases in the deep ocean~~: higher~~. Higher spatial resolution is needed to properly resolve the very narrow geometry of the Strait of Gibraltar and it is one way to better simulate the position of the Gulf Stream, although other factors also play an important role. The latter improvement reduces warm SST biases over North Atlantic areas that are in contact with the deeper ocean (Fig. 7b–d), which in turn reduces the warming in the deep ocean. While a strong resolution-dependence was also shown by Marzocchi et al. (2015), there are additional ways for getting a more realistic Gulf Stream separation. These include details of the numerical scheme that can affect current-topography interactions (Penduff et al., 2007) or the representation of non-local dynamics that impact the formation of a northern recirculation gyre along the North American coast, such as the Deep Western Boundary Current downstream of Cape Hatteras (Zhang and Vallis, 2007) and the cold Labrador Current northward of the Gulf Stream front (Sein et al., 2017).

3.4.1 Displacement and tilt of simulated isopycnals

There is a third source of biases, which is responsible for the deep ocean warming in the Southern Ocean. It is related to the fact that the eddy parameterization (GM) has difficulties in representing the slope of the isopycnals, which is determined by the counteracting effects of Ekman pumping and eddy transport (Farneti et al., 2015). As an example, meridional sections along 10.5°E reveal that the strong deep ocean warming in LR seen in Fig. 6a to the West of Cape Agulhas is linked to too steep simulated isopycnals between 40°S and 5045°S (black and grey contours in Fig. 10b, leftpanel) compared to the much flatter observed tilt of the isopycnals (as in PHC; magenta contours). Already at medium resolution (MR), the simulated isopycnals are isopycnal slope is about halved compared to LR and much closer to the observed isopycnals slope (Fig. 10, rightpanel) with strongly reduced temperature biases, indicating suggesting that the explicitly resolved eddies outperform the eddy parameterization as implemented in FESOMapplied in the prototype simulations with AWI-CM (using a default K_{GM}).

Isopycnals in the upper ocean above 200–300 m in MR ($\sigma_1 = 30.5, 30.8$) are also much closer to the observed state from PHC than in LR (compare red contours to magenta contours in Fig. 10b), associated with an interior bias dipole of warmer/colder temperatures in LR (left panel) and a more homogeneous (cold) bias pattern in MR (right panel). Interestingly, the surface representation (SST bias) of this warm/cold interior bias to the west of Cape Agulhas and a similar dipole-like bias in the Brazil-Malvinas Confluence region are cleanly separated into their warm and cold parts by the $\sigma_1 = 30.5$ isopycnal surface contour (red contour in Fig. 10a, leftpanels) in LR. This suggests that these biases can be largely explained by could be caused by shifted water masses as indicated by the erroneous northward shift of the $\sigma_1 = 30.5$ contour, leading to a warm bias on its northern side and to a cold bias on its southern side. Flattening the slope would result in a southward shift with potentially reduced biases. Indeed, the too steep isopycnals on the coarse mesh with active eddy parameterization, as these surface biases are strongly diminished in MR (Fig. 10a) with better resolved eddies and the associated flatter isopycnals., which are a close fit to the target contours from PHC (Fig. 10b).

3.5 Implications for model initialization

The five simulations in this study are initialized from rest with zero velocities, prescribing long-term mean temperature and salinity fields for boreal winter from PHC. This leads to a fast initial adjustment of geostrophic currents, usually based on a smoothed rather smooth climatology as done in this study, while in reality, e.g., zonal fronts will move up and down throughout the year. After this first phase of fast adjustment, which takes months to one year and is also influenced by the topography as represented on the model grids, significant biases are already apparent after the first years (not shown). As an example, the warm/cold bias pattern along the eastern coast of North America (Fig. 7b), which is related to the too northerly course of the Gulf Stream in AWI-CM-LR and REF, fully develops within a couple of years. We are confident that a focus on (and good understanding of) the initial bias development could lead to significantly improved models, as the later stages are likely

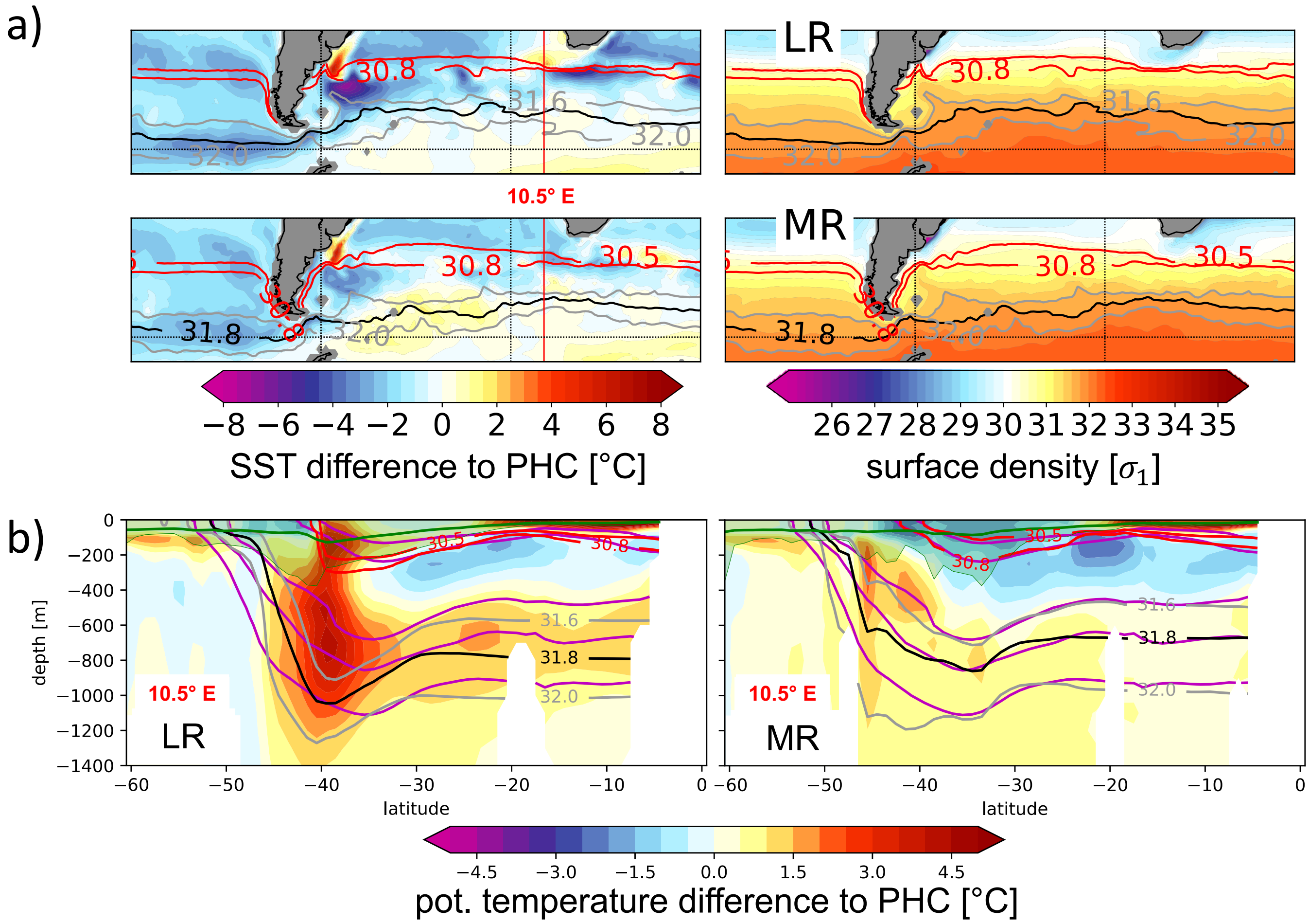


Figure 10. a, (right) The Southern Ocean σ_1-density structure at the ocean surface in LR and MR (years 71–100). Black and gray contours indicate outcropping areas for typical isopycnal surfaces found in the deep ocean around 1000 m; red contours represent shallower isopycnal surfaces with a maximum depth of about 200 m. **a, (left)** Sea surface temperature (SST) biases in the 2 simulations (years 71–100) with respect to PHC. Simulated σ_1-density contours are overlaid (identical to the contour levels in the **right panels**). A meridional section at 10.5 °E is highlighted with a vertical red line. **b)** Meridional section at 10.5 °E, to the west of Cape Agulhas, showing the potential temperature bias with respect to PHC in **(left)** LR and **(right)** MR (years 71–100). Contours show simulated σ_1-density contours that are representative for the deep ocean between ≈600 and 1000 m (gray and black; 31.6, 31.8, 32.0) and for the surface ocean until a maximum depth of about 200 m (red; 30.5, 30.8). In contrast to LR, the tilt of the isopycnals in MR is a close fit to the 'target' σ_1-contours from PHC (given in magenta). The average (maximum) mixed layer depth in the 30-yr window is overlaid with a green line (green shading).

dominated by slow developments in the deep ocean, following these fast initial 'damages'. At higher resolutions like MR and HR, when initialized from zero velocities, it could also become important to temporarily hold the 3D temperature and salinity fields close to a (seasonally varying) climatology as long as the circulation and eddy field is still developing.

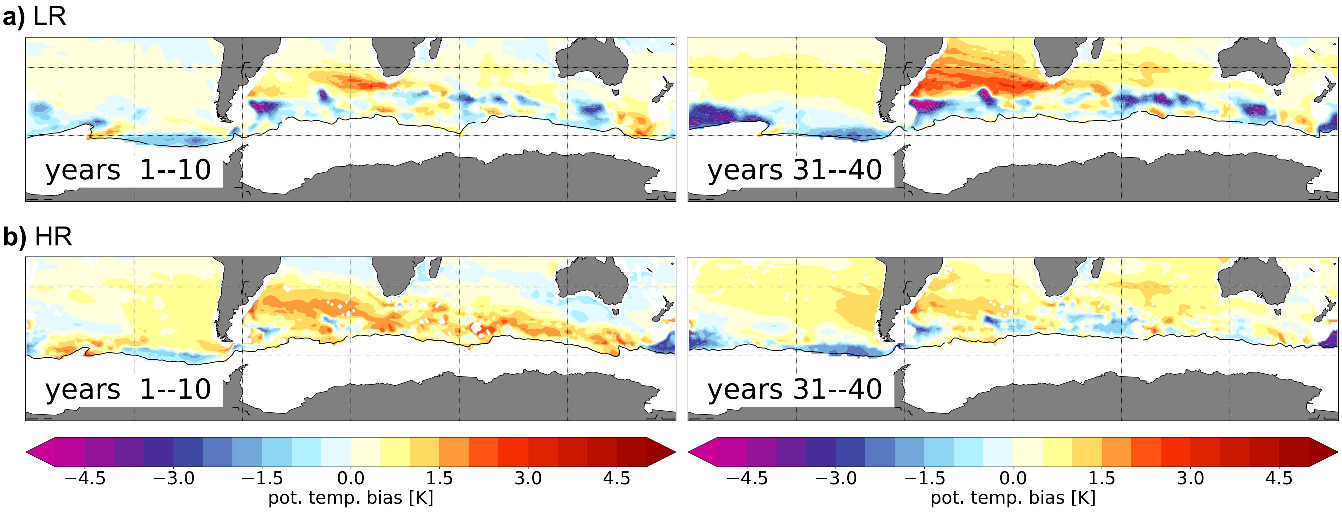


Figure 11. Southern Ocean maps of along-isopycnal potential temperature biases [K] in **a)** LR and **b)** HR for years 1–10 and 31–40. Black contours show the outcropping location of the σ_1=31.8 isopycnal. For animations of the bias development with a 10-yr running window, see supplementary animations S1 and S2 (Rackow et al., 2018b).

Interestingly, in HR, an initial movement of the 31.8 isopycnal surface contour in the Southern Ocean towards the equator ~~is evident (supplemental animation S2), leading apparently~~ apparently leads to larger initial biases than in LR (~~supplemental animation S1~~Fig. 11, left), and then it returns back to the south after 20 or 30 years. ~~After this period~~In years 31–40, the biases seem to recover and are again smaller than in LR ~~.~~ (Fig. 11, right). We hypothesize that this is (i) due to the westerly winds that quickly steepen the isopycnals, thus increasing baroclinicity; and (ii) due to the slowly developing eddy field that later flattens the isopycnals, which again shifts the outcropping region back towards the south. The time scale for the development of the Southern Ocean eddy field is several tens of years (Allison et al., 2010), which fits the behaviour described above. In contrast, the eddy parameterization in LR is active from the start, which keeps the isopycnals initially closer to the observed state, only to be outperformed by the HR simulation with explicitly resolved eddies in the later stages of the simulation.

4 Conclusions

It has been found that CMIP5 models tend to show a strong anomalous warming and salinization in the deep North Atlantic Ocean. Although being substantial in magnitude, to our surprise the deep ocean biases in CMIP5 models did not receive a lot of attention yet. While one could argue that this bias is 'well-hidden' from the atmosphere and therefore not as critical for climate simulations as surface biases, it can impact the outcropping and position of isopycnals. This could lead to a wrong mapping of the deep ocean to the surface and as a consequence to erroneous projections of the heat uptake of the deep ocean. Here we exploit the fact that the AWI-CM at low CMIP5-type resolutions reproduces the behaviour seen in CMIP5 models. We show how the deep ocean bias develops from the surface and how it propagates along relevant isopycnal layers into the deep ocean. Along-isopycnal analyses are common oceanographic diagnostics to trace sources and pathways of temperature and salinity anomalies (e.g. Alban et al., 2001; Nonaka and Sasaki, 2007); and they could be further applied in climate models to determine pathways of anthropogenic heat uptake by the ocean. We found that the deep bias seen in AWI-CM-LR and REF is systematically reduced when moving to successively higher resolutions ~~. Our~~ (10 km and higher) in eddy-active regions, a capability supported by FESOM1.4's use of multi-resolution ocean grids. Although there is certainly scope for improved eddy parameterizations, our results thus highlight the benefit of using high-resolution ocean components in climate modelling.

It should be mentioned that the flexibility of unstructured multi-resolution ocean grids comes with its own challenges: How best to distribute a given number of computational grid points over the globe in climate simulations? While in the past, more idealized approaches to the distribution of the spatial resolution had been performed at AWI (e.g. the refinement towards 0.25° along the equator in REF, or resolution increases over the whole Arctic in LR), the medium- and high-resolution meshes follow a more objective global strategy by focusing resolution in regions of strong observed eddy variability. As a consequence, for example the nominally coarsest mesh, REF, features the highest resolution in the tropical Pacific Ocean among all meshes. Despite the fact that the resolution change in the five meshes is thus not strictly systematic over the global ocean, there is a systematic increase of spatial resolution in the North Atlantic. Since we only consider 100-yr simulations in this study, we do not expect resolution changes in the other basins to impact the simulation of the North Atlantic and the conclusions of our study.

Potentially, the chosen vertical mixing scheme could also impact biases in the deep ocean. However, we could not identify a clear dependence of deep ocean biases on the vertical mixing schemes used in CMIP5 models~~, with the three worst performing CMIP5 models~~ : the three models with the strongest absolute error at a depth of 1000 m (GISS-E2-R, MPI-ESM-LR, GFDL-CM3; see mean absolute potential temperature error in 1000 m in Table 1) ~~using~~ use either KPP or PP mixing (Huang et al., 2014, their Table 1). This suggests that spatial resolution provides an alternative way to reduce long-standing deep ocean biases.

We identified two major sources for the deep ocean biases in the ~~deep~~ Atlantic ocean. The first source is the Strait of Gibraltar, which is likely to be a geometric issue related to the spatial discretization of this narrow strait (15 km) at relatively coarse resolution that is typical for CMIP5 models (about 100 km), ~~leading~~ and that often leads to increased Mediterranean outflow ~~.~~ (e.g., Sterl et al., 2012). Much more systematic efforts are required to tune the horizontal and at the same time the vertical discretization of the Strait of Gibraltar. The warm and saline biases originating from this area ~~thus~~ largely disappear with higher resolution in AWI-CM-MR/HR~~.~~, probably due to lower spurious numerical mixing and an improved representation of the bathymetry to the West of the Strait of Gibraltar (Fig. 9), which should add to the realism of the simulated plume. At the highest resolutions considered here, the bias in the proximity of the Strait of Gibraltar changes sign towards a too cold and fresh anomaly. Ongoing tests suggest a similar sensitivity to the chosen vertical viscosity/diffusion, as it can also affect the exchange by changing the friction between Atlantic and Mediterranean waters (not shown). We suspect that besides local resolution increases using multi-resolution grids, the incorporation of (the effect of) tides in climate models ~~might be a necessary step~~ and the addition of an overflow parameterization might be necessary steps to further improve the model performance.

The second source in the low-resolution configurations is the north-eastern North Atlantic, where erroneous downwelling associated with ~~an~~ typically anomalously deep MLD (Danabasoglu et al., 2014, their Fig. 13) communicates biased surface conditions into deeper layers. The signal then further propagates along isopycnal layers with the sub-polar gyre circulation into the deep Atlantic around 1000 m. This source of the deep ocean biases is largely diminished in the higher resolution configurations, which better simulate the separation of the Gulf Stream and the North Atlantic Current; and, in fact, we could ascribe the improvement in the deep ocean to smaller SST biases over ocean regions that are in contact with the deeper layers around 1000 m.

In the Southern Ocean, there is a third source of deep ocean warming that is related to a displacement of isopycnals, which are locally too steep on the coarse meshes with active default eddy parameterization. Thus, outcropping often happens too far to the north compared to observations~~. Explicitly resolved eddies~~, so that denser water masses will be in contact with atmospheric conditions (fluxes) that are usually in contact with lighter waters, which can impact water mass transformation. Compared to parameterized eddies (with the default GM coefficient), explicitly resolved eddies in the prototype simulations tend to flatten the isopycnals stronger, which reduces sub-surface biases as well as their surface representations locally, e.g. to the West of Cape Agulhas and in the Brazil-Malvinas Confluence region. ~~It~~ Since we were using a default GM coefficient for all simulations, it can be argued that a regional tuning of GM with a horizontally varying coefficient (Visbeck et al., 1997; Danabasoglu et al., 2012) could lead to a better simulation of the Southern Ocean in low-resolution AWI-CM configurations.

~~It should be mentioned that the flexibility of unstructured multi-resolution ocean meshes comes with its own challenges: How best to distribute a given number of computational grid points over the globe in climate simulations? While in the past, more idealized approaches to the distribution of the spatial resolution had been performed at AWI (e.g. the refinement towards 0.25° along the equator in REF, or resolution increases over the whole Arctic in LR), the medium- and~~ Moreover, high-resolution ~~meshes follow a more objective global strategy by focusing resolution in regions of strong observed eddy variability. As a consequence, for example the nominally coarsest mesh, REF, features the highest resolution in the tropical Pacific Ocean among all meshes. Despite the fact that the resolution change in the five meshes is thus not strictly systematic over the global ocean, there is a systematic increase of spatial resolution in the North Atlantic. Since we only consider 100-yr simulationsin this study, we do not expect resolution changes in the other basins to impact the simulation of the North Atlantic and the conclusions of our study.~~ simulations and their effective K_{GM} could also serve as a template for the regional tuning of low-resolution simulations.

The remaining biases between ± 20–40°N/S, seen in meridional sections along 30.5 °W through the Atlantic, show a consistent warm/cold pattern in the vertical direction. Griffies et al. (2015) also study surface and interior temperature bias maps and show that "where the upper portion of the gyres is cool, the deeper portion is warm". They conclude that mean vertical heat transport from the upper ocean into the interior ocean by the time-mean currents is too strong in their 1° (and to some extent in their 0.25°) configurations, or rather it is not sufficiently compensated by the upward transport from mesoscale eddies. Apparently, typical current eddy parameterizations are not sufficient to offset the downward heat transport from the mean circulation. This implies a possible limitation of our focus of high spatial resolution only in areas of strong eddy activity in AWI-CM-MR and -HR (mainly over the western boundary currents and in the Southern Ocean) since resolution could be important even in the gyre centers to get a realistic magnitude of vertical eddy transports.

The Hovmoeller diagrams for the potential temperature and salinity in the North Atlantic Ocean reveal strongly reduced drifts in the interior ocean at medium and high resolutions, which fits previous findings (von Storch et al., 2016; Hewitt et al., 2017). However, one cannot rule out the possibility that the higher resolution configurations could be drifting only slower towards an equally large equilibrium error, and it remains to be seen whether the strong improvements seen over the 100yr-timescale will last on multi-centennial timescales. Even so, a slower drift at higher ocean resolution is certainly very beneficial for efforts related to ocean reanalysis, and seasonal, interannual, and decadal prediction.

Overall, we have shown major improvements when using medium-resolution (MR) and high-resolution (HR) meshes on representing the hydrography in the deep ocean ~~.~~ around 1000 m. These improvements at depth do not come at the expense of degradations in other climatically relevant fields, as shown by a performance index analysis (Appendix B), but rather improve both the ocean and atmospheric simulation. These grids are partly eddy-resolving and partly at most eddy-permitting, so that eddy parameterizations still need to be applied locally. This calls for dedicated in-depth analyses of eddy heat fluxes (and budgets) and their representation on multi-resolution unstructured grids in future studies. Owing to the competitive speed of 6 simulated years per day, the MR mesh can be used for our CMIP6 standard configuration AWI-CM-MR (with T127 atmosphere), and the HR mesh is used in the HighResMIP project. Next steps will be the development of frontier climate simulations (e.g. AWI-CM-XR) with meshes of 6 or more millions of surface grid points and higher-resolution atmospheres (T255 or higher). With FESOM1.4's finite-volume successor FESOM2 (Danilov et al., 2017), which is ~ 3 times faster and more resource-efficient, running this class of flagship meshes will become possible even for coupled simulations. The corresponding coupled model with its tentative name AWI-CM2 is close to its test phase, and we expect a major step change in the quality of the simulated climate at these resolutions.

This paper does not document AWI's final CMIP6 pre-industrial control simulations that will likely undergo additional changes to the model configuration and further tuning. ~~Additionally, they~~ Tuning could potentially affect the deep ocean simulation, although the global top-of-the-atmosphere (TOA) balance in particular appears not to be directly related to the magnitude of North Atlantic deep ocean biases (not shown). Additionally, the final simulations will use updated ozone forcing that had not yet been available at the time of writing. However, we deem it very important to report on significant improvements during the model development cycle that could also be of interest for other groups developing high-resolution models, in order to document identified sensitivities of model biases to the various possible sources in global coupled climate models.

5 Code availability

The source code and used configuration (namelists) for the coupled FESOM model that is part of AWI-CM1.0 is archived at http://doi.org/10.5281/zenodo.1342014 (Rackow et al., 2018a). The ECHAM6 source code is maintained by the Max Planck Institute for Meteorology and freely available to the public at large (http://www.mpimet.mpg.de/en/science/models/mpi-esm/echam/). External access to the ECHAM6 model is provided through their licensing procedure (http://www.mpimet.mpg.de/en/science/models/license/). If you are interested in the full coupled model including the ECHAM6 sources, you need to register on the MPI-ESM user page (https://www.mpimet.mpg.de/en/science/models/mpi-esm/users-forum/) and then download the

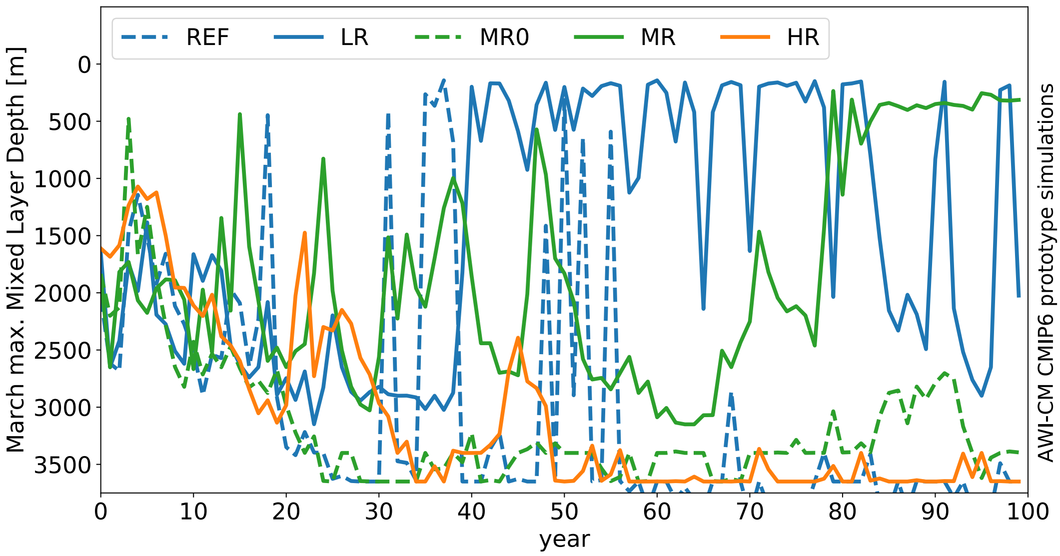


Figure 12. Maximum mixed layer depth [m] for March in the Labrador Sea for the five 100-yr simulations with AWI-CM. The simulated mixed layer starts to diverge after about 20-30 years into the coupled simulations. At the end of the simulation (years 71—100), LR and MR have the lowest mixed-layer while REF, MR0, and HR simulate overly deep mixed-layers in the Labrador Sea.

complete coupled AWI-CM model (rev140 was used in this study) from the SVN repository at https://swrepo1.awi.de/svn/awi-cm/trunk[at]140. After registering, the code can be accessed using the open-source subversion software (http://subversion.apache.org/). Updated code for AWI-CM will be available through the same link. Mesh partitioning in FESOM is based on the METIS Version 4.0 package developed at the Department of Computer Science & Engineering at the University of Minnesota (http://glaros.dtc.umn.edu/gkhome/views/metis). METIS and the pARMS solver (Li et al., 2003) are separate libraries which are freely available subject to their licenses. The OASIS3-MCT coupler is available for download at https://portal.enes.org/oasis.

6 Data availability

The video supplements S1 to S4 are archived at Zenodo, http://doi.org/10.5281/zenodo.1323334 (Rackow et al., 2018b). The data of the five simulations (years 71–100) can be publicly accessed at the DKRZ cloud at https://swiftbrowser.dkrz.de/public/dkrz_035d8f6ff058403bb42f8302e6badfbc/Rackow_DeepBias_GMD2018/. The Polar Science Center Hydrographic Climatology (PHC3.0; Steele et al., 2001) is used for comparison and is freely available online (http://psc.apl.washington.edu/nonwp_projects/PHC/Data3.html).

Appendix A: Surface conditions in LR and MR

The applied model version of AWI-CM (rev140) has too high simulated variability in the Labrador Sea, causing occasional "on" and "off" episodes of deep convection in the Labrador, which can mask changes at the surface on a decadal time-scale (Sidorenko et al., 2015; Rackow et al., 2016). In the whole Labrador Sea, LR and MR show only cold SST biases (not shown) for years 71-100, while the other three configurations (REF, MR0, HR) are on the warmer side (Fig. 7, left panelsb–d). As mentioned above, the LR and MR behavior can be explained by the occurrence of strongly reduced deep convection in those years (green and blue solid lines in Fig. 12) associated with too high sea-ice coverage, leading to the strong cold SST biases. To draw definite conclusions at the surface for the LR and MR configuration is thus more difficult than for the deep ocean analysis. We therefore focused the surface analysis in section 3.3.2 on the other low-, medium-, and high-resolution simulations (REF, MR0, and HR). A separate branch of development at AWI is dealing with this issue of too high variability in the Labrador Sea, and in preliminary tests with a newer AWI-CM version that uses a different mixing scheme in the ocean (KPP; Large et al., 1994) and newer versions of ECHAM6 (ECHAM 6.3.02p4/6.3.04p1) this issue is gone, and we will report on these simulations in the future.

Appendix B: Computation of oceanic performance indices

Extending on the idea to compute performance indices (PI) that grade climate model simulations of various atmospheric parameters (Reichler and Kim, 2008), performance indices for the ocean are computed in this study as follows: First, FESOM potential temperature and salinity data are interpolated horizontally and vertically to the grid of the PHC climatology. This is done for both climatological winter (DJF) and summer (JJA) means of the last 30 years of the AWI-CM simulations. Afterwards, the absolute winter and summer temperature and salinity errors with respect to the PHC climatology are calculated for each grid point and averaged globally, or over individual ocean basins. The same is done with an ensemble of 21 CMIP5 models for which the three-dimensional temperature and salinity fields were available at the time of download. FESOM absolute errors for winter and summer temperature and salinity are normalized with the mean absolute errors of the CMIP5 ensemble (for each individual ocean basin and globally). In Table 3, we give the average over the two parameters and two seasons globally and for two key ocean areas (North Atlantic and Southern Ocean). We set the southern limit of the North Atlantic as 0N while the northern limit is composed of the 65N latitude line west of Iceland, a straight line from Iceland to Spitsbergen, and a straight line from Spitsbergen to the northern tip of Norway (as shown in Fig. 4f). The Southern Ocean is defined here as the ocean area south of 40S. The atmospheric PI are computed as detailed in Sidorenko et al. (2015) and Appendix 1 and 2 in Rackow et al. (2016).

Acknowledgements. D. Sein's work was supported by the PRIMAVERA project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 641727. N. Koldunov benefited from funding by the German Research Foundation through the Collaborative Research Centre TRR 181 "Energy Transfer in Atmosphere and Ocean". D. Sidorenko and Q. Wang were supported through the regional climate initiative REKLIM. Some of the research underlying this report was also funded by

Table 3. Oceanic performance indices (PI) for the global ocean, two important areas (North Atlantic and Southern Ocean), and PI for key atmospheric parameters

AWI-CM configuration	Oceanic PI			Atmospheric PI*
	Global Ocean	North Atlantic	Southern Ocean	
REF (T63)	0.87	0.98	0.68	1.03
LR (T127)	0.72	0.80	0.74	0.87
MR (T127)	0.64	0.62	0.62	0.81
MR0 (T127)	0.61	0.57	0.62	0.79
HR (T127)	0.66	0.63	0.62	0.80

*PIs below (above) 1 indicate that a model performs better (worse) than the average of the considered CMIP5 models (Sidorenko et al., 2015)

the Federal Ministry for Education and Research (Germany) with the support code 01DJ15029 and 01DJ16016. The responsibility for the content of this publication lies with the authors. All plots except Figure 2 have been made with the 2D graphics environment Matplotlib (Hunter, 2007) using Jupyter (Kluyver et al., 2016). We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 1 of this paper) for producing and making available their model output. All simulations have been performed at the German Climate Computing Center (DKRZ).

References

Adcroft, A., Scott, J. R., and Marotzke, J.: Impact of geothermal heating on the global ocean circulation, Geophysical Research Letters, 28, 1735–1738, doi:10.1029/2000GL012182, https://doi.org/10.1029/2000GL012182, 2001.

Alban, L., Ragu, M., and J., B. A.: A model study of temperature anomaly propagation from the subtropics to tropics within the South Atlantic Thermocline, Geophysical Research Letters, 28, 1271–1274, doi:10.1029/2000GL011418, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000GL011418, 2001.

Allison, L. C., Johnson, H. L., Marshall, D. P., and Munday, D. R.: Where do winds drive the Antarctic Circumpolar Current?, Geophysical Research Letters, 37, doi:10.1029/2010GL043355, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010GL043355, 2010.

Biastoch, A., Sein, D., Durgadoo, J. V., Wang, Q., and Danilov, S.: Simulating the Agulhas system in global ocean models – nesting vs. multi-resolution unstructured meshes, Ocean Modelling, 121, 117 – 131, doi:https://doi.org/10.1016/j.ocemod.2017.12.002, http://www.sciencedirect.com/science/article/pii/S146350031730197X, 2018.

Danabasoglu, G. and Mc Williams, J. C.: Sensitivity of the Global Ocean Circulation to Parameterizations of Mesoscale Tracer Transports, Journal of Climate, 8, 2967–2987, doi:10.1175/1520-0442(1995)008<2967:SOTGOC>2.0.CO;2, https://doi.org/10.1175/1520-0442(1995)008<2967:SOTGOC>2.0.CO;2, 1995.

Danabasoglu, G., Bates, S. C., Briegleb, B. P., Jayne, S. R., Jochum, M., Large, W. G., Peacock, S., and Yeager, S. G.: The CCSM4 Ocean Component, Journal of Climate, 25, 1361–1389, doi:10.1175/JCLI-D-11-00091.1, https://doi.org/10.1175/JCLI-D-11-00091.1, 2012.

Danabasoglu, G., Yeager, S. G., Bailey, D., Behrens, E., Bentsen, M., Bi, D., Biastoch, A., Böning, C., Bozec, A., Canuto, V. M., Cassou, C., Chassignet, E., Coward, A. C., Danilov, S., Diansky, N., Drange, H., Farneti, R., Fernandez, E., Fogli, P. G., Forget, G., Fujii, Y., Griffies, S. M., Gusev, A., Heimbach, P., Howard, A., Jung, T., Kelley, M., Large, W. G., Leboissetier, A., Lu, J., Madec, G., Marsland, S. J., Masina, S., Navarra, A., Nurser, A. G., Pirani, A., y Mélia, D. S., Samuels, B. L., Scheinert, M., Sidorenko, D., Treguier, A.-M., Tsujino, H., Uotila, P., Valcke, S., Voldoire, A., and Wang, Q.: North Atlantic simulations in Coordinated Ocean-ice Reference Experiments phase {II} (CORE-II). Part I: Mean states ", Ocean Modelling, 73, 76 – 107, doi:http://dx.doi.org/10.1016/j.ocemod.2013.10.005, 2014.

Danilov, S., Sidorenko, D., Wang, Q., and Jung, T.: The Finite-volumE Sea ice?Ocean Model (FESOM2), Geosci. Model Dev., 10, 765–789, doi:10.5194/gmd-10-765-2017, 2017.

Downes, S. M., Hogg, A. M., Griffies, S. M., and Samuels, B. L.: The Transient Response of Southern Ocean Circulation to Geothermal Heating in a Global Climate Model, Journal of Climate, 29, 5689–5708, doi:10.1175/JCLI-D-15-0458.1, https://doi.org/10.1175/JCLI-D-15-0458.1, 2016.

Farneti, R., Downes, S. M., Griffies, S. M., Marsland, S. J., Behrens, E., Bentsen, M., Bi, D., Biastoch, A., Böning, C., Bozec, A., Canuto, V. M., Chassignet, E., Danabasoglu, G., Danilov, S., Diansky, N., Drange, H., Fogli, P. G., Gusev, A., Hallberg, R. W., Howard, A., Ilicak, M., Jung, T., Kelley, M., Large, W. G., Leboissetier, A., Long, M., Lu, J., Masina, S., Mishra, A., Navarra, A., Nurser, A. G., Patara, L., Samuels, B. L., Sidorenko, D., Tsujino, H., Uotila, P., Wang, Q., and Yeager, S. G.: An assessment of Antarctic Circumpolar Current and Southern Ocean meridional overturning circulation during 1958–2007 in a suite of interannual CORE-II simulations, Ocean Modelling, 93, 84–120, doi:https://doi.org/10.1016/j.ocemod.2015.07.009, http://www.sciencedirect.com/science/article/pii/S1463500315001183, 2015.

Gent, P. R. and McWilliams, J. C.: Isopycnal mixing in ocean circulation models, J. Phys. Oceanogr., 20, 150–155, 1990.

Griffies, S. M., Winton, M., Anderson, W. G., Benson, R., Delworth, T. L., Dufour, C. O., Dunne, J. P., Goddard, P., Morrison, A. K., Rosati, A., Wittenberg, A. T., Yin, J., and Zhang, R.: Impacts on Ocean Heat from Transient Mesoscale Eddies in a Hierarchy of Climate Models, Journal of Climate, 28, 952–977, doi:10.1175/JCLI-D-14-00353.1, https://doi.org/10.1175/JCLI-D-14-00353.1, 2015.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J.-S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, Geoscientific Model Development, 9, 4185–4208, doi:10.5194/gmd-9-4185-2016, https://www.geosci-model-dev.net/9/4185/2016/, 2016.

Hallberg, R.: Using a resolution function to regulate parameterizations of oceanic mesoscale eddy effects, Ocean Modelling, 72, 92 – 103, doi:10.1016/j.ocemod.2013.08.007, 2013.

Hewitt, H. T., Bell, M. J., Chassignet, E. P., Czaja, A., Ferreira, D., Griffies, S. M., Hyder, P., McClean, J. L., New, A. L., and Roberts, M. J.: Will high-resolution global ocean models benefit coupled predictions on short-range to climate timescales?, Ocean Modelling, pp. –, doi:https://doi.org/10.1016/j.ocemod.2017.11.002, https://www.sciencedirect.com/science/article/pii/S1463500317301774, 2017.

Huang, C. J., Qiao, F., and Dai, D.: Evaluating CMIP5 simulations of mixed layer depth during summer, Journal of Geophysical Research: Oceans, 119, 2568–2582, doi:10.1002/2013JC009535, http://dx.doi.org/10.1002/2013JC009535, 2014.

Hunter, J. D.: Matplotlib: A 2D graphics environment, Computing In Science & Engineering, 9, 90–95, doi:10.1109/MCSE.2007.55, 2007.

Izquierdo, A. and Mikolajewicz, U.: The role of tides in the spreading of Mediterranean Outflow waters along the southwestern Iberian margin, Ocean Modelling, 133, 27–43, doi:https://doi.org/10.1016/j.ocemod.2018.08.003, https://doi.org/10.1016/j.ocemod.2018.08.003, 2018.

Izquierdo, A., Kagan, B. A., Sein, D. V., and Mikolajewicz, U.: Modelling in the Strait of Gibraltar: from Operational Oceanography to Scale Interactions, Fundamental and Applied Hydrophysics. Russian Academy of Sciences: Saint Petersburg. ISSN 2073-6673, 9, 15–24, http://hydrophysics.info/?p=3098&lang=en, 2016.

Jakobsson, M., Macnab, R., Mayer, L., Anderson, R., Edwards, M., Hatzky, J., Schenke, H. W., and Johnson, P.: An improved bathymetric portrayal of the Arctic Ocean: Implications for ocean modeling and geological, geophysical and oceanographic analyses, Geophysical Research Letters, 35, doi:10.1029/2008GL033520, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008GL033520, 2008.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., and Willing, C.: Jupyter Notebooks – a publishing format for reproducible computational workflows, in: Positioning and Power in Academic Publishing: Players, Agents and Agendas, edited by Loizides, F. and Schmidt, B., pp. 87 – 90, IOS Press, 2016.

Large, W. G., McWilliams, J. C., and Doney, S. C.: Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization, Rev. Geophys., 32, 363–403, doi:10.1029/94rg01872, 1994.

Large, W. G., Danabasoglu, G., Doney, S. C., and McWilliams, J. C.: Sensitivity to Surface Forcing and Boundary Layer Mixing in a Global Ocean Model: Annual-Mean Climatology, Journal of Physical Oceanography, 27, 2418–2447, doi:10.1175/1520-0485(1997)027<2418:STSFAB>2.0.CO;2, https://doi.org/10.1175/1520-0485(1997)027<2418:STSFAB>2.0.CO;2, 1997.

Li, Z., Saad, Y., and Sosonkina, M.: pARMS: a parallel version of the algebraic recursive multilevel solver, Numerical Linear Algebra with Applications, 10, 485–509, doi:10.1002/nla.325, https://onlinelibrary.wiley.com/doi/abs/10.1002/nla.325, 2003.

Marzocchi, A., Hirschi, J. J.-M., Holliday, N. P., Cunningham, S. A., Blaker, A. T., and Coward, A. C.: The North Atlantic subpolar circulation in an eddy-resolving global ocean model, Journal of Marine Systems, 142, 126–143, doi:https://doi.org/10.1016/j.jmarsys.2014.10.007, http://www.sciencedirect.com/science/article/pii/S0924796314002437, 2015.

Nonaka, M. and Sasaki, H.: Formation Mechanism for Isopycnal Temperature–Salinity Anomalies Propagating from the Eastern South Pacific to the Equatorial Region, Journal of Climate, 20, 1305–1315, doi:10.1175/JCLI4065.1, https://doi.org/10.1175/JCLI4065.1, 2007.

Penduff, T., Le Sommer, J., Barnier, B., Treguier, A.-M., Molines, J.-M., and Madec, G.: Influence of numerical schemes on current-topography interactions in 1/4° global ocean simulations, Ocean Science, 3, 509–524, doi:10.5194/os-3-509-2007, https://www.ocean-sci.net/3/509/2007/, 2007.

Rackow, T., Goessling, H. F., Jung, T., Sidorenko, D., Semmler, T., Barbi, D., and Handorf, D.: Towards multi-resolution global climate modeling with ECHAM6-FESOM. Part II: climate variability, Climate Dynamics, doi:10.1007/s00382-016-3192-6, http://dx.doi.org/10.1007/s00382-016-3192-6, 2016.

Rackow, T., Sein, D., Semmler, T., Danilov, S., Koldunov, N., Sidorenko, D., Wang, Q., and Jung, T.: The AWI Climate Model (AWI-CM), version 1.0, doi:10.5281/zenodo.1342014, https://doi.org/10.5281/zenodo.1342014, 2018a.

Rackow, T., Sein, D., Semmler, T., Danilov, S., Koldunov, N., Sidorenko, D., Wang, Q., and Jung, T.: Sensitivity of deep ocean biases to horizontal resolution in prototype CMIP6 simulations: video supplements, doi:10.5281/zenodo.1323333, https://doi.org/10.5281/zenodo.1323333, 2018b.

Redi, M. H.: Oceanic Isopycnal Mixing by Coordinate Rotation, Journal of Physical Oceanography, 12, 1154–1158, doi:10.1175/1520-0485(1982)012<1154:OIMBCR>2.0.CO;2, https://doi.org/10.1175/1520-0485(1982)012<1154:OIMBCR>2.0.CO;2, 1982.

Reichler, T. and Kim, J.: How Well Do Coupled Models Simulate Today's Climate?, Bulletin of the American Meteorological Society, 89, 303–312, doi:10.1175/BAMS-89-3-303, https://doi.org/10.1175/BAMS-89-3-303, 2008.

Sein, D. V., Danilov, S., Biastoch, A., Durgadoo, J. V., Sidorenko, D., Harig, S., and Wang, Q.: Designing variable ocean model resolution based on the observed ocean variability, Journal of Advances in Modeling Earth Systems, 8, 904–916, doi:10.1002/2016MS000650, http://dx.doi.org/10.1002/2016MS000650, 2016.

Sein, D. V., Koldunov, N. V., Danilov, S., Wang, Q., Sidorenko, D., Fast, I., Rackow, T., Cabos, W., and Jung, T.: Ocean Modeling on a Mesh With Resolution Following the Local Rossby Radius, Journal of Advances in Modeling Earth Systems, 9, 2601–2614, doi:10.1002/2017MS001099, http://dx.doi.org/10.1002/2017MS001099, 2017.

Sidorenko, D., Wang, Q., Danilov, S., and Schröter, J.: FESOM under coordinated ocean-ice reference experiment forcing, Ocean Dynamics, 61, 881–890, 2011.

Sidorenko, D., Rackow, T., Jung, T., Semmler, T., Barbi, D., Danilov, S., Dethloff, K., Dorn, W., Fieg, K., Goessling, H. F., Handorf, D., Harig, S., Hiller, W., Juricke, S., Losch, M., Schröter, J., Sein, D. V., and Wang, Q.: Towards multi-resolution global climate modeling with ECHAM6–FESOM. Part I: model formulation and mean climate, Climate Dynamics, 44, 757–780, doi:10.1007/s00382-014-2290-6, http://dx.doi.org/10.1007/s00382-014-2290-6, 2015.

Steele, M., Morley, R., and Ermold, W.: PHC: A Global Ocean hydrography with a high-quality Arctic Ocean, J. Climate, 14, 2079–2087, 2001.

Sterl, A., Bintanja, R., Brodeau, L., Gleeson, E., Koenigk, T., Schmith, T., Semmler, T., Severijns, C., Wyser, K., and Yang, S.: A look at the ocean in the EC-Earth climate model, Climate Dynamics, 39, 2631–2657, doi:10.1007/s00382-011-1239-2, https://doi.org/10.1007/s00382-011-1239-2, 2012.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., Brokopf, R., Fast, I., Kinne, S., Kornblueh, L., Lohmann, U., Pincus, R., Reichler, T., and Roeckner, E.: Atmospheric component of the MPI-M Earth System Model: ECHAM6, J. Adv. Model. Earth Syst., doi:10.1002/jame.20015, 2013.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bull. Am. Meteorol. Soc., 93, 485–498, 2012.

Timmermann, R., Danilov, S., Schröter, J., Böning, C., Sidorenko, D., and Rollenhagen, K.: Ocean circulation and sea ice distribution in a finite element global sea ice–ocean model, Ocean Modell., 27, 114–129, 2009.

Valcke, S.: The OASIS3 coupler: a European climate modelling community software, Geosci. Model Dev., 6, 373–388, doi:doi:10.5194/gmd-6-373-2013, 2013.

Valcke, S., Craig, T., and Coquart, L.: OASIS3-MCT User Guide, OASIS3-MCT 2.0, Technical Report, TR/CMGC/13/17, CERFACS/CNRS SUC URA No 1875, Toulouse, France, 2013.

Visbeck, M., Marshall, J., Haine, T., and Spall, M.: Specification of Eddy Transfer Coefficients in Coarse-Resolution Ocean Circulation Models, Journal of Physical Oceanography, 27, 381–402, doi:10.1175/1520-0485(1997)027<0381:SOETCI>2.0.CO;2, https://doi.org/10.1175/1520-0485(1997)027<0381:SOETCI>2.0.CO;2, 1997.

von Storch, J.-S., Haak, H., Hertwig, E., and Fast, I.: Vertical heat and salt fluxes due to resolved and parameterized meso-scale ddies, Ocean Modelling, 108, 1 – 19, doi:https://doi.org/10.1016/j.ocemod.2016.10.001, http://www.sciencedirect.com/science/article/pii/S1463500316301111, 2016.

Wang, C., Zhang, L., Lee, S.-K., Wu, L., and Mechoso, C. R.: A global perspective on CMIP5 climate model biases, Nature Clim. Change, 4, 201–205, http://dx.doi.org/10.1038/nclimate2118, 2014a.

Wang, Q., Danilov, S., and Schröter, J.: Finite element ocean circulation model based on triangular prismatic elements, with application in studying the effect of topography representation, J. Geophys. Res., 113, 2008.

Wang, Q., Danilov, S., Sidorenko, D., Timmermann, R., Wekerle, C., Wang, X., Jung, T., and Schröter, J.: The Finite Element Sea Ice-Ocean Model (FESOM) v.1.4: formulation of an ocean general circulation model, Geoscientific Model Development, 7, 663–693, doi:10.5194/gmd-7-663-2014, http://www.geosci-model-dev.net/7/663/2014/, 2014b.

Wang, Q., Ilicak, M., Gerdes, R., Drange, H., Aksenov, Y., Bailey, D. A., Bentsen, M., Biastoch, A., Bozec, A., Böning, C., Cassou, C., Chassignet, E., Coward, A. C., Curry, B., Danabasoglu, G., Danilov, S., Fernandez, E., Fogli, P. G., Fujii, Y., Griffies, S. M., Iovino, D., Jahn, A., Jung, T., Large, W. G., Lee, C., Lique, C., Lu, J., Masina, S., Nurser, A. G., Rabe, B., Roth, C., y Mélia, D. S., Samuels, B. L., Spence, P., Tsujino, H., Valcke, S., Voldoire, A., Wang, X., and Yeager, S. G.: An assessment of the Arctic Ocean in a suite of interannual CORE-II simulations. Part I: Sea ice and solid freshwater, Ocean Modelling, 99, 110 – 132, doi:https://doi.org/10.1016/j.ocemod.2015.12.008, http://www.sciencedirect.com/science/article/pii/S1463500315002449, 2016a.

Wang, Q., Ilicak, M., Gerdes, R., Drange, H., Aksenov, Y., Bailey, D. A., Bentsen, M., Biastoch, A., Bozec, A., Böning, C., Cassou, C., Chassignet, E., Coward, A. C., Curry, B., Danabasoglu, G., Danilov, S., Fernandez, E., Fogli, P. G., Fujii, Y., Griffies, S. M., Iovino, D., Jahn, A., Jung, T., Large, W. G., Lee, C., Lique, C., Lu, J., Masina, S., Nurser, A. G., Rabe, B., Roth, C., y Mélia, D. S., Samuels, B. L., Spence, P., Tsujino, H., Valcke, S., Voldoire, A., Wang, X., and Yeager, S. G.: An assessment of the Arctic Ocean in a suite of interannual CORE-II simulations. Part II: Liquid freshwater, Ocean Modelling, 99, 86 – 109, doi:https://doi.org/10.1016/j.ocemod.2015.12.009, https://www.sciencedirect.com/science/article/pii/S1463500315002450, 2016b.

Wu, W., Danabasoglu, G., and Large, W.: On the effects of parameterized Mediterranean overflow on North Atlantic ocean circulation and climate, Ocean Modelling, 19, 31–52, doi:https://doi.org/10.1016/j.ocemod.2007.06.003, https://doi.org/10.1016/j.ocemod.2007.06.003, 2007.

Zadra, A., Williams, K., Frassoni, A., Rixen, M., Adames, Á. F., Berner, J., Bouyssel, F., Casati, B., Christensen, H., Ek, M. B., Flato, G., Huang, Y., Judt, F., Lin, H., Maloney, E., Merryfield, W., van Niekerk, A., Rackow, T., Saito, K., Wedi, N., and Yadav, P.: Systematic Errors in Weather and Climate Models: Nature, Origins, and Way Forward, Bulletin of the American Meteorological Society, 0, null, doi:10.1175/BAMS-D-17-0287.1, https://doi.org/10.1175/BAMS-D-17-0287.1, 2017.

Zhang, R. and Vallis, G. K.: The Role of Bottom Vortex Stretching on the Path of the North Atlantic Western Boundary Current and on the Northern Recirculation Gyre, Journal of Physical Oceanography, 37, 2053–2080, doi:10.1175/JPO3102.1, https://doi.org/10.1175/JPO3102.1, 2007.