Reviewer 1

This is my first review for this paper. The authors present the EC-Earth3P-VHR model configuration, a high-resolution global climate model developed for HighResMIP, featuring atmospheric resolution of about 16km and 8km oceanic resolution. The model shows improvements in key regions like the Gulf Stream and the Equator compared to lower resolution models, with reduced biases in some areas but increased biases in others, such as a larger warm bias over the Southern Ocean. The model also demonstrates better air-sea coupling in tropical regions. However, it tends to overestimate the oceanic influence on atmospheric variability at mid-latitudes. Overall, the EC-Earth3P-VHR configuration appears to offer enhanced opportunities to study climate variability and change on regional and local scales.

First of all, the paper is in my view well-written, understandable, and has basically no typos. The figures are all high quality and well done. A description of the EC-Earth configuration for HighResMIP is clearly within the scope of GMD.
We thank the Reviewer for the appreciation on our work. In the following, we answer each specific point (in blue).

My only minor comments are with respect to highlighting some key results better, and better embedding the study into previous work, also outside of Europe. I have provided some references below for that purpose that the authors can decide to include at their convenience, and also gave suggestions for potential additional figures that could make the study even stronger. Overall, the study in its present form is already very interesting, it lists key results that are encouraging for fellow coupled high-res modellers, and is worthy of prompt publication. I am providing line-by-line comments below that I'd suggest having included before the paper can be accepted.

##################################

Line-by-line comments:

l.59-60 I would suggest to cite relevant papers for these projects, for example Hoffmann et al. 2023 (https://doi.org/10.1016/j.cliser.2023.100394) for Destination Earth or Hohenegger et al. 2023 (https://doi.org/10.5194/gmd-16-779-2023) and Rackow et al. 2024 (already cited elsewhere in the study) for nextGEMS. There should be something from MetOffice for PRIMAVERA as well
The references have been added to the reviewed manuscript.

l.64 have been "proven"
Modified, although both proved and proven can be used.

l.84 For single precision, there are other earlier examples, e.g. Váňa et al. 2017 (https://doi.org/10.1175/MWR-D-16-0228.1); as Destination Earth and nextGEMS were listed, there is also Sarmany et al. 2024 (https://doi.org/10.1145/3659914.3659938) for IO considerations
Both references have been added to the reviewed manuscript.

l.91 Another extreme example next to Gutjahr et al. 2019 is AWI's CMIP6 climate model (e.g. Rackow et al. 2019, https://doi.org/10.5194/gmd-12-2635-2019 , Semmler et al. 2020 https://doi.org/10.1029/2019MS002009), where their ocean is locally finer than 10km and has been run with a 100km atmosphere. There might be other examples in CMIP

The references have been added to the reviewed manuscript.

l. 104 Regarding "High-resolution modelling usually relies on single-model component":
I think there are several examples for relatively high resolution in both components, e.g. Chang et al. 2020 (https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2020MS002298), Small et al. 2014 https://opensky.ucar.edu/islandora/object/articles:14403 , the high-res studies from the South Korean group (https://ibsclimate.org/research/ultra-high-resolution-climate-simulation-project/ and listed references there, e.g. https://www.science.org/doi/10.1126/sciadv.abd5109), and the DestinE, EERIE, PRIMAVERA and nextGEMS results of course as well

We agree with the Reviewer that there are indeed examples of high-resolution modeling in both the atmosphere and ocean. However, many of these examples are relatively recent (from 2020 on). The point the paragraph tries to make is that a part of the community has relied and still relies on forced single-model components (e.g., atmosphere-only and ocean-only models), and sometimes regional configurations, for their high-resolution studies, because either they do not have the computing resources, or because those configurations are more appropriate to address their research questions. Some of these studies are as recent as the ones listed by the Reviewer (e.g., Swingedouw et al., 2022) These approaches come to a cost, nonetheless, which is usually i) relying on boundary conditions generated by low-resolution models or observations, and ii) lacking a global perspective and climate model interactions. The introduction to the paragraph has been modified to make our intentions clearer.

References:

Swingedouw, D., Houssais, M.N., Herbaut, C., Blaizot, A.C., Devilliers, M. and Deshayes, J.: AMOC recent and future trends: a crucial role for oceanic resolution and Greenland melting?. Frontiers in Climate, 4, 838310, https://doi.org/10.3389/fclim.2022.838310, 2022.

l. 137 "of the" -> "of"

Modified.

l. 138 Is this OpenIFS or IFS?

As indicated in the manuscript, it is IFS in its 36r4 cycle.

l. 161 240 s and 720 s has large white spacing

Modified.

l. 198 What are the novel source code changes?

The section has been reformulated in the revised manuscript, with more details about the changes in the source code.

l. 200 Which model workflow software?

We refer to the Auto-EC-Earth model workflow, which is used to run simulations at BSC. This was added to the revised manuscript.

l. 204/205 Is there a parallel version of this available now? Is this linked as part of your document?
The coupler model version, OASIS-MCT 3.0, did not have parallelization in weight generation at that time. More recent versions of OASIS3 (such as the one in EC-Earth4) support OpenMP work sharing. This speeds up the process and allows EC-Earth4 to couple meshes on the fly. This is now indicated in the revised manuscript.

l. 209 Can you write technical details of "the network" here? Otherwise this does not tell much.
Added

l. 279-281 This inconsistency could trigger a bigger adjustment potentially. From your experience, does this lead to a cooling or warming initially that gets levelled out during the spinup?
The difference between temperature potential temperature and conservative temperature is very close to zero for the range of salinity values in the ocean (see Figure 2 in McDougall et al., 2021). Practical salinity can take lower values than absolute salinity (~0.1 psu as in Pawlowicz, 2013). The adjustment to this difference, however, will depend on the preferred state of the model. In regions that tend to show negative salinity biases, initializing from a lower salinity values might lead to a smaller adjustment. Regions with positive salinity biases might, by contrast, would show stronger adjustments. The global effect is in the end difficult to assess. Nonetheless, and as shown for the second Reviewer, the drift in the globally averaged ocean salinity is no larger than 0.00005 psu/yr in the last 50 years of the control simulation in the three configurations, being the smallest in VHR. This result suggests that any potential adjustment due to the incorrect initial conditions is soon levelled out during the spin-up phase.

References
McDougall, T. J., Barker, P. M., Holmes, R. M., Pawlowicz, R., Griffies, S. M. and Durack, P. J.: The interpretation of temperature and salinity variables in numerical ocean model output and the calculation of heat fluxes and heat content. Geoscientific Model Development, 14(10), 6445–6466, https://doi.org/10.5194/gmd-14-6445-2021, 2021.
Pawlowicz, R.: Key physical variables in the ocean: temperature, salinity, and density. Nature Education Knowledge, 4(4), 13, 2013.

l. 296 "but for" -> "except for"
Modified.

l. 310 "will enjoy" please rephrase
Done.

l. 314/315 This is a very encouraging result; in principle, this is what modelling groups have been hoping for to see with higher resolution. This point could maybe be highlighted, potentially with a figure, and be included in the abstract?
These lines describe the results of Figure 3; we therefore do not think an additional figure is needed. Nonetheless, the result has been mentioned in the Abstract, as suggested.

l. 389 delete “bias” in this line?
Removed.

l. 401 40 deg N seems far away from polar influence, are you sure about this statement?
The linkage is suggested between the warm subsurface bias and the sea ice anomaly at subpolar latitudes in the Labrador Sea, between 45–60 °N. The connection is therefore less remote. As explained in the manuscript in the lines following that, the excess of sea ice weakens the surface oceanic circulation (subpolar gyre), leading to a heat accumulation to the south near the surface in the intergyre region (at around 40 °N). Also, the excess of sea ice and the associated weak oceanic deep mixing in the Labrador Sea would reduce the vertical exchange between surface waters, which are cooled down by the atmosphere, and the subsurface, leading to an anomalous warming as well.

Figure 13: Hard to see anything on those maps. Could you try with other colors or try a different (shorter) range?
The Figure has been modified to increase clarity.

l. 459/460 The plus/minus refers to what, standard deviation of monthly values?
Clarified.

l. 493-495 This seems like another key result that is very encouraging and not covered with a separate figure.
This result refers to Figure 4, which is now indicated in the revised manuscript. Although we could add another figure, zooming in over certain mountain regions, it would not add additional information to that in Figure 4. Therefore, and for the sake of conciseness, we have decided to refrain from doing so. This result, nonetheless, has been highlighted in the Abstract and Conclusions.

l. 535 Sections -> sections
Changed.

l. 561 The lack of ocean current feedback comes a bit as a surprise here and could be covered earlier in the model description as to how the coupling is implemented
Added to the reviewed manuscript.

l. 580/581 Maybe give another example here if you know it (e.g. US or South Korean references mentioned above if matching), but this statement might indeed be correct
References have been added for completeness.

l. 585 excessive whitespace after “performed”
Changed.

l. 619/620 Another study in this direction would be Sein et al. 2017 (https://doi.org/10.1002/2017MS001099). The authors argue that resolution over the shelf areas

northward of the Gulf Stream front is key, an area where the cold Labrador water from the north meets the warm Gulf Stream water.
The reference has been included, and the text updated.

l. 687/688 “and VHR does it faster and appears more stable after 100 years than HR and LR” Another key result, see above
We agree. This result is now highlighted in the Abstract.

Reviewer 2

The paper introduces and validates a high-resolution configuration of the EC-Earth model, utilizing the IFS T1279 atmosphere and NEMO ORCA12 ocean components. The validation confirms that the model is technically ready for use and, in many aspects, shows reduced biases compared to its lower-resolution predecessors, although some tuning is still required, particularly for the Southern and Arctic Ocean regions. The authors definitely did a great job. I therefore recommend this paper for acceptance after minor revisions, which I have summarized below.
We thank the Reviewer for the comments on our work. In the following, we answer each specific point (in blue).

My main concern is that the plots should be updated to meet the standards of the 202Xs. This is especially relevant for Figures 7 and 14. Based on visualizations I have seen on YouTube and in several presentations, the presented plots appear outdated.
In our opinion, figures follow many of the current standards in plot making, such as clean and concise design, accessible color schemes, clear labels, and high data-ink ratio. Nonetheless, we have modified certain figures as requested by both Reviewers, either to increase readability, or to add new information, or both. High-quality rendered figures will also be provided during the publications process.

L43: Although the message is clear I would reformulate this sentence which currently states that biases lead to something improved.
Clarified.

L155: Regarding the effective resolution for the ORCA grid, could you cite the appropriate reference or tell how you compute it?
A reference has been added.

L166: independent on which grid, atmospheric or ocean? In general which type of remapping between ocean and atmosphere do you use? Is it conservative? Do you conserve the global freshwater balance? If not, how large is the imbalance?
Coupling variables are communicated between the different model components via the OASIS3-MCT coupler, using the parallel SCRIP library as interpolation method. Some variables are conservative, to ensure global freshwater and energy conservation. The conversion process requires MPI collective communications, to fix potential imbalances and ensure that any possible imbalance is negligible over

time. This approach is described in detail for the CMIP6 configuration of EC-Earth (Döscher et al., 2022). The reference has been added to the reviewed manuscript.

References

Döscher, R., et al.: The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6, Geosci. Model Dev., 15, 2973–3020, https://doi.org/10.5194/gmd-15-2973-2022, 2022.

I am also curious if you couple the ocean velocities for computation of the wind stress as this could explain some of the biases. Also I would list the number fluxes and surface fields which are being exchanged.

The fields that are passed from the atmosphere to the ocean are:

- Momentum fluxes for the ocean and sea ice on the U and V grids of the NEMO model.
- Non-solar and solar radiation over the ocean, total evaporation, and precipitation (which is conserved).
- Solar and non-solar radiation over sea ice, the sensitivity to temperature of non-solar energy flux over sea ice, and evaporation over sea ice (which is not conservative).

These are the ocean fields passed to the atmosphere:

- SST; temperature, albedo, fraction, and thickness of the sea ice; and snow thickness over ice.

Such coupling characteristics are the same as those described for the CMIP6 version of EC-Earth, which are described in Döscher et al. (2022) and thus do not include ocean velocities.

References

Döscher, R., et al.: The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6, Geosci. Model Dev., 15, 2973–3020, https://doi.org/10.5194/gmd-15-2973-2022, 2022.

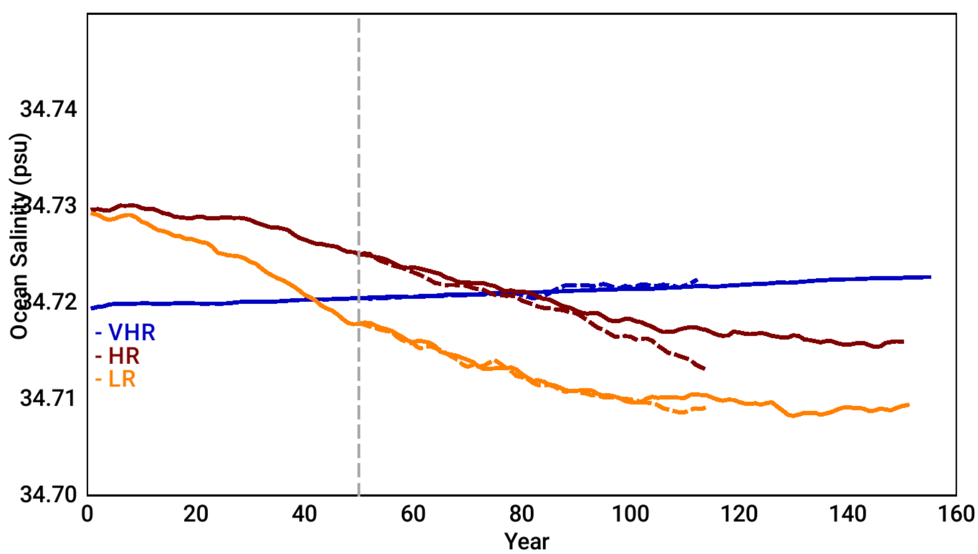
Fig. 2 Does this include the I/O cost? How many cores were assigned to OIFS, and how many to NEMO? What would the scaling for the individual uncoupled components look like, or at what number of grid points per core would scaling saturate according to the technical reports of these two models? If I understand correctly, for a 50x50 distribution in NEMO, this would be (4322*3059)/(4800/2) > 5,500 grid points per core. Given that NEMO scales linearly until around ~500 grid points per core, the saturation is reached much earlier than expected, possibly due to coupling issues or I/O. Is this correct?

We thank the Reviewer for pointing this out. The results do include the I/O cost, and that is one of the reasons why it does not fully scale. The saturation is reached much earlier than expected, because of the coupling (exchange between atmosphere and ocean due to waiting times and costly conservative exchanges), and also because of the IFS I/O approach. In EC-Earth3, while NEMO uses a parallel and asynchronous I/O server (XIOS), IFS cycle36 does not use it, which leads to the model treating the flush of the output serially. This is computationally more expensive and the reason why IFS saturates sooner. We therefore consider we do not need an independent scalability test for each model component.

L303: You say here that the spin up for 50 years is insufficient but from the plot we clearly see that even 150 years are insufficient. In Fig. 3 (lower panel) I note that the warming drift in VHR is the smallest for the global ocean but it is the largest for the upper ocean (upper panel). Does this imply that the overly large warming drift in the upper ocean is compensated by a negative drift in the deep ocean?

We agree that the extent of the spin-up is insufficient to equilibrate the full ocean, as discussed in the Discussion section. Concerning the drifts in VHR, in both the global ocean (Fig. 3b) and the upper-100-m (Fig. 3a), VHR has the smallest drifts (solid blue lines) out of the three configurations. This is also described with trend values in the manuscript, at the beginning of Section 3.1: "Across all three model resolutions, the length of the spin-up (50 years) appears to be insufficient to equilibrate the full ocean (Fig. 3b); in fact; the ocean temperature is still drifting about 0.001–0.002 °C/yr towards warmer conditions at the end of the control simulation (computed over the last 50 years) in the three configurations. In the upper ocean, however, VHR shows a substantially smaller warming drift than the other two configurations: about 0.00005 °C/yr compared to 0.0025 °C/yr and 0.0062 °C/yr in HR and LR, respectively (computed over the last 50 years; Fig. 3a). "

Regarding my previous comment, is there drift in global salinity or the system is conserved?
The Figure below shows the evolution of the mean salinity in the upper 100 m and in the whole ocean in the spin-up, control, and historical simulations. The drift in the full ocean in the three configurations is no larger than 0.00005 psu/yr in the last 50 years. VHR again shows the smallest drift out of the three. The result points to very small adjustment in the ocean at the end of the control period, similar or smaller than the changes over the historical period (dashed lines). Although we have not added this extra figure to the revised manuscript, the results are now mentioned in Section 3.1.



Response Figure 1. Mean global oceanic salinity (in psu) in the LR (yellow), HR (red), and VHR (blue) models in the spin-up runs (0–50-year period), control runs (50–150-year period; solid lines), and historical runs (50–114-year period; dashed lines). The vertical dashed line marks the end of the spin-up period.

L342: While the precipitation bias over the ocean, especially the "dipole structure" bias, will be mixed by ocean eddies, likely resulting in a low ocean salinity bias, the precipitation bias over land, especially in northeastern South America, will lead to heavily reduced runoff from land, as seen in the Amazon River. I observe this bias in all three simulations. What is the origin of this bias? Is it also present in other CMIP-type models?
Monteverde et al. (2022) study the precipitation biases in CMIP6 models over the Amazonia, including the CMIP6 version of EC-Earth. They find that models fail to reproduce the seasonal cycle in

precipitation, as well as extreme events. The three configurations of EC-Earth show a similar bias in their precipitation, which suggest similar mechanisms might be at play. This has been added to the reviewed manuscript.

References

Monteverde, C., De Sales, F. and Jones, C.: Evaluation of the CMIP6 performance in simulating precipitation in the Amazon River basin. Climate, 10(8), 122, https://doi.org/10.3390/cli10080122, 2022.

L408: I also observe large deep mixing which takes place in the Arctic Ocean north of the Bering Strait. Seems like the Arctic Ocean is well mixed there which shall not be the case according to my knowledge. Nevertheless the ice patterns look nice but you didn't provide the sea thickness. Could you augment Fig. 7 with sea ice thickness such that the story becomes more consistent. As you say thick ice and the excessive brine rejection are the likely reasons for what is happening.

L654: In my opinion, the negative aspects are also strongly reflected in the biases observed in the Arctic Ocean, which should be addressed in the discussion section.

We answer the last two comments together, because they are related. The Reviewer is right in pointing to the excess of ocean mixing in the Arctic in VHR. As proposed, we have updated Figure 7, so now the shading represents the bias in sea ice thickness in VHR with respect to PIOMAS. As we discussed in the initial manuscript, the sea ice excess could be connected with the excess of ocean mixing in the central Arctic. Too thick ice would lead to overly brine rejection, explaining the positive surface salinity bias. Since more salinity makes the waters denser, this will lead to a less stratified water column, which is easier to mix. The results are now also highlighted in the Discussion Section and Abstract.