This manuscript can be viewed as a continuation of two previous papers by the same group Matthes et al. (2020) and Martin and Biastoch (2023). The main novelty lies in the use of a new atmospheric model, two different ocean resolutions, and an improved coupling strategy. The authors quantify the impact of ocean resolution (within a regional refinement) on the mean state, the AMOC, and the northward heat transport in the North Atlantic but the analyse of the atmospheric contribution and the explanation of the physical mechanisms is weak.

Overall, I believe that this work could be published in *Geoscientific Model Development* (GMD), but only after major revisions. My main comments are detailed below.

A) Technical aspects

The use of a new atmospheric model together with a new coupling interface clearly falls within the scope of GMD. However, several important technical aspects require a more detailed description.

1) Model configurations

Although the OpenIFS and NEMO configurations have been described in previous publications, I find the description of configurations in the present manuscript far too brief.

- A summary of the main physical parameterizations used in the atmospheric model would be welcome.
- For the ocean model, Eqs. (2) and (3) suggest that viscosity is represented using a biharmonic operator whereas diffusion uses a harmonic operator. This should be stated explicitly. It should also be specified whether these operators are applied horizontally or along isopycnal surfaces.
- The mesh refinement factor within the AGRIF zoom is five, whereas the time-step refinement is only three (instead of five). How do the authors explain that such a large time step remains numerically stable? Have they observed any numerical noise within the refined region, for example near the ocean bottom or along the Gulf Stream?
- Matthes et al. (2020) report using a Gent-McWilliams coefficient of 2000 $m^2 s^{-1}$, whereas the present manuscript states that 1000 $m^2 s^{-1}$ was used (as in Martin and Biastoch, 2023). Which value was actually used in this study?

2) Coupling interface

The description of the coupling interface is also insufficient.

At the very least, the manuscript should list all variables exchanged between the atmosphere and the ocean, especially in the presence of sea ice. How is the coupling with sea ice performed? For example, is sea-ice thermodynamics computed within the ocean or the atmospheric model? Which strategy is used to ensure flux conservation in partially ice-covered regions where the ice fraction evolves during a coupling time step?

Lines 161–162: "*... fields to the OASIS coupler once per model time step and the fields are then sent from OASIS to both NEMO and AGRIF.*"
This wording is misleading and could suggest that OASIS is run as a standalone executable, as was the case for versions earlier than OASIS3-MCT. In reality, OpenIFS accumulates coupling fields at every model time step through the OASIS library without MPI communication. MPI exchanges between OpenIFS and NEMO, mediated by OASIS, only occur at the coupling frequency, not at every model time step.

3) Interpolation schemes

The authors use interpolation schemes that are rather unusual, as they preserve neither integral quantities nor derivatives.

- As a consequence, heat and freshwater fluxes (evaporation minus precipitation) are not strictly conserved across the air-sea interface. Have the authors evaluated the resulting losses of mass and energy through the coupling interface? For example, comparing maps of the fluxes computed on the atmospheric and ocean grids would allow the geographical distribution of positive and negative interpolation errors to be quantified.
- Likewise, the divergence and, more importantly, the curl of the wind stress, which play a critical role in driving the ocean circulation, may be significantly degraded by the interpolation method used in this study. Have the authors compared the wind stress curl over the North Atlantic computed on the atmospheric and ocean grids?

B) Scientific results

1) Atmospheric analysis

Most of the analyses focus on the ocean, although the simulations are fully coupled. The manuscript lacks a thorough analysis of the atmospheric response. For example example lower atm temperature bais, e.g. at 850hPa compared to ERA5 could be added to Fig 2. The comparison between LH and HH should clearly identify the atmospheric processes responsible for the differences between these two simulations, which differ only in the resolution of their atmospheric component.

On several occasions, the authors speculate about the underlying mechanisms even though the atmospheric model output could be used to test these hypotheses directly:

- Line 234: the differences are "*likely*" attributed to cloud radiative forcing. Why not simply show the cloud radiative forcing? Alternatively, analyze separately the upward and downward radiative components, or the shortwave and longwave fluxes at the top and bottom of the atmosphere.
- Line 241: the authors state that they "*speculate*". Why not plot the polar jet and verify whether this hypothesis is correct?
- Figure 9: the net surface heat flux combines several different physical processes into a single quantity. It would be much more informative to present the turbulent heat fluxes and the net radiative fluxes separately. Furthermore, separating shortwave and longwave radiation would greatly help the interpretation, particularly over sea-ice regions where albedo changes play an important role.

2) Initialization and model drift

Unless I missed it, I could not find any description of how the ocean initial state was generated. Was it taken from a long coupled simulation, as in Martin and Biastoch (2023)? From the simulations of Kjellsson et al. (2020)? Or from an observational climatology such as the World Ocean Atlas? In any case, the simulations presented here are probably too short to robustly evaluate the deep ocean circulation. The model is most likely still undergoing a significant adjustment phase (see, for example, Griffies et al., 2025). I fully understand that the authors may not have had sufficient computational resources to perform a much longer spin-up. However, this issue cannot simply be dismissed by the brief statement given in lines 195–197.

As stated by the authors, no retuning was performed for the HH simulation, while the top-of-atmosphere radiative balance differs from the other simulations (line 236). This strongly suggests that the HH experiment is not in radiative equilibrium and is therefore drifting, potentially in a different manner from LL and LH. Under these circumstances, it is difficult to support the statement that there is a "*common drift in all simulations*" (line 197).

At the very least, the manuscript should present the top-of-atmosphere radiation budget for each simulation together with time series of several key diagnostics, for example the mean atmospheric temperature between 500 and 100 hPa north of 60°N, or the mean ocean temperature between 500 and 1000 m south of 40°N.

3) Mediterranean Outflow

As stated in the manuscript, the width of the Gibraltar Strait in ORCA05 is fixed to the observed value (20 km).bFigure 5 appears to suggest that the Gibraltar Strait is unrealistically deep in ORCA05. This is not correct. Its depth (284 m, if I am not mistaken) is also constrained to the observed value thanks to partial bottom cells, meaning that the geometry of the strait is already realistic even in ORCA05. Although the LH bathymetry is more detailed, the effective cross-sectional area of the strait—which primarily controls the volume transport—should remain very similar in LL and LH. Moreover, the lateral friction coefficient can also be adjusted to reproduce the observed Mediterranean outflow.

Therefore, the lower resolution of ORCA05 alone cannot explain the "*unrealistically large outflow.*" The results shown here appear to reflect model (wrong) tuning at a given resolution rather than a genuine resolution effect.

4) Absence of current feedback

As explained in lines 108–109, the coupling interface does not account for the current feedback. This is surprising given that the objective of the study is precisely to investigate the impact of ocean mesoscale resolution on the coupled system.

From a technical point of view, including ocean currents in the wind-stress calculation is straightforward to implement and computationally inexpensive. Numerous studies have demonstrated the importance of this process, making its omission increasingly difficult to justify, particularly in a study focusing on the effects of high ocean resolution.

Ocean modelers often dislike this process because it damps mesoscale eddies, whose explicit simulation is computationally expensive. However, can one justify neglecting a physically important process simply because doing so produces a compensating error and artificially maintains stronger eddy activity?

Finally, a minor but important point raises concerns about the care taken in preparing the manuscript. The citation to Zhai et al. (2010) in line 109 is incorrect. That paper does not discuss the *eddy-killing* effect associated with including ocean currents in the computation of wind stress.

5) Relation to previous studies

Several previous studies are closely related to the present work. The authors cite Matthes et al. (2020) and Martin and Biastoch (2023), but they should also cite Taillandier et al. (2014), which was a pioneering study on this topic, albeit using forced rather than coupled simulations.

Many of the differences found between LL and LH, particularly regarding the AMOC and northward heat transport, had already been reported in Taillandier et al. (2014) and Matthes et al. (2020) for example. The authors should therefore discuss much more explicitly what is genuinely new in the present study. Are these simulations simply confirming previous findings, or do they provide new insights? More specifically, what scientific advantages does OpenIFS provide compared with ECHAM6?

The discussion would benefit from explicit comparisons with equivalent figures in previous studies, for example:

- Figure 2 with Figure 7 of Taillandier et al. (2014) and Figure 3 of Martin and Biastoch (2023).
- Figure 8 with Figure 6 of Taillandier et al. (2014).
- Figure 6 with Figures 15 and 16 of Taillandier et al. (2014).

Griffies, S. M., Adcroft, A., Beadling, R. L., Bushuk, M., Chang, C.-Y., Drake, H. F., et al. (2025). The GFDL-CM4X Climate Model Hierarchy, Part I: Model Description and Thermal Properties. *Journal of Advances in Modeling Earth Systems*, *17*(10), e2024MS004861. https://doi.org/10.1029/2024MS004861

Talandier, C., Deshayes, J., Treguier, A.-M., Capet, X., Benshila, R., Debreu, L., et al. (2014). Improvements of simulated Western North Atlantic current system and impacts on the AMOC. *Ocean Modelling*, *76*, 1–19. https://doi.org/10.1016/j.ocemod.2013.12.007