Response to Referee Comment 1 (RC1)

egusphere-2026-1289, Ghosh et al.

We thank the reviewer for the positive assessment and the helpful questions. Below, the reviewer comments are given in blue and our responses in black. Line numbers refer to the discussion paper.

1. Could you devote a paragraph to other groups with similar projects? ICON is one, but there are probably also efforts in Japan, the US and China.

We have added a paragraph to the Introduction describing comparable efforts: ICON-Sapphire, nextGEMS and Destination Earth in Europe, the NICAM global cloud- and storm-resolving work in Japan, the E3SM/SCREAM and CESM high-resolution efforts in the United States, and the eddy-resolving LICOM and FGOALS configurations in China. The paragraph ends by stating how our simulations are complementary, namely that they combine a kilometre-scale atmosphere with an eddy-rich ocean in a coupled configuration integrated over a centennial period.

Changes: new paragraph in the Introduction, after line 119.

2. SST is clearly improved compared to the 'standard' 1 deg. setups, but with precipitation I am not quite sure. Could you add one extra panel for SST and precipitation each, that shows the IFS-FESOM biases in a coarser set-up, like 1/2 or 1 degree?

We agree that this would be the most direct demonstration of the benefit of resolution. Unfortunately, no lower-resolution IFS-FESOM historical simulation performed under the same protocol is currently available, and producing one is not feasible within this revision. The closest available comparison is the performance index in Table 1, which compares the simulation with the CMIP6 multi-model mean for both near-surface temperature and precipitation. It shows that precipitation is indeed among the fields where the improvement is smallest, in line with the reviewer's impression. We have made this limitation explicit in the Discussion:

> "In particular, no lower-resolution IFS-FESOM simulation performed under the same protocol is currently available, so the comparisons with lower-resolution models presented here rely on the CMIP6 multi-model mean (Table 1) and on published HighResMIP results rather than on a controlled resolution experiment."

Changes: limitations paragraph of the Discussion, lines 1109–1112.

3. It appears as if the model is simply switched on and then analyzed, but it probably required years of engineering to get there. For the non-modeller, could you describe some of the key steps?

We agree that this was far from a matter of simply switching the model on. Most of the underlying development of the coupled IFS-FESOM system at kilometre scale, including the single-executable coupling, the ocean mesh and the MultIO/FDB output infrastructure, is documented in detail by Rackow et al. (2025), and we now state this explicitly and credit that work. On top of this, the present centennial-scale experiment required additional work, which we now describe in a new paragraph at the end of Section 2.5: installing and optimising the model on the JUWELS and Levante systems; addressing limitations of the IFS for integrations of this length, in particular in the numerical representation of model time and the associated restart handling, which only became apparent once the simulations extended well beyond the multi-year runs the configuration was originally designed for; adapting the MultIO output configuration to the variables and frequencies needed for climate analysis; and developing procedures to generate the initial conditions for the control, historical and scenario simulations from the end of the coupled spin-up.

Changes: new paragraph at the end of Section 2.5, after line 254.

4. The FESOM grid promised a better Gulf Stream separation. Has this been achieved and could you be explicit about it?

Yes. We now state this explicitly in Section 4.5:

> "With respect to the Gulf Stream itself, the simulation shows a clear improvement over eddy-parametrized and eddy-permitting configurations: the current separates from the coast close to Cape Hatteras and the associated band of elevated SSH variability extends eastwards as a coherent jet, instead of continuing along the continental slope as in coarser FESOM configurations. The remaining deficiency is therefore not the separation point but the subsequent north-eastward pathway of the North Atlantic Current, which is too zonal."

Changes: lines 806–809.

5. There appears to be no AABW in the Atlantic. Could you show the global MOC as well? And how strong is the ACC?

We thank the reviewer for this comment. Figure 18a covers the Atlantic from about 35°S to 80°N and therefore shows the northward penetration of Antarctic Bottom Water into the Atlantic rather than its formation, which takes place around Antarctica, outside the domain shown.

Within this domain the abyssal cell is present rather than absent. Below about 3500 m the overturning streamfunction becomes negative, reaching a maximum strength of about −4 Sv between roughly 10°S and 40°N near 4000 m and weakening again below 5000 m. The cell is therefore weak and vertically confined, and it is less pronounced near the southern boundary, where AABW enters the basin, than in the subtropical North Atlantic. Two caveats apply: the streamfunction is computed in depth coordinates, in which the abyssal cell is generally less well defined than in density coordinates, and the deep inflow is concentrated in narrow boundary currents that a zonally integrated streamfunction resolves poorly.

Deep and bottom water formation is a long-standing difficulty in coupled models. No CMIP5 model assessed by Heuzé et al. (2013) exported dense shelf water as bottom water, most forming deep water by open-ocean convection instead, and in CMIP6 the majority of models still form deep and bottom water by open-ocean deep convection too deeply, too often or over too large an area (Heuzé, 2021). Our configuration shows this behaviour in the Weddell Sea, with excessively deep winter mixed layers (Section 4.6), and it does not represent dense overflows explicitly, so the realism of the abyssal cell should be interpreted with caution.

In the revised manuscript we add the global overturning streamfunction as a supplementary figure, which includes the Southern Ocean formation region and the full lower cell, and we report the simulated Drake Passage transport as a measure of the ACC strength, comparing it with the observational range of about 134–173 Sv (Whitworth, 1983; Koenig et al., 2014; Donohue et al., 2016) and the reconciled estimate of about 157 Sv of Xu et al. (2020).

Changes: new paragraph at the end of Section 4.8 (after line 1034), a new supplementary figure, and the Discussion (Section 5).

6. In all studies that I know, the tropical precip biases are similar in coupled and uncoupled mode, whereas in your model, the uncoupled IFS is almost bias free. Is this the result of skilled tuning, or are there key processes that only exist in IFS but not in other models?

We thank the reviewer for this question. We would first like to qualify the impression that the uncoupled IFS is almost bias free. Figure R1 shows the zonal-mean tropical precipitation in MSWEP, in the coupled historical simulation and in IFS-AMIP. The coupled simulation shows the classical double-ITCZ error: a southern maximum that is too strong and displaced towards 10°S, an equatorial minimum that is too dry, and a northern ITCZ that is too weak, so that the southern rainband becomes the stronger of the two (asymmetry index +0.33, against −0.16 in the observations, where the northern band dominates). In IFS-AMIP the profile follows the observations closely south of the equator and at the equator, and the asymmetry index is −0.02. The residual error is a northern ITCZ about 1 mm/day too weak, which leaves the two maxima more comparable in amplitude than observed. The double-ITCZ signature is therefore strongly reduced, but not removed, when the atmosphere is run with observed sea surface temperatures.

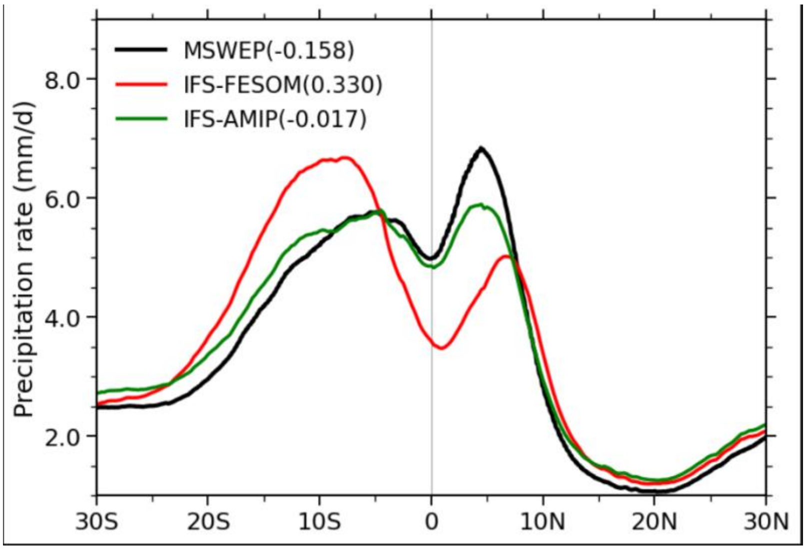

Figure R1. Annual-mean zonal-mean precipitation (mm/day) for the period 1985 to 2014, in MSWEP (black), the IFS-FESOM historical simulation (red) and IFS-AMIP (green). The value in brackets after each label is a simple asymmetry measure, defined here as the ratio of the maximum zonal-mean precipitation in 15°S-2°S to that in 2°N-15°N, minus one. It is zero when the two rainbands are equally strong, negative when the northern ITCZ dominates, as in the observations, and positive when the southern band dominates.

As to why the reduction is larger than in many earlier models, we can only offer a partial explanation. IFS-AMIP is forced with observed SSTs, so the cold-tongue and tropical Atlantic SST errors that drive the double-ITCZ response in the coupled model are absent by construction, whereas in the coupled simulation these errors develop freely over decades and the feedbacks that amplify the double ITCZ are fully active. In addition, the atmospheric component is a direct evolution of an operational forecasting model, whose convection and cloud schemes are continuously evaluated against observations at short lead times. We therefore interpret the remaining tropical precipitation bias in IFS-FESOM as predominantly a coupled-system error, while a weaker version of the bias is also present in the atmospheric component.

Changes: Section 4.3, after line 560, and a new supplementary figure.