Dear Authors,
Your paper has been submitted to a second round of reviews and there are still major concerns with your revised version. Therefore, before making my final decision on the publication or the rejection of your paper, I would like to have your feedback to the second round of reviews. I put here below the remarks made by the reviewers in the second round and I am asking you to reply point by point.

Reviewer 1:

This study focuses on the impact of horizontal resolution and model time step on extreme precipitation over Europe. The authors use AMIP-style simulations with the OpenIFS at various resolutions, covering a 25-year period (1979-2014). Results are compared with daily gridded precipitation observations data from GPCC and ERA5 reanalysis data. Although the authors did address some of the previous remarks/comments and adjusted parts in the revised manuscript, some of the original points raised have not been adequately addressed.

Specific comments:

1. The introduction is substantially improved and sharpened in the revised manuscript.

2. Lines 107-113: Although the authors responded with additional information to previous comment regarding the interpolation method used to convert the model output from native to regular grid, the manuscript has not been adapted to include this information. I would recommend that the authors adjust the manuscript to match their response to previous comments.

Regarding XIOS server output re-gridding, I could only find information in literature that the XIOS server is doing conservative (1st or 2nd) order interpolation, not bilinear interpolation. Could the authors please provide the relevant documentation on the interpolation methods of the XIOS server, either through appropriate referencing of documentation that includes the information on the interpolation methods or by making available the interpolation source code? That would substantially strengthen the reproducibility of the study.

3. I still have strong reservations regarding the use the ERA5 convective and large-scale precipitation as benchmark for comparing the HR, MR and LR convective and large-scale precipitation. Although Lavers et al. (2022) compares ERA5 total precipitation to observations and does indeed find that in Extratropics ERA5 it has lower biases than in the Tropics, there is no comparison in Lavers et al. (2022) against different model output simulations that would show that ERA5 is doing better than OpenIFS or any other model at similar resolution. Neither does Lavers et al. (2022) suggest in their manuscript that ERA5 large-scale precipitation (LSP) or convective precipitation (CP) can be used as proxies for observed large-scale or convective precipitation.

Since the authors own results in this study suggest that ERA5 is doing similarly or slightly (for precipitation percentiles higher than 99%) better than HR and MR experiments against observations, I would argue that assuming that LSP and CP precipitation from ERA5 is better than HR and MR experiments is not valid. Hence, I would recommend that the authors avoid the use of a metric such as RMSE, which hints to proper evaluation against a reliable benchmark, to compares LSP and CP precipitation between ERA5 and LR, MR and HR experiments. Instead, the authors can use and compare the precipitation distributions for LSP and CP and discuss about differences between LSP and CP across the different resolutions or for the different percentiles (as done in Figs S6-S7), without suggesting that ERA5 convective and large-scale precipitation can be used as a valid “observations” (i.e., removing Fig 6 and RMSE comparison in Figs 4,5).

4. Given that the focus of the study is on extreme precipitation, which can be quite sensitive to resolution and tends to occur very locally (e.g., scales less than ~1x1 degree) have the authors investigated whether the conclusions of the study change if precipitation is investigated in native resolution rather than the coarsely interpolated at ~1x1 degrees? Also, if the interpolation from native grid to the regular grid is indeed bilinear instead of conservative as the authors suggest in their response, this would mean that total precipitation in no longer conserved when interpolating to ~1x1 degree in MR and HR experiments. Have the authors checked how this affects the conclusions of this study? The impact of interpolation method on extreme precipitation, and impact of coarsening from native resolution to ~1x1 degrees, should ideally be discussed in the manuscript.

5. Lines 296-298. Assuming directly that convective precipitation decreases with increasing resolution may not be valid for all resolutions. The ratio between convective and large-scale precipitation will change, likely because large-scale precipitation tends to increase with increasing resolution. But so can total precipitation. It would be more accurate if the authors mention that the ratio between convective and large-scale precipitation can change as resolution increases. Also, the origin of large-scale precipitation would depend on the model’s effective resolution, hence it is not accurate to say that large-scale precipitation originates from synoptic storms without mentioning for which resolution this assumption may be valid.

6. Lines 432-434. What is the link here between the extreme precipitation in Europe and tropical cyclone representation? This is not investigated in the current study. Also, the study from Manganello et al. (2011) is rather old if the authors want to make a comparison with ECMWF-IFS system on tropical cyclone representation, as the current operational resolutions is 9 km.

<u>Reviewer 2 :</u>

The paper is scientifically sound and well written. If the following minor comments can be addressed, then I believe the paper is suitable for publication in GMD.

1) Introduction: The sensitivity of climate model performance to horizontal resolution and model timestep is presented as applying to all models. This is probably true to some extent, but the level of sensitivity varies considerably between models and this should be acknowledged. The IFS appearing to be amongst the more sensitive in this regard. I think it’s also work stating in the Introduction that the purpose of parametrizations is to represent processes which are not resolved (spatially or temporally), hence this sensitivity reflects a weakness in the formulation of those models.

2) Line 75: I don't think Jung et al. particularly comment on timestep sensitivity of precipitation. They indicate that the errors in the tropical circulation are smaller at 15 min than 60 min timesteps (opposite of what's written here).

3) Line 76: The Roberts et al. (2018) reference is not listed in the final Bibliography.

4) Section 3: The authors have a number of figures as supplementary material, however the number of figures in the main article is not excessive and the manuscript is reliant on the figures in the supplementary material for the reader to follow (e.g. paragraph starting line 257). I therefore suggest most/all of the figures in the supplementary material are moved to the main paper.

5) Line 297: Given that few convective systems are larger than 25km, I would question whether convective precipitation should decrease with increased resolution over the range of resolutions studied here – this would imply >25km^2 updraughts from explicit representation. (I would expect the resolution sensitivity as one moved to higher resolutions with grid lengths <20km). Does this point to a weakness in the convective parametrization?

6) Paragraph starting line 311: It should be noted in the manuscript that comparing the split between convective and large-scale precipitation in ERA5 is also justified because it shares the same convective parametrization as the IFS. It would be less appropriate for other models for which the definition of convective precipitation is an arbitrary choice and related to the formulation of the convective parametrization and what it handles versus the large-scale scheme (convective core?, anvils?, etc.).

7) Line 403/404: The three references should be listed in the same order as the three example models to which they relate. I also wonder if lines 400-406 would be better in the Introduction?

8) Line 429: The need to re-tune for different resolutions appears to be presented as fact, whereas if model parametrizations were appropriately scale aware, tuning for different resolutions shouldn't be needed.

9) Figure 1: Colors/line styles of lines in the figure don't match what's written in the caption.

10) Map plots (e.g. Figure 2): These would be better as block plots (i.e. each grid box is filled with a color) rather than using a filled coloring which can be misleading e.g. if single grid boxes are notably different to surroundings.

11) Figures 4&5: Given that the vast majority of extreme precipitation over northern Europe is large scale, and over southern Europe is mostly large-scale in DJF, and it is the large-scale precipitation which shows the strongest resolution dependence of the RMSE, would it be better to plot Figures 4 and 5 as lsp/tp?