Response to reviewers

Dear Editor and reviewers,

We appreciate your time and efforts in reviewing our manuscript and providing constructive comments. These comments are all insightful and helpful for improving our study. We have carefully considered the comments and updated the manuscript to address them. Below we provide point-by-point responses in blue. The line numbers refer to the tracked version of our manuscript.

Response to Reviewer#1:

General comments

The preprint authors address an important scientific question of theoretical and practical relevance: modelling the extreme part of precipitation with CMIP-style climate models. The introductory part follows a clear and logic structure by introducing the topic and its relevance, reviewing other model analysis studies in the field, recounting the origins of observational and reanalysis data used for comparison, introducing the details of the model simulations, and finally explaining the statistical terms and quantities. In particular the assessment of the strengths and weaknesses of the observational and reanalysis data sets allows the reader to understand the following analysis.

The article is well written in a clear language with distinct formulations, which allows to easily understand and follow the text.

Specific comments

The introduction is very helpful in introducing the relevant precipitation processes, related biases in CMIP models and (potential) links to spatial and temporal resolution. The detailed description of methods to compute quantiles highlights details that many readers may not be aware of. This is helps to understand and follow the analysis procedure in greater detail.

The analysis of results highlights improvement patterns (e.g. improvements over mountainous regions with higher horizontal resolution, better convection precipitation with shorter time steps), while at the same time not generalising where there is no evidence and also indicating when and why there might be no improvements (e.g. MR to HR). Overall, the conclusions appear convincing and detailed.

Technical comments

Two minor comments: Figures 2, 4 and 5 seem to be at the lower limit w.r.t. their size (when printed on A4 paper). Particularly the fonts in this figures is a bit hard to read.

Thank you for pointing out these issues, we adjusted the size and fonts in these figures (see page 20-24).

The text refers to figures S2, S3, S4, and S5 but the referee could not find these figures. However, it is possible that I am just not aware of some common practise in this journal. It is also noted that the code for those figures is included in the referenced Jupyter notebooks.

Thank you and sorry to hear that you could not find supplementary figures S2, S3, S4 and S5. Actually, this is included in the additional material which you can see and find in the supplementary material at: https://gmd.copernicus.org/preprints/gmd-2024-66/

(You can find the supplementary materials on the top right corner, which has a red sign 'Download')

Response to Reviewer#2:

General comments

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 study has its merits and is covering the very relevant topic of precipitation extremes in Europe, the approaches followed in evaluation and analysis of the results require substantial changes and reconsideration.

Specific comments

1. The general introduction of the topic this study addresses and the context and background information provided by previous studies is rather confusing and lacks focus. The authors cover a very wide range of studies, from CMIP-based studies (300-50km, multi-year runs) to even regional climate and seasonal studies at km-scale resolution. It can be improved if the authors focus their introduction in the temporal and spatial scales relevant for the current study, and discuss current limitation of the existing approaches and where their research fits into this context. Most of the information needed is already there and can be re-written in a more concise way.

   Thank you for this helpful comment. We have rewritten the introduction (line 41-91).

2. The novelty of the approach followed by the authors should be further highlighted. For example what is the difference of this study versus the one done by Strandberg & Lind (2021), which seems to cover exactly what the topic the authors addresses in this study, at even higher horizontal resolutions and with multiple models.

   Thank you for this important suggestion. There are differences between this study and the one done by Strandberg & Lind (2021). Firstly, we analyzed the effect of both horizontal resolution and model time step on extreme precipitation using atmospheric model only, while Strandberg & Lind (2021) only focused on the horizontal resolution's impact on precipitation in coupled model simulations. Secondly, we also compared the large-scale and convective precipitation in different resolution and time step configurations, which was not studied in Strandberg & Lind (2021). Now we have added more information to make it more clear in line 82-89.

3. The experimental setup could be further improved. What is the motivation behind using the same timestep for MR and HR experiment. I would expect the timestep for the 50km run to be 20 min. This would certainly have an impact on precipitation extreme with the MR experiments. Why not test the impact of model timestep across all model resolutions (i.e., including MR and HR experiments). This would make for a

rather interesting and novel study, where the authors could identify if timestep sensitivity for climate simulation of extreme precipitations changes across resolutions.

Thank you and this is really a good point, I agree that it would be very interesting to identify if timestep sensitivity for simulation of extreme precipitation changes across resolutions. However, we usually do not use long time step in HR in OpenIFS, because it is not reasonable for some physical processes (e.g. advection) and we will get errors from semi-Lagrangian scheme. The time step for HR is usually 15 minutes or shorter. Different time steps are only tested for LR because the motivation of experimental setup is to investigate which time step at low resolution is suitable for coupled model simulations (Savita et al., 2024). Computational cost is also a limit that we can not test all time steps for high resolutions.

The reason we keep the same time step for LR, MR and HR experiment is that we wanted to investigate how the model precipitation will change due to changes in model horizontal resolution.

4. Regarding the model evaluation, although the use of the GPCC dataset is appropriate the use of ERA5 as reference benchmarks for evaluating RMSE of convective precipitation is not appropriate. My understanding is that ERA5 does not explicitly assimilate convective and large-scale precipitation rates, only total precipitation estimates. Also the authors show in Figures 1-3, that ERA5 is not substantially better (and not significantly different) at capturing precipitation at any percentiles compared to MR and HR simulations as RMSE for ERA5 lies often inside the CIs of MR and HR. Hence this does not justify using it as an evaluation benchmark for convective precipitation to compute RMSE scores. Although the discussion about difference in convective vs large-scale precipitation is interesting and could be including as a indication of differences (preferably showing them as PDFs all the different precipitation types) for each experiment, the current evaluation done in section 3.2 is not appropriate.

   Thank you for this valuable comment. ERA5 does not assimilate total precipitation, which is calculated from the combined large-scale and convective precipitation. Observation datasets such as GPCC and GPCP do not provide convective and large-scale precipitation separately. ERA5 is not the truth, the reason we took ERA5 as benchmarks is that ERA5 has smaller biases in simulating precipitation in extra-tropics (Lavers et al., 2022), we would expect the convective and large-scale precipitation also have smaller biases. In section 3.2, we do not focus much on how the model simulations are close to ERA5, we more care about whether the differences of each simulation and ERA5 (represented by RMSE) are significantly different to each other. That means we more focus on the error bar (Fig. 6) overlaps with others, but not their values. We also updated some texts (line 491-498, 518).

   As you suggested, we added a frequency distribution plot of cp and lsp in the supplementary Fig. S6 & 7. We also added an additional paragraph to discuss the

physical processes involved (line 444-459). This also clarify the comment 5 you mentioned.

5. During the analysis of the results in section 3.2 the authors often provide remarks about convective and large-scale precipitation differences between the different resolutions for summer and winter, without discussing the physical processes involved. For example it would help to clarify that convective precipitation is generally related with not explicitly resolved convective motions, and deep convection systems with scales smaller than the effective resolution of the model (e.g., Mediterranean Hurricanes or MCS) which tend to contribute more precipitation around the Mediterranean. On the other hand, large-scale precipitation is likely to originate from large-scale synoptic storms at these resolutions. As the effective resolution of the model increases the ratio between convective and large-scale precipitation will tend to change, since more convective motions are resolved rather than parametrized. Again plotting a frequency distribution for precipitation types for the different experiments will really help here.

   Thank you very much and please see the response to comment 4.

6. Line 53: What to the authors mean by "lack of observations"? Does this mean that there is no assimilation of precipitation observations in climate models, or that we lack observations of precipitation to built better parametrization schemes?

   Thank you for pointing out this issue. We mean that we don't have long-term observations for precipitation which are very important for evaluating the climate models. We rewrote the sentences to make it clear (line 46-50).

7. Line 133-136. What time of interpolation is used here to convert the data to the regular grid and also from the native resolution for MR and HR experiments to the intermediate resolution of 0.45 degrees. Is it similar to the one mentioned in lines 170-174?

   Thank you for your comment. Bilinear method is used by XIOS to convert the data from native grid data to regular grid, also from native resolution for MR and HR to the intermediate resolution of 0.45 degrees. It is not the same as the one mentioned in lines 170-174, which is the second-conservative method. We redid one-year (1979) simulation using the HR model and saved the output on a native grid. We converted the native grid data to regular grid using bilinear and conservative method separately. We found that these two distributions are similar, the bilinear one has slightly more extreme events than the conservative one. The 99th percentile precipitation for the two distributions are very close (17.8 mm/d for bilinear and 17.4 mm/d for conservative method), therefore, we think the interpolation method would not have a large effect on the final conclusion.

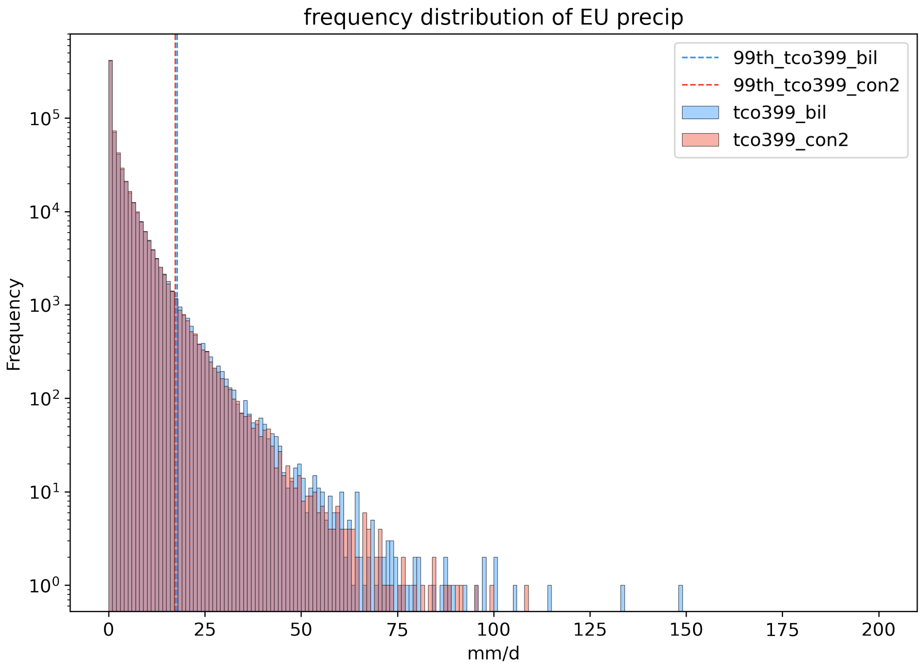

Figure. Frequency distribution of European precipitation converted from native grid to regular grid by bilinear (blue) and conservative (red) method in 1979. The dash lines are the 99th percentile value for each distribution.

8. Line 260-261: How is this sentence related to the overestimation of precipitation over the Alps in the LR experiments? Lavers et al. (2022) is about a single storm in ERA5, rather than the consistent multi-year estimations of 99th precipitation percentiles. Also ERA5 resolution is 31km, which is very different than the LR experiment and closer to the HR experiment, and based on Figure 2 I can't find a substantial overestimation of precipitation in the northern side of the Alps for ERA5 or HR.

   Thank you very much for your suggestion. We agree that both the time scales and resolutions are different between the single storm in ERA5 and consistent multi-year estimations of 99th precipitation percentiles in LR. It is not appropriate to compare them, therefore, we deleted that sentence.

9. Line 271: Have authors checked the data from GPCP (instead of GPCC) to see if the high values near Slovenia are also present in that dataset. This may help with diagnosing the source of the bias in precipitation over that region.

   Thank you for this idea. We checked the data from both GPCP and EOBS, and found the similar negative bias near Slovenia in EOBS, but not in the GPCP. We have added this information in the manuscript (line 391-399) and a plot in supplementary Fig. S8.

10. Line 433: I think Jung et al. (2012) does not discuss at all changes in tropical cyclone intensity, but only extratropical cyclone intensity.

    Thank you for pointing out this issue and it is fixed now (line 589-594).