REPLY TO THE COMMENTS OF THE REFEREE #1

Dear reviewer,

First of all, we would like to thank you for the time you have spent reviewing our manuscript. We strongly appreciate the constructive comments and valuable feedback made. We have carefully addressed the reviewer's comments and suggestions. Below are our point-by-point responses to the comments in blue.

Thank you very much again for your review.

Author and Co-Authors

COMMENTS FROM REVIEWER#1

The paper presents an evaluation of extreme precipitation from two regional climate models; one is a regional model at 12km resolution, the other is a convection-permitting model at 3km resolution. The evaluation is based on different observational products: two grid products at 1day and 1hour temporal resolution, and a network of rain gauges (about 190 at daily resolution and 10 at hourly resolution). Seasonality and relation with elevation are also explored, at regional (by contrast with the gridded product) and local scale (by contrast with rain gages). The main findings indicate generally better performance for the HCLIM3 than the HCLIM12 model, more clearly at hourly resolution.

The study is of interest on the general topic of evaluation of extreme precipitation from convection-permitting models, which may help in better understanding how to use them in practical applications. In my opinion, it gives an incremental advancement in this field more than novelty, considering the regional scale (Norway) and the use of metrics commonly used in these kind of studies (Rx1d, Rx1h, return levels). It is well written, but I found difficult to get the main messages on the results because of the total length, number of figures and panels. The topic is of interest, and fitting the journal scopes, but I suggest a few major revisions and some minor before publication in HESS. My comments are listed below.

- Major 1. You used 8 regions. I wonder if less can be used, pooling together smaller ones. I say this based on two considerations: 1) extension is very different across the regions, and regional evaluation of models are then based on quite different number of grid points (for example, how many grid points for the two small regions in the south?) and number of daily rain gages (just 4-11-14 for three regions!); 2) it is difficult to follow the explanations and figures with comparisons on 8 regions, 2 models, 4 seasons, 2 durations … and get a message on the results; maybe having less could help.

Reply: Thanks for your valuable comment. Based on your suggestions, we conducted additional analyses by merging the two smaller regions in the south (southern and south-western) and the two regions with fewer rain gauges (northern-coastal and northern-

inland). We have updated Figure1, 2 and 4 accordingly (please see the revised figures below). Our finding from these new analyses indicated that:

(1) HCLIM3 performed worse than HCLIM12 in seasonal and annual daily maximums (Rx1d) in the newly merged regions. This contrasts with our original findings using eight regions, where the deficiencies of HCLIM3 were primarily observed in the south-western region alone at regional scale. The discrepancies are noticeable in the annual and seasonal Rx1d biases (Figures R1 and R2) and in the biases of extreme annual Rx1d at various return levels (Figure R3). For example, the positive bias in the south-western region is compensated by the negative bias in the southern region in the merge region of S-SW (southern and south-western) during winter, leading to a different interpretation of HCLIM3's added value compared to HCLIM12. Overall, HCLIM3 does not show added value in the south-western region seasonally and merging it with southern region masks the potential benefits of HCLIM3 observed in the southern region alone.

(2) Hanssen-Bauer et al. (2006) originally divided Norway into 13 regions, which were further reduced to 8 regions based on shared characteristics, as suggested by Konstali and Sorteberg (2022) and Michel et al. (2021). Given Norway's complex coastal climate and geological conditions, 8 regions represent the minimum number necessary to capture the climatic variability adequately.

(3) The overall goal of the supported Impetus4Change (I4C) project is to improve the usability of climate information and services at the local and regional levels. To provide

climate information for each administrative region in Norway directly, we divide Norway into 8 regions.

Therefore, we decided to retain the original eight regions in our study, as recommended by Konstali and Sorteberg (2022) and Michel et al. (2021).

We acknowledge that the current presentation of results may be overwhelming and require better synthesis to clarify conclusions and convey a clear message. To address this, we have improved the presentation of our findings in the revised manuscript. For example, we identified redundant in Figure 5, 9, and 12 compared to Figure 13, 14, and 15, which already include precipitation frequency data. Consequently, we have removed Figure 5, 9, and 12 to reduce redundancy and enhance clarity. This streamlining aims to make our conclusions more concise and easier to follow.

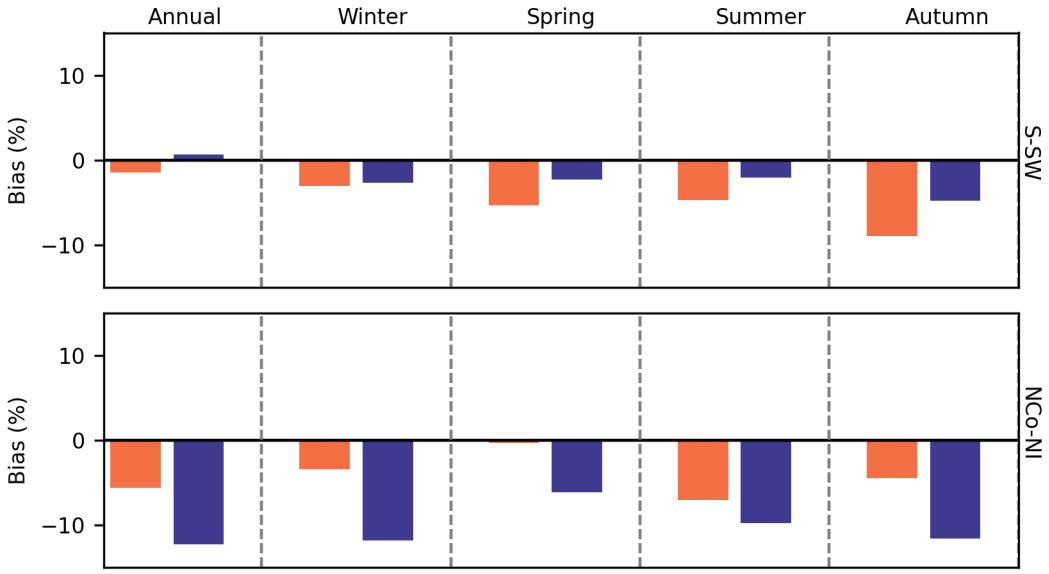

Figure R1: The percentage bias of seasonal Rx1d from HCLIM3 and HCLIM12 to SeNorge for merged southern and south-western (S-SW), northern-coastal and northern-inland (NCo-NI) regions.

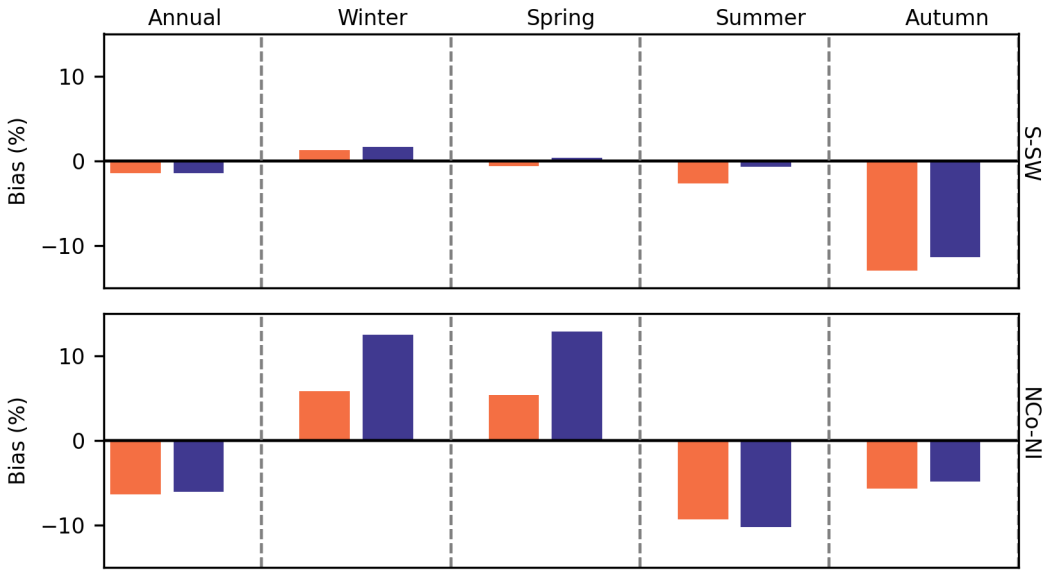

Figure R2: The percentage bias of seasonal Rx1d from HCLIM3 and HCLIM12 to daily in-situ observation for merged southern and south-western (S-SW), northern-coastal and northern-inland (NCo-NI) regions.

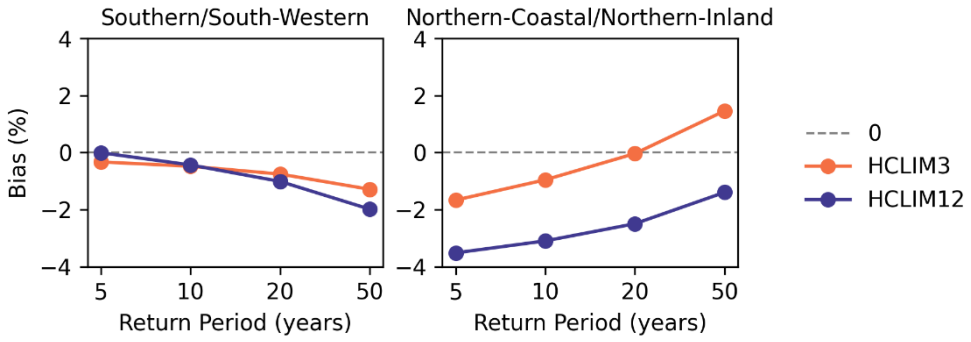

Figure R3: The bias of extreme annual Rx1d exceeding the 5-year to 50-year over eight regions between seNorge and HCLIMs (i.e., HCLIM3 and HCLIM12).

- Major 2. Regional scale is here referred to the analysis using the gridded products; local scale is referred to the analysis based on rain gages. They show different results, but I wonder how much this is due to the use of different observation products. How is SeNorge-models comparison sampled on the same rain gage points? Or, how is seNorge compared to rain gages? Differences you highlight in your text (e.g lines 337, 352, 384, etc) could be due to the use of a different observational benchmark. Moreover, when comparing gridded products and

climate model, you compare values from a same size grid (12x12km), when comparing model and rain gages, the comparison is made on a 12x12km grid and a point measurement. I suggest to add further analysis comparing seNorge with rain gages, or extracting the comparison between seNorge vs models on the same locations of rain gages, or to add some considerations in the discussion.

Reply: We fully agree with your comment. We have analyzed and compared seNorge with rain gages by interpolating the seNorge to the same location of rain gauges. We have added it in the discussion in the revised manuscript.

- Major 3. Biases (e.g. figure 2, 6) are shown in mm (absolute differences). Maybe relative differences (modelled-observed)/observed could be more meaningful considering that precipitation has a big range across the regions: a bias of 5 mm is different on a 30 mm or a 80 mm daily precipitation! My suggestion is to update maps and plots of biases with relative bias (%), and to revise description of results and comments on “magnitude” of bias based on this.

Reply: Thanks for your comment. We have updated all the plots with relative bias (%).

Minor comments

1. title: I suggest hourly in place of sub-daily, considering that you just evaluate 1h

   Reply: Thanks for the comment, the titles were changed for: Enhanced Evaluation of hourly and Daily Extreme Precipitation in Norway from Convection-Permitting Models at Regional and Local Scales.

2. Line 120: these results on orographic effect “were based on the annual maxima”… yes. But also yours are based on Rx1d (see line 222-223). So … why do you highlight these about other studies at line 120, if then you do the same? I suggest to remove.

   Reply: Thanks for your comment. We have removed it in the revised manuscript.

3. Line 121-122. “The dependence on seasonality … need the evaluation based on season”. Of course! Maybe you wanted to tell something different here and I didn’t get it. Please calrify/modify.

   Reply: Thanks for your comment. We appreciate your feedback. What we intended to convey is that the performance of CPRCMs is heavily influenced by seasonal variations, which necessitates evaluating the orographic effects of seasonal extremes in addition to annual extremes. This is crucial for understanding how different seasons impact model performance and the resulting hydrological responses.

Line 195. Not clear: you say here you averaged the indices in the region … then in caption of figure 2 you write the bias is calculated at each grid point (this makes sense to me). So, when do you use the averaged indices?

Reply: Thanks for your comment. To clarify, the bias is initially calculated at each grid point within the region. After calculating the bias for each grid point, we then average these biases across all grid points within each region to obtain a regional average bias. This approach ensures that the regional bias is representative of all grid points in the region.

4. Line 206. Specify somewhere in the section that this is done for daily data (both gridded and rain gages), while just on 10 rain gages at 1h duration. More importantly, later in the text (lines 367, 372,..) you mention uncertainty … and it is never explained before in the paper how it is evaluated. Add it in the methodology.

   Reply: Thanks for your comment. The uncertainty means the long whisky. We have rewritten it more clearly in the revised manuscript.

5. Figure 2 (same for figures 6 and 7). The caption mentions absolute bias but describes a relative bias calculation. Please correct.

   Reply: Thanks for your comments. We have corrected it to percentage bias (simulations minus observations, divided by observations) in the revised manuscript.

6. Figure 2c (same for figures 6 and 7). I suggest to add another row with the regional bias for annual Rx1d, not just seasonal, in order to have a synthesis of what is shown in the maps in panel a.

   Reply: We fully agree with your comment. We have added the row with regional bias for annual Rx1d in the revised manuscript.

7. Figure 4. Same y-axis limits could help in comparing the bias… and maybe you can revise the description of results better considering the different magnitude of the bias (example at lines 290-291).

   Reply: We agree. We have updated the plot with same y-axis limits.

8. Figure 4. I can't understand why Northern-Coastal has so big bias for HCLIM3, considering that figure 3 shows a tendency of underestimation of the empirical distribution, similar to Northern-Inland, for which you find big underestimation of return levels. And for the Norther-Inland, why underestimation of return levels is bigger for HCLIM3, while the distribution in figure 3 show more underestimation for HCLIM12? … please check the correctness of the results.

   Reply: Thanks for your comment and for pointing out the discrepancies. Upon review, we recalculated the return levels for all regions and found that the calculations for the Northern-Coastal and Northern-Inland regions were indeed incorrect. As a result, we have corrected these errors and updated the results and corresponding plot. We apologize for the confusion. Thank you again for bringing this to our attention.

9. Line 370. For 50yr return-periods they seem identical, not larger bias for HCLIM3. I would remove it

   Reply: Thanks for your comment. The 50yr return-periods has now removed.

10. Line 439-440. This is based on just 10 points. This can't be considered a general finding, I suggest to mention the limit of the analysis.

    Reply: We fully agree with your comment. We have added the analysis of the limitation in the corresponding results and discussions.

11. Line 483, section 4.5. You show the slope of precipitation with elevation as absolute value, mm/km. I strongly suggest to calculate and show it as relative slope, for example with respect to the average value of Rx. Because 1mm/km has a different magnitude for Rx1d and Rx1h. Then I suggest to revise your discussion considering this … (e.g. I see very weak relation of Rx1d with elevation, so I'm sure you can really speak about reverse orographic effect …also at line 629)

    Reply: We fully agree. We have recalculated the relative slope and updated it in the revised manuscript.

12. Line 494. "Significant"? Based on a specific test? Maybe "relevant"…

    Reply: Thanks. We have replaced the significant with relevant.

13. Line 561-562. Not very informative consideration …. Could you elaborate more on this? Or delete …

Reply: Thanks for your comment. We have deleted it.

14. Line 580-582. I see here two contrasting points. 1) You mention *underestimation* for return levels, but for Rx1d in figure 8 I see bias around zero, while for Rx1h you have evaluation on just 10points. 2) Then you say this is in line with results in Malawi (!!!!) finding *overestimation*. I can't really understand your reasoning here.

Reply: Thanks for your comment. We apologize for any confusion caused by the presentation of our results. To clarify the two points you raised:

(1) We recognize that our initial explanation might have been unclear. For the analysis, we retained only those stations with less than 10% missing data from 1999 to 2018, resulting in a total of 10 hourly stations across Norway. Although the number of rain-gauge is small, the data quality is high and the time series is extensive. We understand that using only 10 stations introduces some uncertainty, and we acknowledge this in our revised manuscript. We have expanded our discussion to address the limitations and potential impacts of this limited dataset on our findings.

Additionally, the SeNorge2 gridded hourly data covers only an 8-year period, making it difficult to compare directly with the station-based results. However, the evaluation results derived from the 20-year station data are consistent with those from the 8-year gridded data, both indicating a significant underestimation by HCLIM12. This consistency between the short-term gridded data over the region and the long-term

station data highlights the robustness of our findings regarding HCLIM12's performance.

(2) We realize that referencing Malawi, a region with different climatic conditions, was inappropriate for drawing direct comparisons. In the revised manuscript, we have instead cited results from Thomassen et al. (2023), who studied similar return levels using the same HCLIM3 and HCLIM12 models in Denmark, which is geographically closer to Norway and more relevant in terms of climate. This citation provides a more appropriate comparison, highlighting that our findings are consistent with those observed in a comparable northern European context.

"Thomassen et al. (2023) compared the performance of HCLIM3 and HCLIM12 based on local rain-gauge data in Denmark, and found that HCLIM12 indeed underestimate the hourly extreme event and HCLIM3 agree well with observation. Our results at local scale corroborates these underestimation results limited at some local place (SN58900, SN44300, SN76539) regarding to the hourly extremes (Fig. 11), and the bias from HCLIM3 is closer to 0 in other 7 local places. Furthermore, the HCLIM3 in capturing the daily extremes (Fig. 8) at local scale tend to almost 0 bias. In addition, the return levels from HCLIM12 underestimate hourly extreme precipitation in all local places (Fig. 11) and daily extremes in southern, middle-inland, middle-coastal and northern-inland regions (Fig. 8). It should be noted that the limitation of Rx1h at 10 points, however, the obvious underestimation of RCMs in simulating the return level of Rx1h in Norway have also been indicated in Médus et al. (2022)."

15. Line 587. I can't understand the meaning of "weakening the superior" …

    Reply: Thanks for your comments. We have rewritten this sentence:

    The hourly and daily precipitation extremes from HCLIM3 at most stations show more realistic results than at the regional scale especially for hourly extremes at 10 points, supporting the hypothesis of damped extremes at regional scale uncovering the superiority from CPRCMs at some regions (eastern and middle-coastal). However, the added value from CPRCMs may also disappear for some regions (western and middle-coastal regions) at local scale compared to regional scale.

16. Line 635. Dallan et al. 2023 analyzed annual Rx1d: this can't be related with the seasonal Rx1d. I suggest to rephrase in some way: "An unclear relation of Rx1d with elevation at regional scale was also seen from the study of Dallan et al. (2023), in which, they analyzed annual Rx1d based on CPRCMs and in-situ observation over Alpine"

    Reply: Thanks for your comments. We have rewritten it according to your suggestions.

17. Line 640. I suggest to add a few recent references on orographic enhancement at daily scale observed in different regions (e.g. Formetta et al. 2021 https://doi.org/10.1016/j.advwatres.2021.104085 and Amponsah et al. 2022 https://doi.org/10.1016/j.jhydrol.2022.128090); same at line 651 for the reverse orographic effect, adding also Formetta et al 2021, considering they explored durations from subhourly to daily.

Reply: Thanks for your suggestions. We have added your suggested references in the revised manuscript.

18. Please also revise your conclusions accordingly to the modifications you will do in the revised version of the manuscript

Reply: Thank you for your feedback. We will carefully revise our conclusions to reflect the changes and updates made to the manuscript.

Thanks very much for your input, which helps us improve the quality and clarity of our manuscript.

References

Hanssen-Bauer, I., Tveito, O. E., & Szewczyk-Bartnicka, H. (2006). Comparison of grid-based and station-based regional temperature and precipitation series (Tech. Rep. No. 04). Norwegian Meteorological Institute.

Konstali, K. and Sorteberg, A.: Why has Precipitation Increased in the Last 120 Years in Norway?, Journal of Geophysical Research: Atmospheres, 127, 10.1029/2021jd036234, 2022.

Médus, E., Thomassen, E. D., Belušić, D., Lind, P., Berg, P., Christensen, J. H., Christensen, O. B., Dobler, A., Kjellström, E., Olsson, J., and Yang, W.: Characteristics of precipitation extremes over the Nordic region: added value of convection-permitting modeling, Nat. Hazards Earth Syst. Sci., 22, 693-711, 10.5194/nhess-22-693-2022, 2022.

Michel, C., Sorteberg, A., Eckhardt, S., Weijenborg, C., Stohl, A., and Cassiani, M.: Characterization of the atmospheric environment during extreme precipitation events associated with atmospheric rivers in Norway - Seasonal and regional aspects, Weather and Climate Extremes, 34, 100370, 10.1016/j.wace.2021.100370, 2021.

Thomassen, E. D., Arnbjerg-Nielsen, K., Sørup, H. J. D., Langen, P. L., Olsson, J., Pedersen, R. A., and Christensen, O. B.: Spatial and temporal characteristics of extreme rainfall: Added benefits with sub-kilometre-resolution climate model simulations?, Quarterly Journal of the Royal Meteorological Society, 149, 1913-1931, https://doi.org/10.1002/qj.4488, 2023.