We organize this Response to Reviewers document in the following way. Reviewers' comments are numbered as RxCy, where X is the reviewer number, and Y is the reviewer's comment number. Our responses are reported in blue, and new or modified text highlighted in *italics red*.

Reviewer 1 Comments and Responses

General Comments

R1C1: The given paper studies how bias of simulated extreme precipitation, from both convection permitting models (CPM) and their driving regional climate models, changes with spatial and temporal aggregation over complex terrain. It thereby constitutes an important addition to the current scope of literature, where CPM evaluation is commonly performed at the native resolution only. While added value at high resolution has been proven throughout a variety of studies, it remained unexplored how this improvement in model performance reflects at areal aggregation of up to several thousands of kilometers and at temporal aggregation of up to 24 h. Such scales are however highly relevant for hydrological modelling in large river basins and of long-duration flood drivers. Given the high computational demand of CPMs, knowledge on how and whether added value of CPMs reflects at higher areal and temporal scales is of high practical relevance. A better understanding of the performance of CPMs, such as potential changes in the sign and magnitude of model bias with aggregation, furthermore pinpoints to physical inconsistencies in the current climate models and thereby supports model development. In conclusion, the presented paper bridges an important research gap, is of direct practical relevance and advances our understanding of CPM performance. It is therefore greatly welcomed as a contribution to HESS. The paper is well structured, presents the findings in a clear and concise manner and is of high scientific quality. The current version of the manuscript would benefit from a sharper contextualisation and a more in-depth discussion on the scales of hydrological processes and their importance in the generation of different flood types through a revision of its introduction, as outlined below. The methodology section leaves a few open questions, which are addressed in the following. The results and their discussion are of high quality, with some rather technical revisions needed before publication.
Response: We thank you sincerely for the thoughtful evaluation of our work and your constructive suggestions, whose implementation has strengthened the manuscript. We have addressed each of your comments and provide our responses below.

Specific Comments

R1C2: P. 1/ L. 17: It would be relevant to mention over which study area you're conducting the analyses, namely over Switzerland, as it gives the reader an idea of the physiography at hand.
Response: We agree and included a note on the study domain, that is Switzerland, in the opening of the abstract. *"Here, we assess how well CPM simulations represent areal precipitation extremes over Switzerland across durations from 1 to 24 h and spatial scales from ~10 to 5000 km^2."*

R1C3: P. 1/ L. 20: The information of the length of the used time series could be useful, as it is a proxy for the robustness of the results (20 years).
Response: We agree and have mentioned the record length of the observational reference in the abstract. *"We use 20 years (2005–2024) of hourly precipitation from Switzerland's high-resolution radar–gauge product as a reference…"*

R1C4: P. 1/ L. 23: With only the abstract, it is not clear what the exact coefficient is that is under-/ overestimated (it is the 20-y extreme precipitation return level).
Response: Thanks, we have clarified this in the abstract. *"For 1–3 h extremes, CPM bias in 20-year return levels strongly depends on the spatial scale, shifting from a ~15% underestimation at native resolution to near-zero bias at ~400 km^2 and to ~20% overestimation at ~4000 km^2."*

R1C5: P. 24/ L. 24: The statement is given that: „RCMs consistently underestimate precipitation extremes across all spatial scales“. The RCMs' bias is not quantified in the abstract, yet at the end (L. 27) it is concluded that CPMs outperform the RCMs for short-duration extremes. It would be helpful to underline this statement with a number for the RCMs' bias. This allows the reader to follow the conclusion and gain an estimate of the degree of added value by CPMs.
Response: Thanks for pointing this out, now we have explicitly mentioned the RCMs' bias with corresponding numbers. *"RCMs consistently underestimate 20-year return levels across all spatial scales, with biases ranging from ~40% underestimation at native resolution (~144 km^2) to ~10% underestimation at the largest aggregation scales (~5000 km^2)."*

R1C6: P. 1/ L. 28: This statement could be refined by specifying for which catchments and flood types in particular scale-dependent biases are of major importance.
Response: We have refined the statement as: *"Overall, CPMs offer important added value for representing short-duration extremes relevant to flash floods and debris flows in small to medium-size Alpine catchments (10–10^3 km^2), but scale-dependent biases must be accounted for when translating CPM outputs into flood-risk assessments at these scales."*

R1C7: P. 2 / L. 32: One important parameter for flood generation, modified by spatial averaging, is indeed extreme precipitation intensity. However, it is not the only one, with precipitation volume also being a defining factor. This aspect could be mentioned here.
Response: We have added the aspect regarding the effect of precipitation volume on flood generation to the end of the opening paragraph in the Introduction. The new sentence reads as: *"Across these scales, both the intensity of extreme precipitation and the total precipitation volume accumulated over the catchment shape the flood response, with their relative importance depending on catchment size, storm duration, and antecedent conditions* (Viglione and Blöschl, 2009)*."*

R1C8: P. 2/ L. 34 f.: The given statement does not only apply to complex terrain, but is valid universally. However, there is indeed great value in conducting the analyses over mountainous terrain, as you highlight in L. 82 ff. In the introduction, you could elaborate more deeply on what's special in complex terrain and why your works finds particular relevance here.

Response: We have modified the opening paragraph in the Introduction to elaborate more on the relevance of complex topography in this context. The paragraph now reads as: *"A defining characteristic of extreme precipitation is its inherent space and time scale dependence. Extremes measured at a point location differ systematically from extremes averaged over a catchment area, and as the spatial averaging scale increases, extreme precipitation intensities generally decrease. The rate of this reduction depends on storm type, storm organization, and terrain complexity* (Breinl et al., 2020; De Michele et al., 2001; Gericke and Pietersen, 2020; Svensson and Jones, 2010)*. While these scaling relationships apply universally, they are particularly pronounced in regions of complex terrain, where orographic effect, and sharp climatic gradients across mountain ranges produce strong spatial heterogeneity in storm structure. In such regions, the spatial scales relevant for flood hydrology span several orders of magnitude, from a few square kilometers in small Alpine or urban catchments to several thousand square kilometers in large river basins. Across these scales, both the intensity of extreme precipitation and the total precipitation volume accumulated over the catchment shape the flood response, with their relative importance depending on catchment size, storm duration, and antecedent conditions* (Viglione and Blöschl, 2009)*."*

R1C9: P: 2/ L. 52 f. & L. 58: A more in-depth discussion on the scales of hydrological processes and their importance in the generation of different flood types, in conjunction with the meteorological scales would be needed. CPMs are known to offer added value for the representation of small-scale heavy summer storms, with Hortonian overland flow being a major driver of resulting flash floods in fast responding catchments. Over longer time scales and larger catchments with longer response times, long-duration flood drivers take the lead, with saturation-excess overland flow being a major player. A discussion on how flood drivers change with scale and where CPMs are expected to offer added value for hydrological impact modelling, would be very valuable. Conclusionary, the types of catchments and floods for which the work presented here has primary relevance can be refined (c.f. comment to P. 1/ L. 28).
Response: Thanks for this suggestion. We have expanded the introduction to clarify how flood-generating mechanisms differ across catchment scales and durations, and where CPMs are expected to add value for hydrological impact modelling. This led to the modification of the entire fourth paragraph of the Introduction, which now reads as follows:

"Despite the growing evidence that CPMs improve sub-daily precipitation statistics (Estermann et al., 2025; Fosser et al., 2024; Lucas-Picher et al., 2024)*, most evaluation studies still emphasize grid-point metrics and station-based comparisons, while hydrological applications typically require areal precipitation over scales spanning up to several thousands of square kilometers* (Rasmussen et al., 2012)*. The relevance of potential improvements obtained by using CPMs instead of coarser resolution climate data for flood hazard applications depends strongly on the flood type and catchment scale considered. In small, fast-responding Alpine catchments, flash floods and debris flows are typically driven by short-duration, high-intensity convective precipitation, with characteristic response times of a few hours* (Borga et al., 2014; Marchi et al., 2010)*. For these flood types, the spatial organization and intensity of hourly precipitation extremes — which CPMs explicitly resolve — directly control flood magnitude and*

timing. As catchment size increases, response times lengthen and flood-generating precipitation durations shift toward sub-daily to multi-day accumulations, with saturation-excess processes and antecedent soil moisture playing a more important role (Viglione and Blöschl, 2009). For these flood types, large-scale synoptic forcing dominates the precipitation mechanism, and the added value of explicitly resolving convection may be less critical. However, it is still not well known how CPM skill in reproducing precipitation extremes changes with areal aggregation, nor whether CPMs and their driving regional climate models (RCMs) behave differently under the same spatial averaging. Clarifying these scale-dependent behaviours is necessary if CPMs are to be used confidently for analysing climate change impacts on flood hazard. At the same time, CPM simulations remain computationally demanding (Ban et al., 2021; Schär et al., 2020)*, and it is important to understand at which spatial and temporal scales they provide clear practical benefits over RCMs* (Kendon et al., 2021; Lucas-Picher et al., 2024)*."*

R1C10: P. 2/ L. 60: Without having read the study, it is ambiguous whether the „spatial and temporal scales" mentioned refer to scales of modelling or aggregation. I would hence suggest rewriting this sentence.
Response: We have rephrased the sentence to make it explicit that we refer to aggregation scales. *"This study sheds light on the dependence of climate model biases on spatial and temporal aggregation scales by evaluating how well a multi-model CPM ensemble from the CORDEX-FPS framework represents precipitation extremes…"*

R1C11: P. 3/ L. 66: A specification which types of biases are quantified would make it clearer for the reader. Furthermore, the term of "event duration" might be misleading, since it refers to the temporal aggregation window, not however to the duration of the studied rainfall events.
Response: We have clarified these terms in the study objectives by replacing "event duration" with "precipitation accumulation duration" and specifying that biases refer to extreme precipitation return levels. *"In particular, we (i) assess how spatial patterns and the magnitudes of extreme precipitation return levels change from the native grid spacing to aggregated areas, (ii) quantify how CPM and RCM biases in extreme precipitation return levels depend on precipitation accumulation duration and spatial aggregation, and (iii) assess why these biases change with spatial scale."*

R1C12: P. 5/ L. 106: Is there any quantification of the uncertainties of CombiPrecip available? Such work has e.g. been done for the radar-based, gauge-adjusted quantitative precipitation estimates over neighbouring Germany (RADOLAN; Kreklow et al., 2020). If the uncertainties of CombiPrecip are known, this would be valuable information to be added.
Response: Thanks for your suggestion on including information regarding the uncertainty of CombiPrecip in Section 2.2.1. In fact, uncertainty in CombiPrecip has been quantified in several studies, which we now reference explicitly in Section 2.2.1. In response to this comment we have added the following short description*:*

"It is important to note that MeteoSwiss also provides a radar-only precipitation product. For the purpose of our study, CombiPrecip offers a pragmatic compromise between spatial detail and bias control, combining the spatial information of the radar mosaic with gauge-based

adjustment at the ground. Cross-validation against rain gauges shows that this merging reduces the quantitative error of the radar-only product by approximately 40% at hourly aggregation (Barton et al., 2020)*. Residual errors in CombiPrecip nonetheless vary systematically with altitude, with bias becoming increasingly negative at higher elevations* (Ghaemi et al., 2023)*, and agreement with gauge observations for extreme precipitation decreases with increasing event severity* (Panziera et al., 2018)*."*

R1C13: P. 6/ L. 152: Here you could add a note that 20 years is too short for a robust climatological analysis.
Response: We have added a sentence on the relatively short length of the time series for a robust climatological analysis:

"We note that the observational reference period available from CombiPrecip (2005–2024) does not overlap with the CPM historical decade (1996–2005). This mismatch is unavoidable given the start date of the radar–gauge record. We also acknowledge that a 20-year observational record and a 10-year simulation are relatively short for robust climatological analysis of extremes, particularly when considering long return periods. This, however, motivates our use of the non-asymptotic SMEV approach (Section 2.3.3)."

R1C14: P. 8/ L. 178: Areal mean precipitation time series are not only constructed for the climate models but also for the observations, right?
Response: Yes, the same procedure is applied to the observations. We have now clarified this in Section 2.3.2.

"For OBS, CPMs, and RCMs alike, we derive areal mean precipitation time series using a sliding, square window centered on each grid cell (see Fig. 1c)."

R1C15: P. 8/ L. 179: This sentence only becomes clear later on, as here it is ambiguous whether e.g. a window size of 4 is a window with 4 pixels or a window with 4 x 4 pixels. I suggest rephrasing it.
Response: We agree that the original phrasing was ambiguous. We have rephrased the window-size definition to make it more explicit.

"For the OBS and CPM models, window sizes include side lengths $g \in \{1, 2, 3, 4, 5, 6, 7, 9, 11, 13, 15, 17, 19, 21\}$ pixels, corresponding to $g \times g$ pixel windows yielding effective areas from $A \approx 9$ up to $A \approx 3969$ km^2. For the RCM models, given their coarser native resolution, the analysis spans side lengths $g \in \{1, 2, 3, 4, 5, 6\}$ pixels ($g \times g$ windows), corresponding to effective areas from $A \approx 144$ up to $A \approx 5184$ km^2."

R1C16: P. 9/ Eq. 3: The given formula suggests that in the computation of the bias, g is the same number for the RCMs and the observations. With the RCMs at 12 km resolution and the observations at 3 km resolution, as an example g = 3 would mean an effective area of 1296 km^2 for the RCMs, but of 81 km^2 for the observations. In keeping, it is not entirely clear, how the metrics are computed e.g. in Fig. 2 f). The question arises whether the RCMs median at 1296 km^2 is compared to the observations at 1089 km^2 or at 81 km^2 (see preceding comment). A reformulation and some more information regarding the computation of the bias would help clarify this.

Response: Thanks for pointing out the need for clarification. The bias was computed by matching the closest effective areas (in km^2) and not by matching g values across datasets (for CPM-to-OBS, they are the same on both g value and corresponding effective area). We have rewritten Eq. 3 and the surrounding text to make this more explicit and clearer. This yielded the following modifications:

A) *"Return levels $I_{i,A}^{(D)}(T)$ associated with a return period of T-years are determined by inverting the SMEV cumulative distribution function in Eq. 1, for each grid cell i, effective area A, and duration D."*

B) $Bias_{i,A}^{(D)}(T) = \frac{I_{Model,i,A}^{(D)}(T)}{I_{OBS,i,A}^{(D)}(T)}$

C) *"where $I_{Model,i,A}^{(D)}(T)$ is the ensemble-median and individual-member CPM/RCM return levels, and $I_{OBS,i,A}^{(D)}(T)$ is the CombiPrecip (OBS) estimate for grid cell i, effective area A and duration D respectively. Biases are always computed at matched effective areas, however, for the RCMs this means comparing each RCM aggregation window against the OBS aggregation window whose effective area is closest to the RCM's."*

R1C17: P. 10/ L. 236 f.: At what spatial resolution is the partitioning into elevation groups being performed?

Response: The elevation grouping is performed at each spatial aggregation scale separately, with the elevation of each grid cell defined as the median elevation within its g×g window. This ensures that the elevation classification is consistent with the spatial scale at which the precipitation field is aggregated. Also, similarly for the catchment-based analysis, each catchment is assigned its median elevation, as already indicated in the caption of Fig. 7 (Fig. 8 in revised version). We have updated the corresponding text in Section 2.3.4 to make this explanation more explicit.

"For the elevation-based analysis, we partition the domain into three elevation groups with approximately an equal number of grid cells: G1 (259–1000 m), G2 (1000–1700 m), and G3 (1700–3612 m). The elevation assigned to each grid cell depends on the spatial aggregation scale considered: for a window size g, each cell is assigned the median elevation within its g×g window, so that the elevation classification is consistent with the spatial scale at which the precipitation field is aggregated."

R1C18: P. 11/ Fig. 2 & P. 12/ Fig. 3: You have chosen to juxtapose the observations and CPMs with an 11 x 11 window (1089 km^2) to the RCMs with a 3 x 3 window (1296 km^2). What favored this choice against an 12 x 12 window for the observations and CPMs (which would have given an equal effective area of 1296 km^2 for all sets)?

Response: We appreciate the question. The choice of aggregation scales for the spatial-pattern figures (Figs. 2–6 in revised version) was motivated by visualisation rather than by equal area constraints between the OBS/CPM and RCM panels. We selected two scales: (i) the smallest scale for each dataset in our analysis (≈9 km^2 for OBS/CPMs, ≈144 km^2 for RCMs), and (ii) an intermediate scale of an order of 10^3 km^2, representative of small-to-medium sized

Alpine catchments. This range has been used in previous studies as well (Marchi et al., 2010; Reszler et al., 2018). The full spatiotemporal scale dependence of biases is examined in Figs. 6 and 7 (7 and 8 in the revised version) across the entire range up to several thousand km^2. Given this target scale and the discrete window sizes we used in our analysis, the closest g×g windows to ~1000 km^2 were 11×11 (≈1089 km^2) for the OBS/CPMs and 3×3 (≈1296 km^2) for the RCMs.

In general, we aimed for odd number window sizes in our analysis for its simplicity and for the symmetry that it provides around the central cell. Also, given the volume of data we didn't want to include all subsequent g values. We did, however, include some even g values for the smallest areas (g = 2, 4, 6) just in order to have smoother curves at these smaller areas.

R1C19: P. 15/ L. 305: I suggest changing the word „strong", as it suggests higher magnitudes of bias, which are not visible from the plot. It only becomes clear that the heavy underestimation is more widespread, but not necessarily more pronounced.
Response: We agree and have rephrased to reflect the spatial extent rather than the magnitude. *"The driving RCMs show a more spatially extensive and uniform underestimation (≤0.6) for 1h-20y extreme precipitation at ≈144 km^2 …"*

R1C20: P. 16/ L. 322: A specification that it is the CPM ensemble median which shows near-zero bias would be recommended, given the large member spread.
Response: Actually, the term "CPMs/RCMs" there and everywhere else in the text refers to the ensemble median of models. We now make this more explicit in Section 2.3 line 162, where we introduce the abbreviations: *"For brevity, we denote the observation product as OBS (CombiPrecip) and the model ensemble median as CPMs/RCMs."*

R1C21: P. 16/ L. 341: The statement is given that mid elevations show mild overestimation of 20-y return levels by CPMs at both 1-h and 24-h durations. While this is true across all scales at 24-h durations, however at 1-h durations, the bias only grows beyond 1 for areas above 1000 km^2. This could be clarified.
Response: We agree and have refined the description. Also, as per R2C7 we have integrated Fig S6 into Fig 7 (Fig. 8 in revised version) that resulted in the revision of the corresponding paragraph. Please refer to the text and figure in R2C7.

R1C22: P. 20/ L. 400: Please present Fig. S8 in the results section as well. In the current version of the manuscript, if is referred to for the first time in the discussion.
Response: Thanks for your suggestion. We have moved Figure S8 (areal reduction factor) into the main text as Fig. 10, in Section 3.4, as also suggested by Reviewer 2. A short paragraph introducing the ARF results has been added to Section 3.4, and the brief description of the ARF in Section 2.3.5 has been slightly expanded. The overall revision in the manuscript reads as follows:

A) *Section 2.3.5: "In addition to this analysis, an areal reduction factor is calculated across the datasets for the estimated return levels, which summarizes the rate at which the intensity of extreme precipitation decreases with increasing area.*

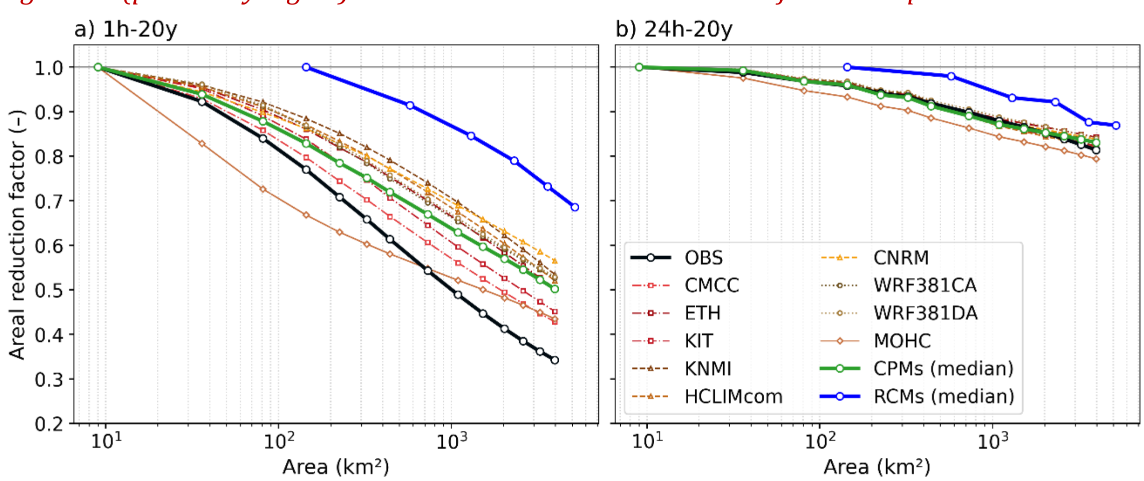
$$ARF_{i,A}^{(D)}(T) = \frac{I_{i,A}^{(D)}(T)}{I_{i,A_{min}}^{(D)}(T)} \qquad (5)$$

Where for each datasets, $I_{i,A}^{(D)}(T)$ is the estimated return level at grid cell i, effective area A and duration D respectively, and $I_{i,Amin}^{(D)}(T)$ is the estimated return level at grid cell i, and duration D for the smallest area A_{min}."

B) *Section 3.4 opening paragraph: "To better explain the spatiotemporal scale dependence of the biases identified in our results, we examine two complementary quantities: (i) the evolution of the within-window coefficient of variation (CV) of annual maximum precipitation with increasing area, which quantifies how spatial heterogeneity changes with aggregation; and (ii) the areal reduction factor (ARF) of extreme return levels, which summarizes how the intensity of estimated return levels decreases as the averaging area increases."*

C) *Section 3.4 new paragraph: "The areal reduction factor for the 20-year return level provides a complementary view (Fig. 10). At 1-h duration (Fig. 10a), OBS shows the most rapid intensity decay with area, reaching ARF values near 0.35 at the largest aggregation scales. The CPM ensemble median decays more slowly than OBS, and individual CPMs span a band above the OBS curve, indicating that simulated 1-h extremes are more spatially coherent than observed ones. The RCM ensemble median exhibits an essentially parallel rate of decay to the CPM ensemble median over the range of common areal scales (above 144 Km^2), but is shifted toward larger areas because of the coarser RCM native grid. At 24-h duration (Fig. 10b), the three datasets decay at very similar rates, with OBS and the CPM ensemble median essentially overlapping; the RCM ensemble median just shifted to the right due to its resolution but follows a similar slope."*

D) *Figure 10 (previously Fig S8) added to the main results section of manuscript:*



"Figure 10: Areal reduction factor (ARF) for 20-year extreme precipitation return levels as a function of areal extent, for (a) 1h and (b) 24h durations. ARF is defined as the ratio of the estimated areal extreme return level at area A to the corresponding return level at the native grid scale, normalised to unity at the smallest scale of each dataset. The black line shows CombiPrecip (OBS), thin coloured lines show individual CPM

simulations, and the thick green and blue lines show the CPM and RCM ensemble medians, respectively."

R1C23: P. 22/ L. 476: A few studies have included analyses on the scale dependence of convection parameterizing RCM bias, e.g. Prein et al. (2016) and Fantini et al. (2018). The study presented here does so for the first time for CPMs. I suggest making this distinction.
Response: Thanks for the suggestion. We have modified the statement as: *"Overall, while previous studies have examined scale-dependence biases in convection-parameterizing RCMs* (Prein et al., 2016; Fantini et al., 2018)*, this study is the first to explore the scale-dependent performance of convection-permitting climate models, demonstrating the added value of CPMs over their driving RCMs in representing short-duration extremes at hydrologically relevant scales."*

Technical Corrections

R1C24: P. 2/ L. 44: The abbreviation CPMs has already been given in L. 42.
Response: Corrected. In the subsequent sentence, only the abbreviation is used.

R1C25: P. 3/ L. 85: Based on the map given in Fig. 1, Ticino would be situated in the Southern Alps, rather than in the south of the climatological region of the Alps. If this is correct, I would suggest capitalising the S (i.e. „the Southern Alps", rather than „the southern Alps").
Response: We agree and have capitalized "Southern Alps" throughout the manuscript where it refers to the physiographic region.

R1C26: P. 7/ Table 1: „WRF3.8.1DA (12km)" and „HadGEM3 (25km)" both need a space between the number and the unit.
Response: Corrected in Table 1: "12 km" and "25 km".

R1C27: P. 9/ L. 233: mean Bias → mean bias
Response: Corrected.

R1C28: P. 13/ L. 271: The observations at the smallest spatial scale are shown in Fig. 3a not in Fig. 3c.
Response: Corrected.

R1C29: P. 15/ L. 301 & Appendix/ Fig. S4: The native resolution RCM resolution is of approximately 144 km^2, not of 114 km^2.
Response: Corrected.

R1C30: P. 15/ L. 302: An area of 1296 km^2 equals to 3x3 RCM grids, not 6x6 grids.
Response: Corrected.

R1C31: P. 15/ L. 310: The abbreviation mAM is not used in the plots, rather „mean AM" is used.
Response: Corrected.

R1C32: P. 16/ L. 328: In the manuscript text, the reference to Fig. S7 comes before any reference to Fig. S5 and Fig. S6.
Response: We have reordered the supplementary figures so that they are referenced in numerical order.

R1C33: P. 19/ Fig. 8 & Appendix/ Fig. S1, S2, S3, S4, S7: I suggest including the abbreviation (AM) into the respective figure captions or into the main text.
Response: We have added the "AM" definition to the relevant figure captions and in the main text upon first use.

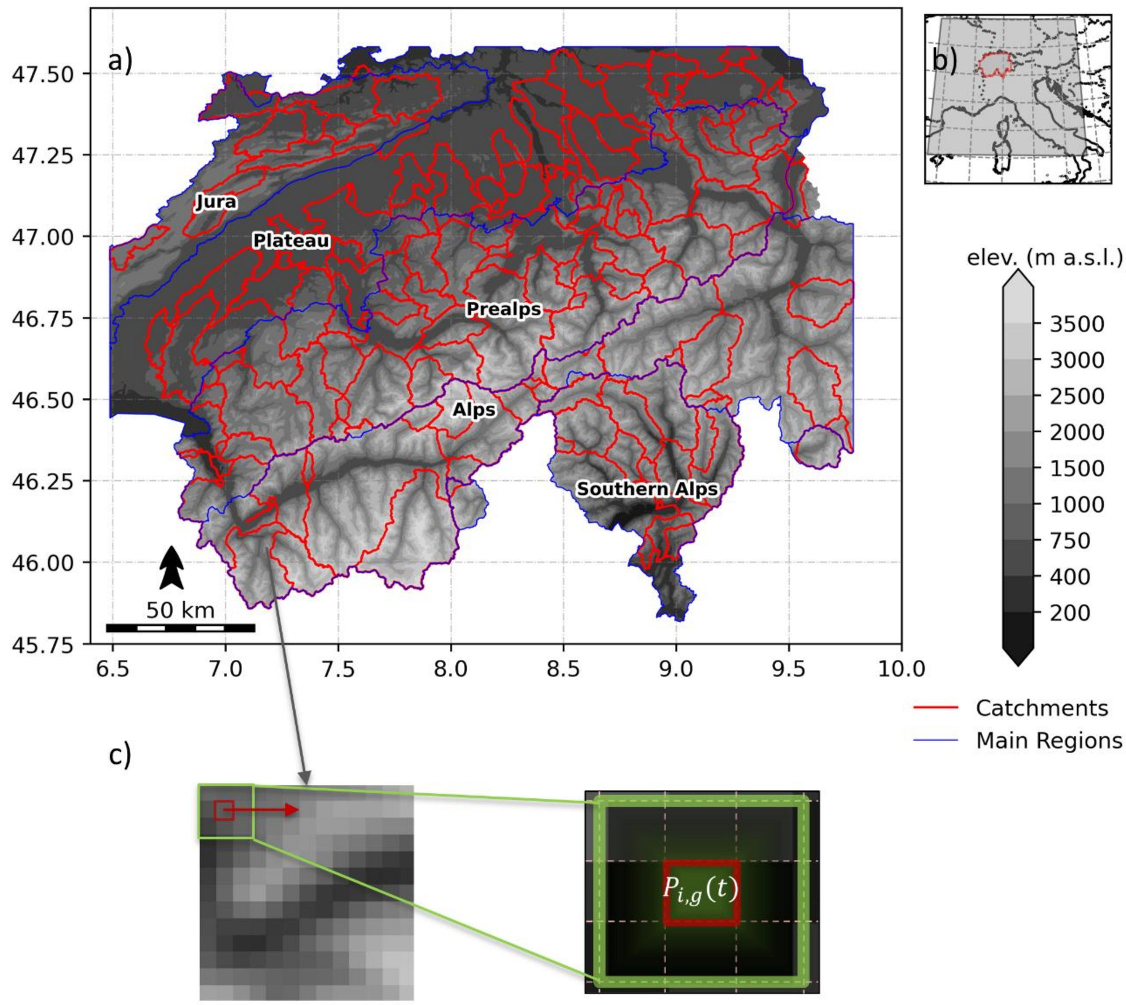
R1C34: P. 22/ L. 462: There is a spare point here.
Response: Corrected.

Reviewer 2 Comments and Responses

R2C1: The manuscript evaluates the added value of convection-permitting models (CPMs) compared to traditional regional climate models (RCMs) for precipitation extremes over the complex terrain of Switzerland. It covers various spatio-temporal scales (spatial scales of 10 to 5,000 km² and durations of 1 to 24 h). Furthermore, Swiss catchments are taken as example for spatial aggregation. The authors assess a multi-model ensemble from the CORDEX-FPS Convection (9 CPMs and 7 RCMs) addressing model uncertainties. As the provided time periods of 10 years are short, they apply the non-asymptotic Simplified Metastatistical Extreme Value (SMEV) framework. The evaluation is conducted in a comparison to a high-resolution observational product combining rain gauges and radar-based measurements (CombiPrecip). The study extends the current literature on CPMs and their added value with its focus on different spatial aggregations reflecting their applicability for hydrological impact modelling. Thereby, the data (climate model simulations and observational reference) and methods (SMEV & evaluation metrics) are timely and very well chosen. Generally, the manuscript is well structured, easy to follow, and well referenced. The quality of the visualization of results is excellent. The figures and results support the statements in the manuscript. Hence, it is from my perspective a valuable contribution to the scientific literature with a potential readership of climate modellers and (hydrological) impact modellers. It fits the scope of HESS (and NHESS), where a revision following RC1 will increase the relationship to hydrological processes.

In addition to the comments raised in RC1, I have a few comments:
Response: We thank you for your positive evaluation and for the constructive suggestions, whose implementation has strengthened the manuscript. Below, we provide our responses to each comment.

R2C2: L100: Fig 1: Nice map; can you add the CPM modelling domain(s) in 1b)?
Response: Thanks for the suggestion, we have added the CORDEX-FPS ALP-3 domain outline to Fig. 1b.



"Figure 1: (a) Study domain used to evaluate simulated precipitation products. The red lines represent the outlines of 143 nested catchments and the blue lines distinguish the main climatological regions of Switzerland (Jura, the Plateau, the Prealps, the Alps and the Southern Alps). (b) location of the study area in Europe and the CPM ALP-3 modeling domain. (c) an illustration of the moving window procedure for the computation of mean areal precipitation."

R2C3: L114: As you later mention undercatch in L426ff: Has the SwissMetNet data as input to CombiPrecip been corrected for undercatch?

Response: No. The SwissMetNet gauge measurements used as input to CombiPrecip are not corrected for undercatch. According to the MeteoSwiss product documentation (MeteoSwiss – Federal Office of Meteorology and Climatology, 2025), the gauge data undergo extensive quality control before their merging with the radar field, but no transfer functions for wind- or solid-precipitation-induced undercatch are applied. This limitation is already discussed in Section 4.3 of the manuscript and is one of the components of observational uncertainty we acknowledge there.

R2C4: L170: (and everywhere else): Present vs. past tense consistence (e.g. in 2.3.1 you use past tense, while before present tense is used)

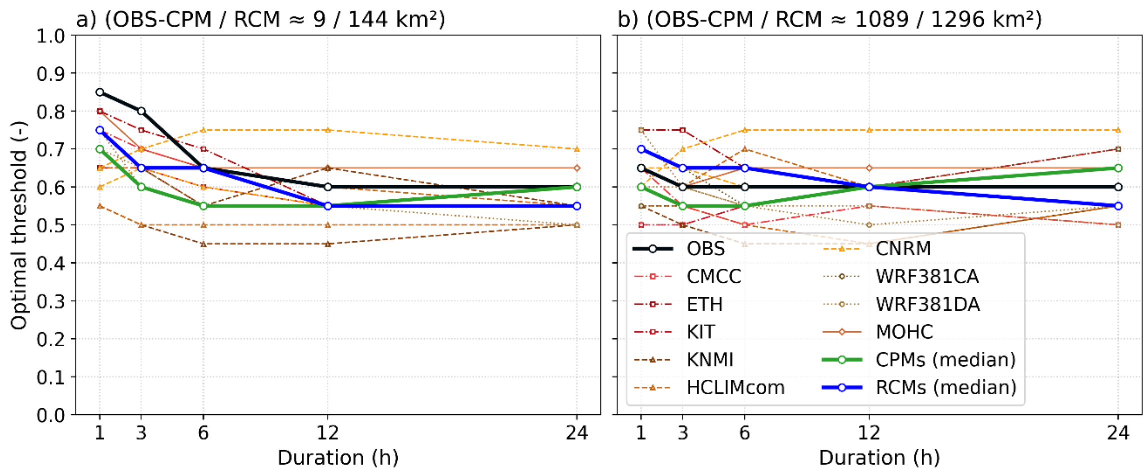
Response: Thanks for pointing this out. We maintained the consistency in the tenses throughout the manuscript.

R2C5: L217: Out of curiosity and own experience with the SMEV: Can you show/visualize test results for left-censoring threshold across the different durations? From my experience, the ideal threshold varies considerably with duration, with higher thresholds for shorter durations. In Dallan et al. (2024b) they also report: "The 90th percentile of the ordinary events is used as the left-censoring threshold for hourly duration, and the 85th percentile for longer durations."

Response: Thank you for your curiosity and question on this point. As you mentioned, the optimal left-censoring threshold can vary with duration, and we did test this systematically across our datasets. For each dataset, duration, and spatial scale, we computed the optimal left-censoring threshold separately for each grid cell, using the Weibull tail test of Marra et al. (2023). To summarize the result for each combination of duration and spatial scale, we then took the 80th percentile of the optimal thresholds across all grid cells within the study domain. This choice was made to have a more robust summary that intentionally emphasizes the upper part of the distribution of optima across the grid cells in the study domain. Below, we show the resulting optimal thresholds across durations for OBS, each individual model, and the CPM and RCM ensemble medians, at the smallest and aggregated spatial scales. We include this figure as Supplementary Information Figure S1 in the revised version, which is referenced in section 2.3.3 as follows:

"Following Dallan et al. (2023, 2024b) and employing the test by Marra et al. (2023), we selected the 85th percentile of the ordinary events as the left-censoring threshold across all durations (refer to Fig S1 for detailed description on the selection of optimal left-censoring threshold)."

As shown in Figure S1, the optimal threshold at the hourly duration spans roughly 0.55–0.85 across datasets, with OBS at 0.85. For longer durations, the spread tightens and most datasets cluster between 0.55 and 0.65. Given the volume of data to be analyzed and to apply SMEV consistently across the full datasets, durations, and spatial scales, we selected the 85th percentile as a uniform left-censoring threshold across all datasets, durations, and spatial scales. This ensures that the chosen threshold is at or above the optimum for every dataset/duration/scale combination, which guarantees that SMEV return levels are robust throughout the analysis.



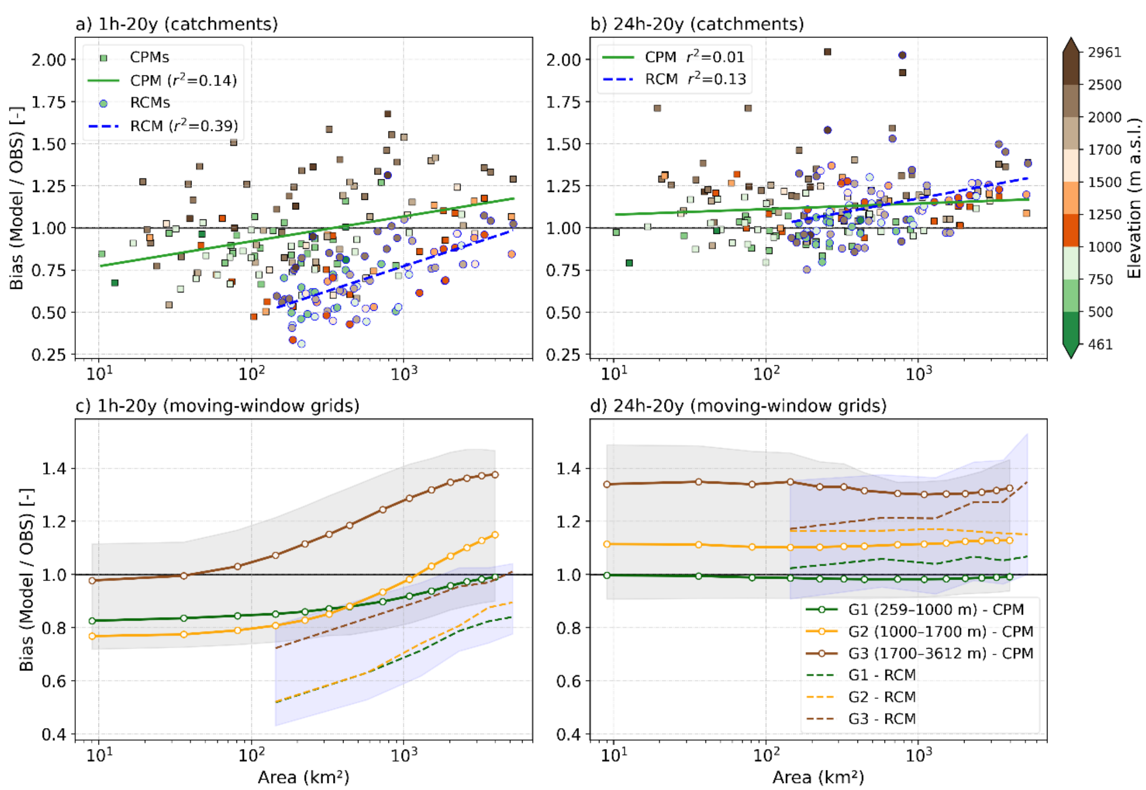
"Figure S1: Optimal left-censoring threshold for the Weibull tail of ordinary events, as a function of accumulation duration. Optimal thresholds were determined at each grid cell using the test of Marra et al. (2023), and are summarized here as the 80th percentile across the grid cells for each combination of dataset, duration, and spatial scale. Lines show results for CombiPrecip (OBS, black), individual CPM and RCM members (see legend), and the CPM and RCM ensemble medians (thick green and blue lines, respectively). (a) Smallest spatial scale: ≈9 km^2 for OBS and CPMs, ≈144 km^2 for RCMs. (b) Aggregated spatial scale: ≈1089 km^2 for OBS and CPMs (11×11 window), ≈1296 km^2 for RCMs (3×3 window). Finally, the 85th percentile is used as a uniform left-censoring threshold for all datasets, durations, and spatial scales in our SMEV application; this corresponds to the upper envelope of the optimal thresholds across the parameter space, ensuring that the chosen threshold lies at or above the optimum for every case considered."

R2C6: L229: Bias is typically defined as percentage bias: (I_Model - I_Obs)/I_Obs * 100%, while your metric is a ratio. I'd appreciate it if you switch to the more common definition of the bias throughout the manuscript.
Response: Thank you and we appreciate the suggestion. We retained the ratio-based formulation because it directly expresses multiplicative differences in return levels and is commonly used in SMEV-based evaluations (e.g., Dallan et al., 2023; Correa-Sánchez et al., 2025).

R2C7: L235ff: Elevation bands are not analyzed later in the main manuscript. Either move their definition to the caption of Fig. S6; or take Fig. S6 to the main article, e.g. integrating it into Fig. 7. I'd prefer the latter. Especially Fig. S6a shows interesting behaviour (e.g. crossing yellow and green lines with increasing area), which should also be discussed in the main manuscript.

Response: Thanks for this suggestion. We have integrated the elevation-stratified moving-window biases (previously Fig. S6) into Fig. 7 (Figure 8 in the revised version) as two new bottom panels (c, d), alongside the already catchment-based biases (panels a and b).

This integration resulted in the following revision of the last paragraph of Section 3.3: *"The elevation coloring shows that high-elevation catchments tend to have larger positive biases and include several strong outliers compared to low-elevation catchments. This pattern* *is confirmed to be more systematic when the moving-window analysis is stratified into three elevation bands. For 1h-20y return levels (Fig. 8c), CPM biases at low (G1) and mid (G2) elevations transition from underestimation (~0.80) at the smallest areal scales to near-unity or modest overestimation at the largest areas. High-elevation biases (G3) instead start near unity at the smallest areas and grow progressively to ~1.4 at the largest areas. The crossing of the G2 (mid) and G1 (low) curves around 100–1000 km² shows that mid-elevation grids show a stronger dependence of bias with area, similar to G3. RCMs (dashed lines in Fig. 8c) show the same ordering across elevation bands but with stronger underestimation at all areal scales, consistent with the catchment-based results in Fig. 8a. For 24h-20y return levels (Fig. 8d), biases within each elevation band are approximately constant across the full range of areal scales, with G3 grids exhibiting a persistent overestimation of ~1.3 and G1 cells remaining near unity. RCM biases follow the same elevation ordering."*



"Figure 8: Spatiotemporal scale and elevation dependence of CPM and RCM biases for 20-y extreme precipitation. Top row: catchment-based bias of CPM and RCM ensemble medians as a function of catchment area, for (a) 1h and (b) 24h return levels, computed across 143 Swiss catchments. Squares show CPM ensemble-median biases and circles show RCM ensemble-median biases; markers are colour-coded by the median catchment elevation. Solid green and blue lines denote the regression line fitted to the catchments' bias for CPMs and RCMs respectively. The bottom row shows moving-window grids stratified by elevation: regional-mean bias of the CPM (solid lines) and RCM (dashed lines) ensemble medians as a function of areal extent, for (c) 1h-20y and (d) 24h-20y return levels, stratified into three elevation bands: G1 (259–1000 m, green), G2 (1000–1700 m, orange), and G3 (1700–3612 m, brown). The upper

and bottom bands of the shaded area (grey for CPMs and blue for RCMs) indicate the corresponding 75th and 25th quartiles of grid points values for the G3 and G1 respectively."

R2C8: L240ff: A recent study by Brunner et al. (2025; disclaimer: I am co-author) uses a similar metric (standard deviation instead of CV) for climate extremes (also 1h-10y precip) in two global km-scale models comparing 10 x 10 km^2 to 100 x 100 km^2. Beyond the similar approach, it might be interesting to you due to the global context. The Alpine region is identified as a "variability hotspot" for precip extremes in Europe, while globally tropical regions show even higher sub-grid variability.
Response: Thank you for pointing us to this study, which provides a useful complementary perspective from a global, kilometre-scale modelling context. We have added a brief reference to Brunner et al. (2025) in Section 4.2.

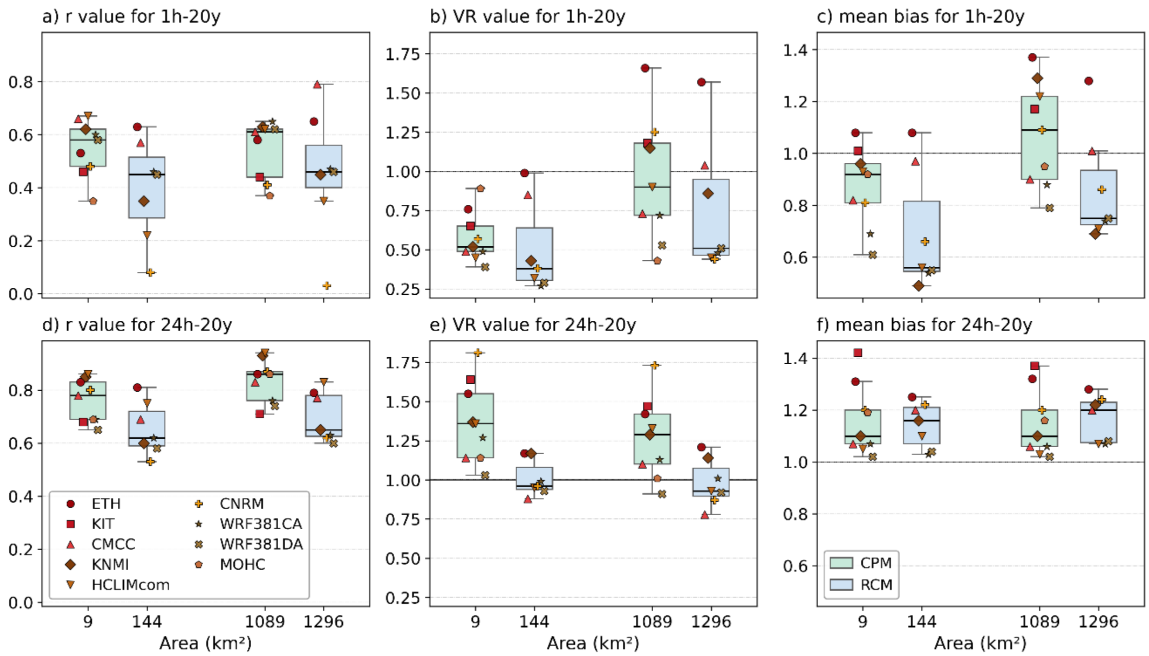
"This overly coherent structure of short-duration extremes has been linked to limitations in microphysics, turbulence, and effective resolution in CPMs (Fosser et al., 2015; Prein et al., 2015), and these resolution-related limitations are particularly pronounced over regions of complex topography (Brunner et al., 2025)*."*

R2C9: L246ff: ARF not shown in the main manuscript, but in Fig. S8. Either move the description to the caption of S8, or add S8 to the main manuscript.
Response: Thanks for your suggestion. We have moved Fig. S8 to the main Results section of the manuscript. Please see comment R1C22 for the subsequent changes made to the text.

R2C10: L251: 1h-20yr: I understand what you mean, but I'd prefer if you introduce the notation once with the long version: "20-year return levels of hourly precipitation (1h-20yr)"
Response: Thanks for pointing this out. *"The spatial distribution of the 20-year return levels of hourly precipitation (1h-20y) for CombiPrecip (OBS), CPMs and RCMs is presented in Fig. 2 for the smallest and for the ≈1000 km^2 spatial scale."*

R2C11: L261ff: Please add a table in the result section, where r, VR, and bias are shown for each model and duration. This would be a valuable overview of climate model performance.
Response: Thank you for this useful suggestion. We have added a new figure (Fig. 4 in the revised version) to Section 3.1 of the results, summarizing the spatial correlation (r), variability ratio (VR), and mean bias for each individual model member, at their smallest area and at the ≈10^3 km^2 areal aggregation scale, for both 1-h and 24-h durations. We also report a table corresponding to the values shown in the new Fig. 4 in SI section.

"As shown in Fig. 4 (see also Table S1 in Supplementary Information), both r and VR vary substantially across individual CPM members for 1h-20y return levels at the native scale (r = 0.35–0.67, VR = 0.39–0.89), with the corresponding RCM statistics being systematically lower. After aggregation to an order of ≈10^3 km^2, most CPMs show improvement in VR values, moving closer to unity. For the 24h-20y return levels, all members show higher correlations and variability ratios closer to unity already at the native scale, indicating a better reproduction of the spatial organization of daily extremes across both ensembles. Mean bias values follow the areal- and duration-dependent patterns examined in detail in Section 3.3. ."



"Figure 4: Per-member spatial correlation (r, panels a and d), variability ratio (VR, panels b and e), and mean bias (panels c and f) of CPM and RCM return levels relative to CombiPrecip (OBS), for 1h-20y (top row) and 24h-20y (bottom row) return levels. Statistics are reported at two spatial scales: smallest area (≈9 km^2 for CPMs and ≈144 km^2 for RCMs) and aggregated area (≈1089 km^2 for CPMs and ≈1296 km^2 for RCMs). VR is the ratio of model-to-OBS spatial standard deviations; mean bias is the model-to-OBS ratio of return levels averaged over the study domain. Box color denotes model family (light green for CPM and light blue for RCM) and the black line marks the median. Individual models are shown with a unique marker-color combination as indicated in the legend."

"Table S1: Per-member spatial correlation (r), variability ratio (VR), and mean bias of CPM and RCM return levels relative to CombiPrecip (OBS), for (a) 1h-20y and (b) 24h-20y return levels. Statistics are reported at two spatial scales: smallest area (≈9 km^2 for CPMs and ≈144 km^2 for RCMs) and aggregated area (≈1089 km^2 for CPMs and ≈1296 km^2 for RCMs). VR is the ratio of model-to-OBS spatial standard deviations; mean bias is the model-to-OBS ratio of return levels averaged over the study domain. Em-dashes indicate cases where the driving RCM is not available (KIT and MOHC; see Table 1)."

(a) 1h-20y return level

Model	CPM ≈ 9 km² – RCM ≈ 144 km²						CPM ≈ 1089 km² – RCM ≈ 1296 km²					
	r		***VR***		**mean bias**		***r***		***VR***		**mean bias**	
	CPM	**RCM**	**CPM**	**RCM**	**CPM**	**RCM**	**CPM**	**RCM**	**CPM**	**RCM**	**CPM**	**RCM**
CMCC	0.66	0.57	0.49	0.85	0.81	0.97	0.61	0.79	0.73	1.04	0.90	1.01
ETH	0.53	0.63	0.76	0.99	1.08	1.08	0.58	0.65	1.66	1.57	1.37	1.27
KIT	0.46	—	0.65	—	1.01	—	0.44	—	1.18	—	1.17	—

Model	CPM ≈ 9 km² – RCM ≈ 144 km²						CPM ≈ 1089 km² – RCM ≈ 1296 km²					
	r		***VR***		**mean bias**		***r***		***VR***		**mean bias**	
	CPM	**RCM**	**CPM**	**RCM**	**CPM**	**RCM**	**CPM**	**RCM**	**CPM**	**RCM**	**CPM**	**RCM**
KNMI	0.62	0.35	0.52	0.43	0.96	0.49	0.63	0.45	1.15	0.86	1.29	0.69
HCLIMcom	0.67	0.22	0.45	0.32	0.93	0.57	0.62	0.35	0.90	0.45	1.22	0.71
CNRM	0.48	0.08	0.57	0.38	0.81	0.66	0.41	0.03	1.25	0.44	1.09	0.86
MOHC	0.35	—	0.89	—	0.92	—	0.37	—	0.43	—	0.95	—
WRF381CA	0.60	0.46	0.49	0.27	0.69	0.54	0.65	0.47	0.72	0.48	0.88	0.74
WRF381DA	0.58	0.45	0.39	0.29	0.61	0.55	0.62	0.46	0.53	0.51	0.79	0.75

(b) 24h-20y return level

Model	CPM ≈ 9 km² – RCM ≈ 144 km²						CPM ≈ 1089 km² – RCM ≈ 1296 km²					
	r		***VR***		**mean bias**		***r***		***VR***		**mean bias**	
	CPM	**RCM**	**CPM**	**RCM**	**CPM**	**RCM**	**CPM**	**RCM**	**CPM**	**RCM**	**CPM**	**RCM**
CMCC	0.78	0.69	1.14	0.88	1.07	1.20	0.83	0.77	1.10	0.78	1.06	1.20
ETH	0.83	0.81	1.55	1.17	1.31	1.25	0.86	0.79	1.42	1.21	1.32	1.28
KIT	0.68	—	1.64	—	1.42	—	0.71	—	1.47	—	1.40	—
KNMI	0.85	0.60	1.37	1.17	1.10	1.16	0.93	0.65	1.29	1.14	1.10	1.22
HCLIMcom	0.86	0.75	1.36	0.95	1.05	1.10	0.94	0.83	1.33	0.93	1.03	1.07
CNRM	0.80	0.53	1.81	0.96	1.22	1.22	0.87	0.62	1.73	0.87	1.21	1.24
MOHC	0.69	—	1.14	—	1.21	—	0.86	—	1.01	—	1.16	—
WRF381CA	0.69	0.62	1.27	0.99	1.07	1.03	0.76	0.63	1.13	1.01	1.06	1.07
WRF381DA	0.65	0.58	1.03	0.93	1.02	1.04	0.74	0.60	0.91	0.92	1.02	1.08

R2C12: L345: Can you describe (and/or later discuss) also the "ranks" of RCMs and their according CPMs? E.g. in Fig. 6a, ETH-RCM is the wettest and ETH-CPM as well. However, KNMI-RCM is drier than the RCM-median, while KNMI-CPM is wetter than the CPM-median. You don't need to describe all of these ranks in detail, but draw a general statement about the consistency of the ranking between RCMs and according CPMs.

Response: We very much appreciate this suggestion. We have added a short paragraph to Section 3.3 commenting on the consistency of the rankings between CPMs and their driving RCMs, drawing on both Fig. 6 (Fig. 7 in the revised version) and the new Fig. 4. We highlight that some chains preserve their relative wet/dry rank across the downscaling step (e.g., ETH and the two WRF configurations), while others show a clear shift, with the CPM lying substantially closer to OBS than the driving RCM (notably KNMI, HCLIMcom and CNRM), which shows that the convection-permitting downscaling step can substantially correct the wet/dry bias inherited from the driving RCM, particularly for short-duration extremes.

"Comparing each CPM with its driving RCM across the range of areas (Fig. 7 and Fig. 4) shows that the wet/dry ranking is only partially preserved across the convection-permitting downscaling step, and that this degree of preservation depends on duration. For 24h-20y return levels, the ranking is largely consistent: 5 of 7 paired chains keep the same wet/dry position relative to their ensemble median at both native and aggregated scales (except for the models from CMCC and KNMI). For 1h-20y return levels, the ranking is less consistent, with only 3 of 7 chains preserving their rank at the native scale (ETH and the two WRF configurations). Of the four chains with different ranks at 1-h, HCLIMcom and KNMI are among the wettest CPMs while their driving RCMs are among the driest in the RCM ensemble, whereas CMCC and CNRM show the opposite pattern. After aggregation to ~1000 km^2, rank consistency improves slightly (4 of 7 chains). Overall, a CPM and its driving RCM do not necessarily share the same wet/dry rank within their respective ensembles, and this is more often the case for short-duration extremes than for longer-duration and daily extremes."

R2C13: L365: Any idea why MOHC's High-Res GCM - CPM chain shows such a distinct behaviour of within-window CV?

Response: Yes, and not only for the CV but also for all other metrics that we considered in our analysis the MOHC's model shows a distinct behavior compared to other members in the ensemble (Fig.6, Fig.8, Fig.9, and Fig. S5). The underlying causes cannot be conclusively identified within the scope of the present analysis. However, we note that the MOHC model primarily differs from the rest of the ensemble members in terms of its nesting strategy: MOHC is the only modelling chain that directly nests a high-resolution (25-km) global model into the 2.2-km CPM domain, without an intermediate ~12-km RCM; this means that the CPM receives lateral boundary conditions at a spatial scale that is significantly different with respect to the rest of the ensemble, which can affect how mesoscale convective structures develop both at the boundaries and in the interior. Beyond this feature, each CPM in the ensemble has its own model physics and configuration, which can also contribute to differences in the spatial coherence of simulated extremes. Isolating the relative contribution of these factors would require a much more dedicated process-level analysis beyond the scope of our evaluation.

R2C14: L388 COSMO wet bias is also found by Rybka et al., 2023 for Germany (see Figs 3 & 4 therein), you might want to add this reference.

Response: Thank you for sharing this reference. We have added Rybka et al. (2023) to support our finding of a COSMO-family wet bias. *"In our ensemble, COSMO-based models (ETH, KIT) systematically overestimate, while WRF configurations underestimate observed 1-h extremes across all areas. The COSMO wet bias is consistent with similar findings over Germany* (Rybka et al., 2023)*."*

R2C15: L434: Here, I suggest an extension of the limitation section:

It's not only internal variability across decades; The mismatch of periods might lead to mismatches in the observed large-scale climate modes during these periods, which should however, rather drive deviations in longer-duration (24h)-rainfall extremes than in localized convective short-duration hourly extremes (see e.g. Haslinger et al., 2025:

https://doi.org/10.1038/s41586-025-08647-2). Though, internal climate variability would also be present and a major uncertainty factor when comparing the same time periods. ICV even manifests in RCM simulations of the same model driven by the same lateral boundary conditions (Alexandru et al., 2007: https://doi.org/10.1175/MWR3456.1). In turn, from the perspective of predictability, Judt (2018 / 2020) describes moist convection as the principal driver of "forecast error growth", highlighting the large variability related to this process. Beyond the process level, the effect of ICV on extreme precipitation metrics is also governed by the degree of spatial aggregation (see e.g. Aalbers et al., 2025: https://doi.org/10.1029/2025JD043768). Further, ICV-driven uncertainty is closely linked to the sample size of the statistical assessment, where 10 years are a clear limitation. Even though the SMEV has proven to outperform traditional EVT methods (GEV block maxima / GP POT) on short sample sizes, uncertainties remain large. There you should add a few sentences and discuss this uncertainty.

Response: We thank you for this thoughtful suggestion and for the relevant references. We have expanded the discussion of the limitations of the non-overlapping observational and simulation periods to better address the uncertainties involve.

"A main limitation of this study concerns the observational reference and the simulation period. CombiPrecip remains subject to terrain-relevant uncertainty in complex topography, particularly at higher elevations (Germann et al., 2006; Sideris et al., 2014). In addition, the CombiPrecip record (2005–2024) does not overlap with the CPM historical decade (1996–2005), meaning that differences between the two periods can influence the absolute magnitude of the estimated return levels and, therefore, the apparent Model/OBS bias. Part of this difference reflects shifts in large-scale circulation patterns active during each decade, which can leave a more visible signature in longer-duration (24-h) extremes that are tied to synoptic forcing, and a comparatively smaller one in localised hourly convective extremes (Haslinger et al., 2025)*. In addition, there is the uncertainty related to internal climate variability, which would remain a major source of uncertainty even for matched periods. For instance, even when driven by identical boundary conditions, RCM ensemble members can produce measurably different extreme-precipitation statistics* (Alexandru et al., 2007)*, reflecting the inherent chaotic nature of the atmospheric system. This effect is particularly pronounced for sub-daily extremes, where small-scale uncertainties associated with moist convection grow rapidly and limit the predictability of short-duration precipitation* (Judt, 2018)*. The 10-year length of the CPM simulations adds a further sampling uncertainty; while SMEV is designed to provide stable return-level estimates from short records* (Dallan et al., 2024; Marra et al., 2019)*, it reduces estimation uncertainty but cannot compensate for variability tied to large-scale climate modes. Observational analyses further suggest that heavy-precipitation intensities — especially at short durations (10 min to 3 h) — have increased in Switzerland in recent decades (Bauer and Scherrer, 2024), implying that potential non-stationarity may also contribute to differences between the two periods. These limitations are most relevant for interpreting absolute bias levels, whereas our main conclusions focus on relative behaviour across aggregation scales and durations (e.g., sign changes at 1–3 h and convergence toward scale-invariant biases at ≥6 h), which are expected to be less sensitive to the uncertainties discussed above."*

References

Alexandru, A., De Elia, R., and Laprise, R.: Internal Variability in Regional Climate Downscaling at the Seasonal Scale, Mon. Weather Rev., 135, 3221–3238, https://doi.org/10.1175/MWR3456.1, 2007.

Ban, N., Caillaud, C., Coppola, E., Pichelli, E., Sobolowski, S., Adinolfi, M., Ahrens, B., Alias, A., Anders, I., Bastin, S., Belušić, D., Berthou, S., Brisson, E., Cardoso, R. M., Chan, S. C., Christensen, O. B., Fernández, J., Fita, L., Frisius, T., Gašparac, G., Giorgi, F., Goergen, K., Haugen, J. E., Hodnebrog, Ø., Kartsios, S., Katragkou, E., Kendon, E. J., Keuler, K., Lavin-Gullon, A., Lenderink, G., Leutwyler, D., Lorenz, T., Maraun, D., Mercogliano, P., Milovac, J., Panitz, H.-J., Raffa, M., Remedio, A. R., Schär, C., Soares, P. M. M., Srnec, L., Steensen, B. M., Stocchi, P., Tölle, M. H., Truhetz, H., Vergara-Temprado, J., De Vries, H., Warrach-Sagi, K., Wulfmeyer, V., and Zander, M. J.: The first multi-model ensemble of regional climate simulations at kilometer-scale resolution, part I: evaluation of precipitation, Clim. Dyn., 57, 275–302, https://doi.org/10.1007/s00382-021-05708-w, 2021.

Barton, Y., Sideris, I. V., Raupach, T. H., Gabella, M., Germann, U., and Martius, O.: A multi-year assessment of sub-hourly gridded precipitation for Switzerland based on a blended radar—Rain-gauge dataset, Int. J. Climatol., 40, 5208–5222, https://doi.org/10.1002/joc.6514, 2020.

Borga, M., Stoffel, M., Marchi, L., Marra, F., and Jakob, M.: Hydrogeomorphic response to extreme rainfall in headwater systems: Flash floods and debris flows, J. Hydrol., 518, 194–205, https://doi.org/10.1016/j.jhydrol.2014.05.022, 2014.

Breinl, K., Müller-Thomy, H., and Blöschl, G.: Space–Time Characteristics of Areal Reduction Factors and Rainfall Processes, J. Hydrometeorol., 21, 671–689, https://doi.org/10.1175/JHM-D-19-0228.1, 2020.

Brunner, L., Poschlod, B., Dutra, E., Fischer, E. M., Martius, O., and Sillmann, J.: A global perspective on the spatial representation of climate extremes from km-scale models, Environ. Res. Lett., 20, 074054, https://doi.org/10.1088/1748-9326/ade1ef, 2025.

Correa-Sánchez, N., Dallan, E., Marra, F., Fosser, G., and Borga, M.: Orographic control on bias and uncertainty in extreme sub-daily precipitation simulations from a convection-permitting ensemble, J. Hydrol., 659, 133324, https://doi.org/10.1016/j.jhydrol.2025.133324, 2025.

Dallan, E., Marra, F., Fosser, G., Marani, M., Formetta, G., Schär, C., and Borga, M.: How well does a convection-permitting regional climate model represent the reverse orographic effect of extreme hourly precipitation?, Hydrol. Earth Syst. Sci., 27, 1133–1149, https://doi.org/10.5194/hess-27-1133-2023, 2023.

Dallan, E., Marra, F., Fosser, G., Marani, M., and Borga, M.: Dynamical Factors Heavily Modulate the Future Increase of Sub-Daily Extreme Precipitation in the Alpine-Mediterranean Region, Earths Future, 12, e2024EF005185, https://doi.org/10.1029/2024EF005185, 2024.

De Michele, C., Kottegoda, N. T., and Rosso, R.: The derivation of areal reduction factor of storm rainfall from its scaling properties, Water Resour. Res., 37, 3247–3252, https://doi.org/10.1029/2001WR000346, 2001.

Estermann, R., Rajczak, J., Velasquez, P., Lorenz, R., and Schär, C.: Projections of Heavy Precipitation Characteristics Over the Greater Alpine Region Using a Kilometer–Scale Climate Model Ensemble, J. Geophys. Res. Atmospheres, 130, e2024JD040901, https://doi.org/10.1029/2024JD040901, 2025.

Fantini, A., Raffaele, F., Torma, C., Bacer, S., Coppola, E., Giorgi, F., Ahrens, B., Dubois, C., Sanchez, E., and Verdecchia, M.: Assessment of multiple daily precipitation statistics in ERA-Interim driven Med-CORDEX and EURO-CORDEX experiments against high resolution observations, Clim. Dyn., 51, 877–900, https://doi.org/10.1007/s00382-016-3453-4, 2018.

Fosser, G., Gaetani, M., Kendon, E. J., Adinolfi, M., Ban, N., Belušić, D., Caillaud, C., Careto, J. A. M., Coppola, E., Demory, M.-E., De Vries, H., Dobler, A., Feldmann, H., Goergen, K., Lenderink, G., Pichelli, E., Schär, C., Soares, P. M. M., Somot, S., and Tölle, M. H.: Convection-permitting climate models offer more certain extreme rainfall projections, Npj Clim. Atmospheric Sci., 7, 51, https://doi.org/10.1038/s41612-024-00600-w, 2024.

Gericke, O. and Pietersen, J. P. J.: Estimation of areai reduction factors using daily rainfall data and a geographically centred approach, J. South Afr. Inst. Civ. Eng., 62, 20–31, https://doi.org/10.17159/2309-8775/2020/v62n4a3, 2020.

Ghaemi, E., Gabella, M., Foelsche, U., Sideris, I., and Nerini, D.: The effect of altitude on the uncertainty of radar-based precipitation estimates over Switzerland, Int. J. Remote Sens., 44, 2495–2517, https://doi.org/10.1080/01431161.2023.2203339, 2023.

Haslinger, K., Breinl, K., Pavlin, L., Pistotnik, G., Bertola, M., Olefs, M., Greilinger, M., Schöner, W., and Blöschl, G.: Increasing hourly heavy rainfall in Austria reflected in flood changes, Nature, 639, 667–672, https://doi.org/10.1038/s41586-025-08647-2, 2025.

Judt, F.: Insights into Atmospheric Predictability through Global Convection-Permitting Model Simulations, J. Atmospheric Sci., 75, 1477–1497, https://doi.org/10.1175/JAS-D-17-0343.1, 2018.

Kendon, E. J., Prein, A. F., Senior, C. A., and Stirling, A.: Challenges and outlook for convection-permitting climate modelling, Philos. Trans. R. Soc. Math. Phys. Eng. Sci., 379, 20190547, https://doi.org/10.1098/rsta.2019.0547, 2021.

Lucas-Picher, P., Brisson, E., Caillaud, C., Alias, A., Nabat, P., Lemonsu, A., Poncet, N., Cortés Hernandez, V. E., Michau, Y., Doury, A., Monteiro, D., and Somot, S.: Evaluation of the convection-permitting regional climate model CNRM-AROME41t1 over Northwestern Europe, Clim. Dyn., 62, 4587–4615, https://doi.org/10.1007/s00382-022-06637-y, 2024.

Marchi, L., Borga, M., Preciso, E., and Gaume, E.: Characterisation of selected extreme flash floods in Europe and implications for flood risk management, J. Hydrol., 394, 118–133, https://doi.org/10.1016/j.jhydrol.2010.07.017, 2010.

Marra, F., Zoccatelli, D., Armon, M., and Morin, E.: A simplified MEV formulation to model extremes emerging from multiple nonstationary underlying processes, Adv. Water Resour., 127, 280–290, https://doi.org/10.1016/j.advwatres.2019.04.002, 2019.

Marra, F., Amponsah, W., and Papalexiou, S. M.: Non-asymptotic Weibull tails explain the statistics of extreme daily precipitation, Adv. Water Resour., 173, 104388, https://doi.org/10.1016/j.advwatres.2023.104388, 2023.

MeteoSwiss – Federal Office of Meteorology and Climatology: CombiPrecip precipitation data, 2025.

Panziera, L., Gabella, M., Germann, U., and Martius, O.: A 12-year radar-based climatology of daily and sub-daily extreme precipitation over the Swiss Alps, Int. J. Climatol., 38, 3749–3769, https://doi.org/10.1002/joc.5528, 2018.

Prein, A. F., Gobiet, A., Truhetz, H., Keuler, K., Goergen, K., Teichmann, C., Fox Maule, C., Van Meijgaard, E., Déqué, M., Nikulin, G., Vautard, R., Colette, A., Kjellström, E., and Jacob, D.: Precipitation in the EURO-CORDEX 0.11° and 0.44° simulations: high resolution, high benefits?, Clim. Dyn., 46, 383–412, https://doi.org/10.1007/s00382-015-2589-y, 2016.

Rasmussen, S. H., Christensen, J. H., Drews, M., Gochis, D. J., and Refsgaard, J. C.: Spatial-Scale Characteristics of Precipitation Simulated by Regional Climate Models and the Implications for Hydrological Modeling, J. Hydrometeorol., 13, 1817–1835, https://doi.org/10.1175/JHM-D-12-07.1, 2012.

Reszler, C., Switanek, M. B., and Truhetz, H.: Convection-permitting regional climate simulations for representing floods in small- and medium-sized catchments in the Eastern Alps, Nat. Hazards Earth Syst. Sci., 18, 2653–2674, https://doi.org/10.5194/nhess-18-2653-2018, 2018.

Rybka, H., Haller, M., Brienen, S., Brauch, J., Früh, B., Junghänel, T., Lengfeld, K., Walter, A., and Winterrath, T.: Convection-permitting climate simulations with COSMO-CLM for Germany: Analysis of present and future daily and sub-daily extreme precipitation, Meteorol. Z., 32, 91–111, https://doi.org/10.1127/metz/2022/1147, 2023.

Schär, C., Fuhrer, O., Arteaga, A., Ban, N., Charpilloz, C., Di Girolamo, S., Hentgen, L., Hoefler, T., Lapillonne, X., Leutwyler, D., Osterried, K., Panosetti, D., Rüdisühli, S., Schlemmer, L., Schulthess, T. C., Sprenger, M., Ubbiali, S., and Wernli, H.: Kilometer-Scale Climate Models: Prospects and Challenges, Bull. Am. Meteorol. Soc., 101, E567–E587, https://doi.org/10.1175/BAMS-D-18-0167.1, 2020.

Svensson, C. and Jones, D. A.: Review of methods for deriving areal reduction factors, J. Flood Risk Manag., 3, 232–245, https://doi.org/10.1111/j.1753-318X.2010.01075.x, 2010.

Viglione, A. and Blöschl, G.: On the role of storm duration in the mapping of rainfall to flood return periods, Hydrol. Earth Syst. Sci., 13, 205–216, https://doi.org/10.5194/hess-13-205-2009, 2009.