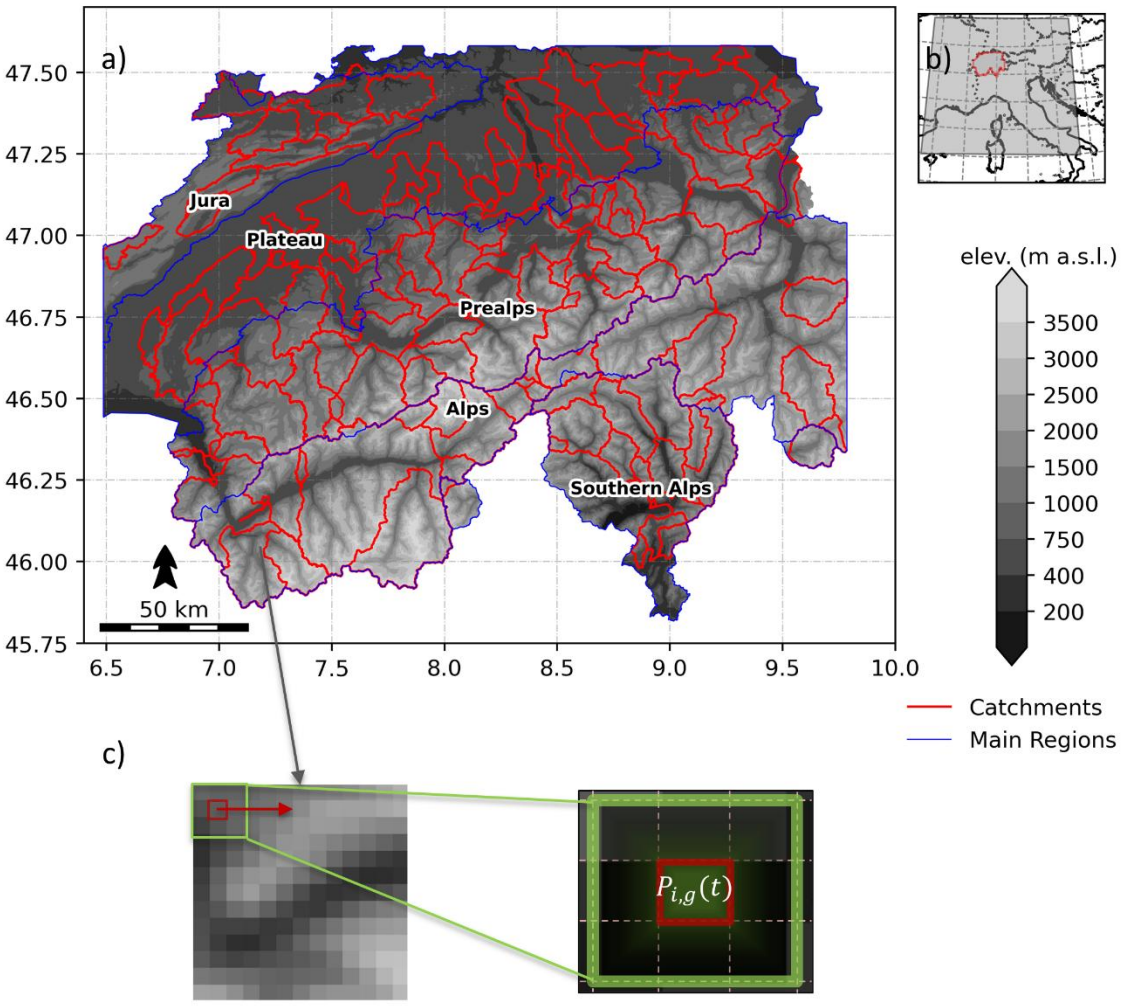
We organize this Response to Reviewers document in the following way. Reviewers' comments are numbered as RxCy, where X is the reviewer number, and Y is the reviewer's comment number. Our responses are reported in blue, and new or modified text highlighted in *italics red*.

Reviewer 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 km^2 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?

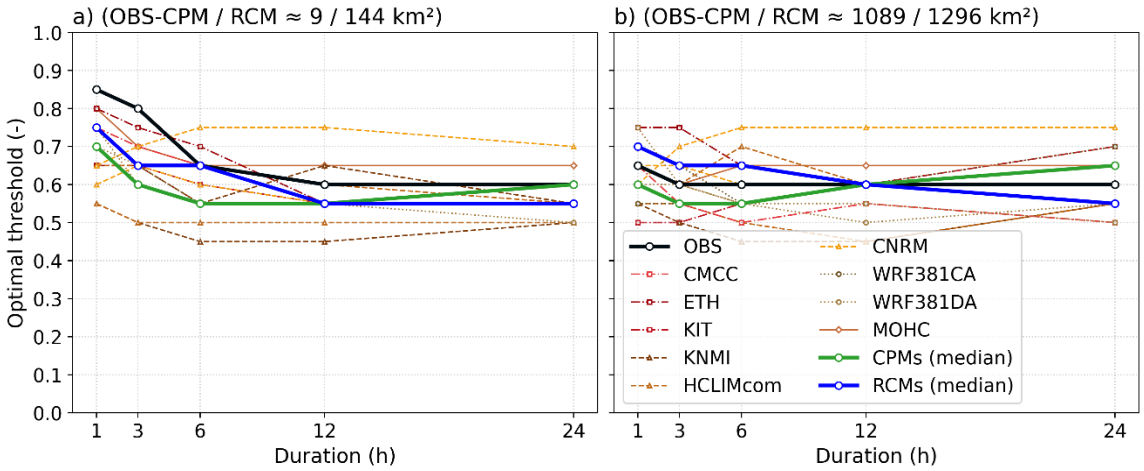
Response: No. The SwissMetNet gauge measurements used as input to CombiPrecip are not corrected for undercatch. According to the MeteoSwiss product documentation (MeteoSwiss – Federal Office of Meteorology and Climatology, 2025), the gauge data undergo extensive quality control before their merging with 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.

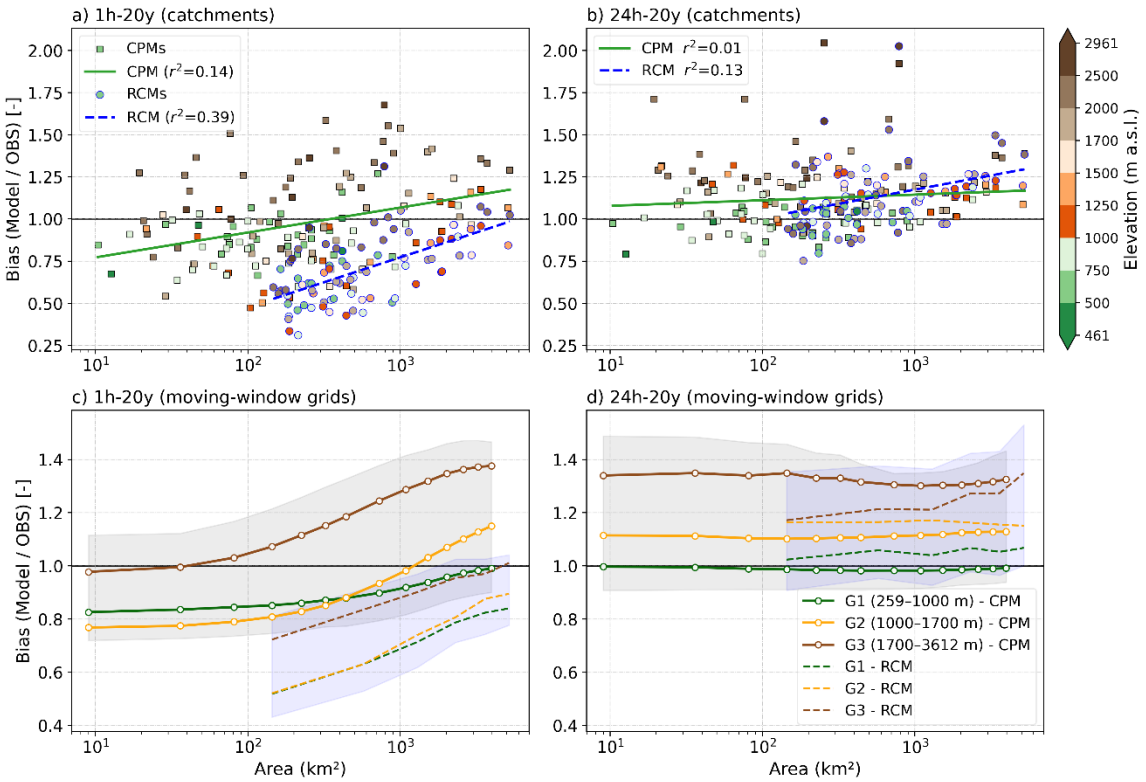
Response: Thank you and we appreciate the suggestion. We retained the ratio-based formulation because it directly expresses multiplicative differences in return levels and is commonly used in SMEV-based evaluations (e.g., Dallan et al., 2023; Correa-Sánchez et al., 2025).

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.

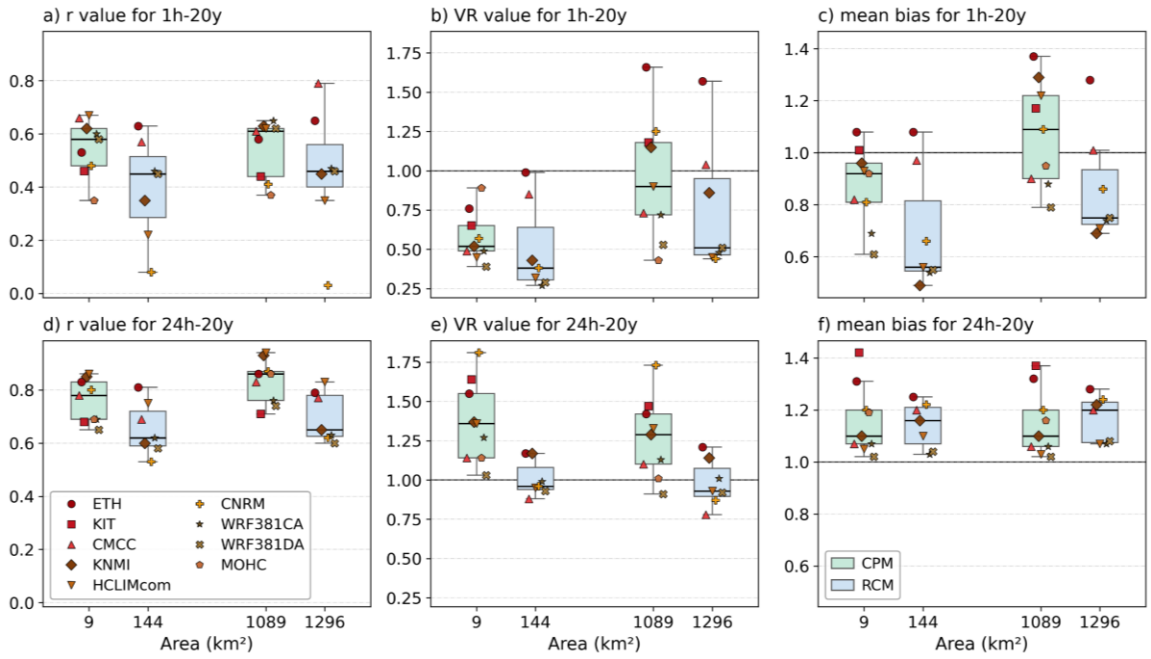
"This overly coherent structure of short-duration extremes has been linked to limitations in microphysics, turbulence, and effective resolution in CPMs (Fosser et al., 2015; Prein et al., 2015), and these resolution-related limitations are particularly pronounced over regions of complex topography (Brunner 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 km² for CPMs and ≈144 km² for RCMs) and aggregated area (≈1089 km² for CPMs and ≈1296 km² 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 km² for CPMs and ≈144 km² for RCMs) and aggregated area (≈1089 km² for CPMs and ≈1296 km² 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., & Laprise, R. (2007). Internal Variability in Regional Climate Downscaling at the Seasonal Scale. *Monthly Weather Review*, *135*(9), 3221–3238. https://doi.org/10.1175/MWR3456.1

Brunner, L., Poschlod, B., Dutra, E., Fischer, E. M., Martius, O., & Sillmann, J. (2025). A global perspective on the spatial representation of climate extremes from km-scale models. *Environmental Research Letters*, *20*(7), 074054. https://doi.org/10.1088/1748-9326/ade1ef

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

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

Dallan, E., Marra, F., Fosser, G., Marani, M., Formetta, G., Schär, C., & Borga, M. (2023). How well does a convection-permitting regional climate model represent the reverse orographic effect of extreme hourly precipitation? *Hydrology and Earth System Sciences*, *27*(5), 1133–1149. https://doi.org/10.5194/hess-27-1133-2023

Haslinger, K., Breinl, K., Pavlin, L., Pistotnik, G., Bertola, M., Olefs, M., Greilinger, M., Schöner, W., & Blöschl, G. (2025). Increasing hourly heavy rainfall in Austria reflected in flood changes. *Nature*, *639*(8055), 667–672. https://doi.org/10.1038/s41586-025-08647-2

Judt, F. (2018). Insights into Atmospheric Predictability through Global Convection-Permitting Model Simulations. *Journal of the Atmospheric Sciences*, *75*(5), 1477–1497. https://doi.org/10.1175/JAS-D-17-0343.1

Marra, F., Amponsah, W., & Papalexiou, S. M. (2023). Non-asymptotic Weibull tails explain the statistics of extreme daily precipitation. *Advances in Water Resources*, *173*, 104388. https://doi.org/10.1016/j.advwatres.2023.104388

Marra, F., Zoccatelli, D., Armon, M., & Morin, E. (2019). A simplified MEV formulation to model extremes emerging from multiple nonstationary underlying processes. *Advances in Water Resources*, *127*, 280–290. https://doi.org/10.1016/j.advwatres.2019.04.002

MeteoSwiss – Federal Office of Meteorology and Climatology. (2025). *CombiPrecip precipitation data*. https://www.meteoswiss.admin.ch/services-and-publications/service/weather-and-climate-products/combiprecip.html

Rybka, H., Haller, M., Brienen, S., Brauch, J., Früh, B., Junghänel, T., Lengfeld, K., Walter, A., & Winterrath, T. (2023). Convection-permitting climate simulations with COSMO-CLM for Germany: Analysis of present and future daily and sub-daily extreme precipitation. *Meteorologische Zeitschrift*, *32*(2), 91–111. https://doi.org/10.1127/metz/2022/1147