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:

L100: Fig 1: Nice map; can you add the CPM modelling domain(s) in 1b)?

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

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)

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.”

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.

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.

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 km² to 100 x 100 km². 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.

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.

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)"

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.

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.

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

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.

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.