

Author response to Anonymous Referee #1 (RC1)

Manuscript egusphere-2026-4111: "Cross-scale characteristics of extreme precipitation events under climate change" by Jingkun Yang, Benjamin Poschlod and Felix Ament

We thank Referee #1 for the careful and constructive review. Referee comments are shown in grey boxes, followed by our response and the planned revision.

Main planned revisions

- **Model evaluation:** new subsection summarising the evaluation of the COSMO-CLM simulations for extreme precipitation (Rybka et al., 2023).
- **Event selection:** xWEI computed for an enlarged pool of candidate events, so that the top-200 events are selected by xWEI.
- **Robustness:** maps of the event locations, analysis of three sub-regions, seasonal breakdown, alternative area weighting.
- **Driving data and scenario:** comparison of the MIROC5- and ERA5-driven simulations for their common period 1979–2000; addition of global warming levels; results framed as conditional on model chain and scenario.
- **Uncertainty:** dGEV parameter distributions, return period cap, leave-one-year-out test, fixed shape parameter, bootstrap confidence intervals.
- **Interpretation:** clearer definition of the "second-order" effect, extended limitations, corrections.
- **Correction:** During the revision process we identified a sign-convention mismatch in the GEV shape parameter between the R package used for fitting and the Python routine used for return-period calculation. We will correct this in the revised manuscript and recompute the according return periods, WEI and xWEI values.

This reply describes the planned revision; line numbers refer to the preprint.

General comments

Referee comment RC1.0

The study compares precipitation event characteristics for COSMO-CLM simulation data across different time periods, from the past to the future under the RCP8.5 scenario. The authors use the dGEV and adapted dGEV distributions to assess extreme rainfall return periods across durations, which is then applied within the xWEI framework to quantify event extremeness and identify the scale of maximum impact relevance.

While the manuscript is well written, and the authors demonstrate a strong statistical understanding, the scientific narrative requires strengthening. Specifically, there is a gap in the fundamental understanding of the climate model products (COSMO-CLM) and hydro-climatological phenomena that underpins the post-processing statistical frameworks. Therefore, there are methodological limitations that arise, which call into question the key findings and interpretations of the results - especially regarding the shift in trend towards more persistent and spatially extensive rainfall systems. These concerns have been detailed systematically below.

Thank you. We will describe COSMO-CLM and its known biases for extreme precipitation in more detail (RC1.2, RC1.5), test the robustness of the shift towards more persistent and extensive events (RC1.1, RC1.3, RC1.4, RC1.9), and phrase our conclusions as conditional on the model chain and scenario (RC1.6, RC1.7).

Specific comments

Referee comment RC1.1

In Line 115, the authors mention the use of a duration specific intensity threshold that is based on absolute values using the DWD severe weather warning threshold. While this is useful as a warning system, it is less than optimal when the focus is multi-hazard hydroclimatic risks. There is a geographic bias when statistically evaluated over large regions. Some areas (like the Alps) may naturally receive more rainfall, the lowland areas receive much lower rainfall on average, but also have higher economic activity (more dense, agricultural activity, etc.), and therefore far more economic damage potential. For this reason, if the study that chooses an absolute threshold with the top 200 extreme events, it would introduce a spatial bias. The authors should consider transitioning to a grid cell specific percentile baseline (like 99th percentile per duration per grid cell) to normalize the spatial heterogeneity or provide a rigorous statistical justification to prove that the current analysis does not skew the resulting datasets towards specific geographic sub-regions.

We agree that the nationwide warning thresholds (W3) yield a spatially non-uniform set of candidate events: W3-based events cluster in mountainous regions, mainly for durations of 12 h and longer, whereas return-period-based events are distributed almost uniformly (Lengfeld et al., 2021). The W3 threshold is used only for candidate-event detection; WEI and xWEI are computed from local return periods and are thus normalised by the grid-cell climatology, as suggested. We keep W3 for consistency with CatRaRE and Voit and Heistermann (2022) and because of its link to the DWD warnings. To show that the top-200 selection is not skewed towards specific regions, we will map the candidate and top-200 events of all periods (Figure R1) and test their regional shares against the area shares. We will also clarify that we quantify the hazard, not exposure or damage potential.

Planned changes in the revised manuscript

- Methods: role of the W3 threshold (detection) versus the locally normalised xWEI.
- Supplement: maps of candidate and top-200 events and test of their regional distribution; limitation in the Discussion.

Referee comment RC1.2

Please clarify what are the statistical scores (such as RMSE, Bias, MAE, FSS, SR, Taylor Diagrams) for extreme precipitation of the COSMO CPM model used, with respect to observed datasets, including, if necessary, spatial maps indicating these scores. Model biases in COSMO are predominant across sub-regions due to terrain, orographic changes, etc. Therefore, a positive bias in the Alps compared to a negative bias in the lowlands, for instance, could skew the results further towards specific sub-regions when using CPM datasets. Furthermore, the performance of the CPM model may drop significantly when considering extreme events. This should be justified in the study with respect to observed data, with the proper use of citations and/or analysis.

The simulations were evaluated by Rybka et al. (2023) against the radar climatology RADKLIM and the gauge-based KOSTRA-DWD, including maps, bias, pattern correlation and Taylor diagrams. We will summarise these scores in the manuscript (Table R1). Extremes of 12 h and longer are reproduced well, while short-duration extremes are moderately overestimated. Here, we note that RADKLIM as radar product underestimates extreme precipitation, while KOSTRA (based on rain gauges; not corrected for undercatch) is the main reference product for extreme intensities.

In the Taylor diagrams, the MIROC5-driven historical simulation performs equal to or better than nine CMIP5 EURO-CORDEX simulations for durations of 3–24 h and return periods of 1–30 years. The overestimation of hourly extremes is reported over both lowlands and mountains (Rybka et al., 2023,

their Figs. 3 and 4), and the dGEV is fitted to the simulation itself, so a time-invariant intensity bias is largely absorbed in the return periods. We will correct L102–103, which understates the short-duration bias.

Table R1. Relative bias and pattern correlation of return levels of the ERA5-driven COSMO-CLM simulation for Germany against RADKLIM and KOSTRA-DWD (from Rybka et al., 2023, their Figs. 1–4).

Duration, return period	Bias RADKLIM	Bias KOSTRA-DWD	Corr. RADKLIM	Corr. KOSTRA-DWD
24 h, 1 year	+7.7 %	+7.9 %	0.87	0.92
24 h, 10 years	+20.3 %	+12.2 %	0.68	0.77
1 h, 1 year	+33.2 %	+23.5 %	0.50	0.69
1 h, 10 years	+64.3 %	+27.1 %	0.18	0.46

Planned changes in the revised manuscript

- Data section: subsection on the evaluation of extreme precipitation with Table R1; correction of L102–103; maps from Rybka et al. (2023) in the Supplement.

Referee comment RC1.3

In Line 223, the authors indicate that each time-period is analyzed independently. However, Voit and Heistermann (2022) studied specific regions, and adapting the framework for a shifting regional analysis every time period must be justified more thoroughly. For instance, does the shift in the extremity of the events still hold if we consider grids fixed over all the time periods analyzed? Or is the changing atmospheric current shifting the locations of the storm at the subregional scale? It may be better if the authors provide a separate sensitivity analysis, by performing an analysis over a few fixed selected sub-regions across all the time periods, or this should be justified thoroughly. In addition, the authors should provide individual maps indicating the locations of the top 200 events for each of the four time-periods.

Voit and Heistermann (2022) did not use fixed regions: they selected the most extreme CatRaRE events across Germany and computed WEI and xWEI in a 200 × 200 km window around each event centroid. We follow this design; “analysed independently” (L223) only means that the top-200 are selected separately for each period, and we will rephrase this. We agree that a change in the regional composition could affect the results. We will therefore show the event locations for all periods (Figure R1) and repeat the analysis for three large, a priori defined sub-regions (e.g., northern lowlands, central uplands, southern Germany), which are large enough not to cut large-scale events.

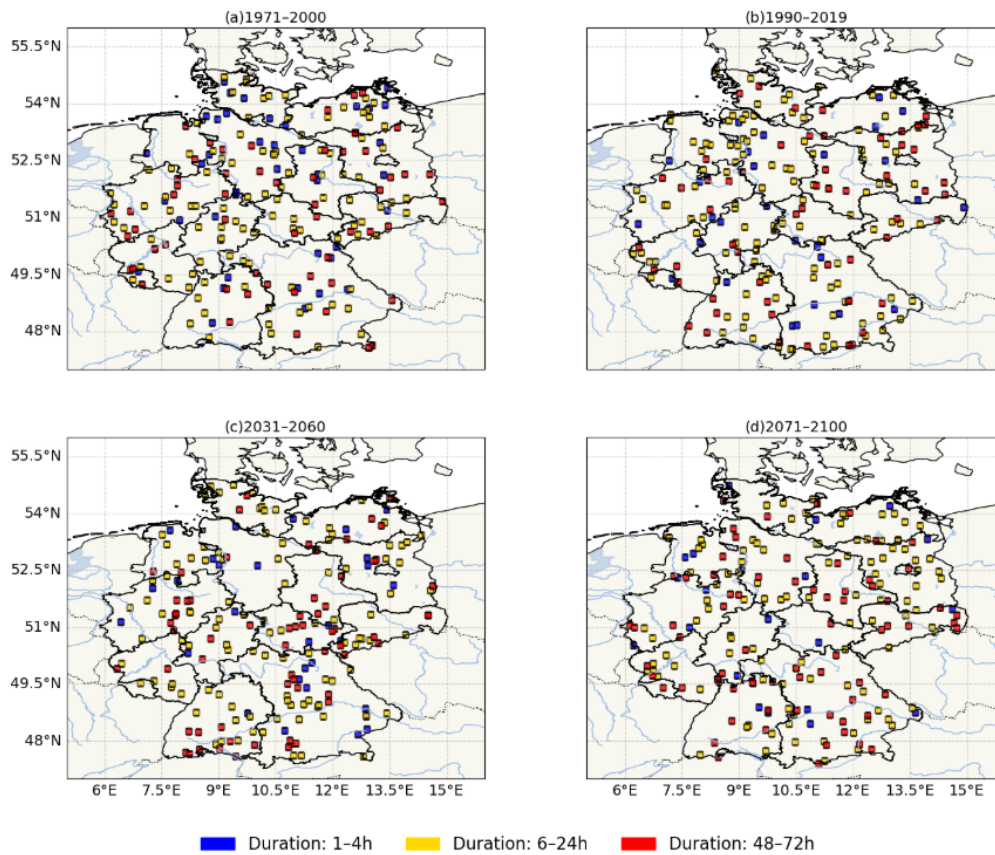


Figure R1 (preliminary). Centroids of the top-200 events in (a) 1971–2000, (b) 1990–2019, (c) 2031–2060 and (d) 2071–2100. Colours: duration of maximum extremeness (WEI), 1–4 h (blue), 6–24 h (yellow), 48–72 h (red). Preliminary: event numbers subject to revision following the corrected event-selection procedure of Reviewer 2 RC2.2.

Planned changes in the revised manuscript

- Rephrasing of L223; event maps (Figure R1) in the Supplement.
- Sub-region analysis (ranked xWEI, duration and area composition, seasonality; fixed and adapted dGEV).

Referee comment RC1.4

An indication that the composition of the top-ranked subset events shifts from smaller, short-lived events to a larger and more persistent events would suggest that the subset is becoming more dominated by synoptic-scale events. Such an analysis should be supported by a seasonal classification, which would clarify by how much the seasons dominated by synoptic scale events change over time, compared to the increase in convective scale events. Combining these scales into the same structural category for analysis provides an incomplete result. Because synoptic scale events are generally much larger than convective events in both duration and volume, an increase in both would skew the representation of the entire subset towards synoptic scale events when selecting only the top 200 most extreme events, due to the sheer mathematical weightage that synoptic scale events possess. The issue of extreme precipitation varying seasonally and regionally has been mentioned in previous studies (Zhong et al., 2025). Please justify.

We agree and will add a seasonal breakdown for both dGEV frames. Preliminary results (Figures R2–R4) show that summer events dominate the top-200 in all periods (fixed dGEV: 160, 146, 151 and 155 of 200), while winter events become rarer. The shift towards longer and larger events occurs within summer rather than through a change in seasonal composition (adapted dGEV: summer events with maximum extremeness at 1–4 h decrease from 39 to 19, those at 48–72 h increase from 35 to 52; Figure

R2). Note that these counts are preliminary and will be updated after the corrected event-selection procedure.

Regarding the weight of large systems: the xWEI weights the area by $\sqrt{A/\pi}$, integrates over $\ln(t)$ up to 72 h and is limited to a 200×200 km window (Voit and Heistermann, 2022), and the same metric is applied to all periods. We will test the sensitivity of the shift to a logarithmic area weighting, which emphasises small and intense events, and discuss the seasonal and regional dependence of changes in extreme precipitation extents (Zhong et al., 2025).

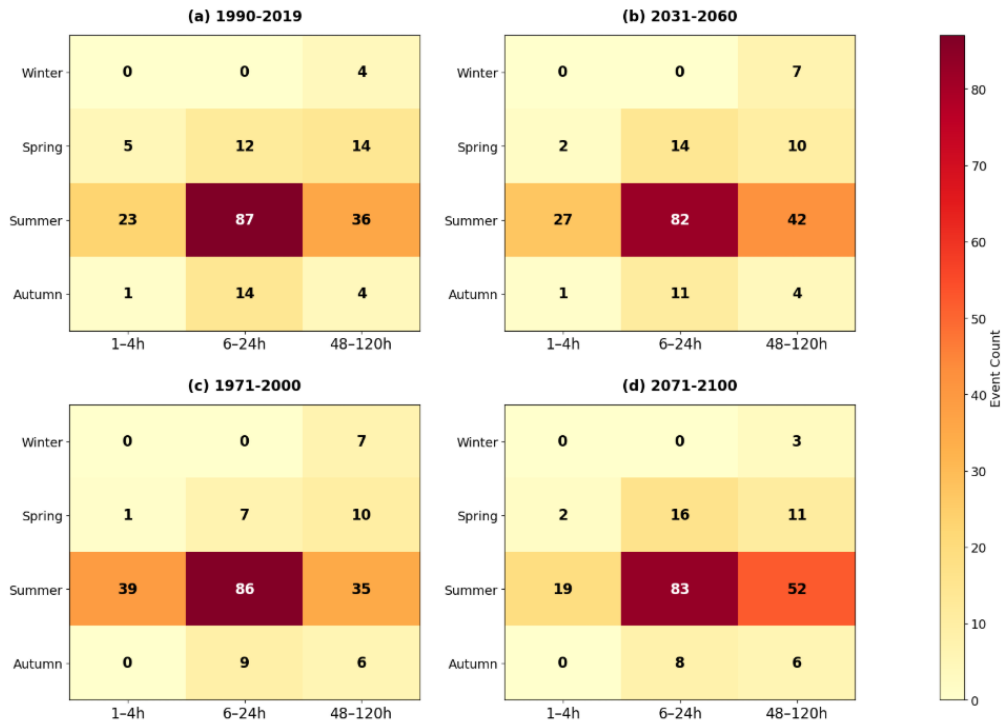


Figure R2 (preliminary). Number of top-200 events per season and duration of maximum extremeness, adapted dGEV: (a) 1990–2019, (b) 2031–2060, (c) 1971–2000, (d) 2071–2100.



Figure R3 (preliminary). As Figure R2, but per area class (km^2).

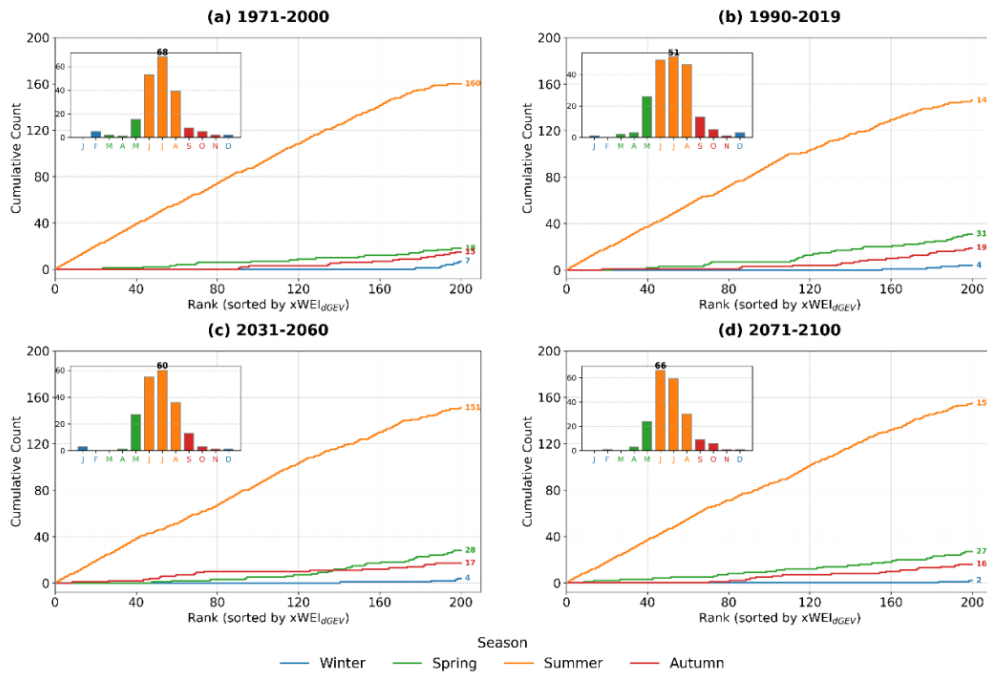


Figure R4 (preliminary). Cumulative number of top-200 events per season versus xWEI rank (fixed dGEV; own fit for 1990–2019). Insets: events per month.

Planned changes in the revised manuscript

- Results: seasonal composition of the top-200 events (fixed and adapted dGEV).
- Sensitivity of the composition shift to a logarithmic area weighting.
- Discussion: weighting of large systems; comparison with Zhong et al. (2025).

Referee comment RC1.5

In continuation with the previous point, the shift of the results from smaller and short-lived events to larger and persistent events is also a known bias for models called the ‘drizzle effect’ or a ‘precipitation persistence’. This effect should also manifest during extreme events. Specifically, modelled datasets often tend to stretch high high-intensity, short-duration events over a longer durations and larger spatial volume (Carlo et al., 2026; Moon et al., 2019). Unless prior validation for this specific COSMO configuration have demonstrated that this smoothing bias is not present, addressing this issue would be critical for the credibility of the results.

Thank you for adding this topic and the additional literature. We agree that this needs to be discussed. The classical drizzle problem (too frequent light rain, too persistent wet days) mainly affects coarse models with parameterised convection, e.g. CMIP5 global models (Moon et al., 2019), and is largely removed in convection-permitting models (Rybka et al., 2023). Carlo et al. (2026), however, find an excessive temporal persistence of heavy rainfall also in two convection-permitting hindcasts over Italy. For our simulations, Rybka et al. (2023) report good agreement for durations ≥ 12 h and an overestimation of short-duration extremes, i.e. the opposite of a smoothing bias; if anything, this favours short events in the top-200. Most importantly, a time-invariant persistence bias affects all periods alike and cannot by itself create the change signal. A state-dependent bias cannot be excluded, and we will state this as a limitation.

Planned changes in the revised manuscript

- Discussion: paragraph on model persistence biases (Moon et al., 2019; Carlo et al., 2026; Rybka et al., 2023); limitation regarding state-dependent biases.

Referee comment RC1.6

The use of RCP8.5 (or SSP5-8.5) may be unjustified. The scenario framework for CMIP7 has recently considered it implausible, and intends to have a high emissions scenario that is expected to result in forcings lower than SSP5-8.5 (Van Vuuren et al., 2026). Given that the novelty of this study is central CPM dataset use for climate projections, it would be beneficial instead to understanding these analyses for more realistic climate projections, such as SSP2-4.5 or SSP3-7.0, rather than SSP5-8.5 that may now be considered obsolete.

We agree that RCP8.5/SSP5-8.5 emissions are no longer considered plausible; the CMIP7 High scenario is expected to have a lower forcing (Van Vuuren et al., 2026). However, the DWD simulations are, to our knowledge, the only multi-decadal convection-permitting projections covering Germany, and they exist only for RCP8.5. To make the results transferable, we will report global warming levels: for the driving MIROC5 run (r1i1p1), about 1.9 °C in 2031–2060 and 3.6 °C in 2071–2100 relative to 1850–1900 (Hauser et al., 2019). Owing to the low climate sensitivity of MIROC5, the far-future period corresponds roughly to the upper end of the CMIP7 scenarios in 2100 (almost 3.5 °C; Van Vuuren et al., 2026). We will frame the conclusions as conditional on scenario and model chain and name the lack of multi-scenario projections as a limitation. We will frame all conclusions as conditional on this single model chain and scenario, and explicitly name the lack of multi-scenario and multi-model projections as a key limitation.

Planned changes in the revised manuscript

- Data section: global warming level of each period; Discussion and Conclusions: warming-level framing; limitation.

Referee comment RC1.7

The authors have noted that the MIROC5 model is considered a median performing model, yet it is among the ‘driest’ in the CMIP5 ensemble and show a notable warm and dry summer bias. It is unclear how this, in combination with the previous point about the extreme SSP5-8.5 scenario, would compound in the case of extreme precipitation events. It would be better if the study used driving models that are median performance specifically in terms of their representation of convective precipitation. The results for an analysis of a dry-driven model subject to an implausible extreme SSP5-8.5 scenario may not give satisfactory results. Moreover, if the model shows a dry bias in summer - a season known for convective scale events - then this would further skew the composition of the results of extreme events towards synoptic scale events as highlighted in Point 4 previously.

We tried to openly discuss the potential limitations of the driving GCM in the preprint. However, we think that the simulations are valuable and suited for our analysis. First, the extremes are generated by COSMO-CLM within the domain: Rybka et al. (2023) find that simulated extremes over Germany depend more on the regional model than on the driving GCM, and the MIROC5-driven convection-permitting simulation outperforms nine CMIP5 EURO-CORDEX simulations driven by various GCMs. Second, we find no under-representation of summer events: they make up 160 of the top-200 events in the MIROC5-driven historical period and 146 in the ERA5-driven present-day period (Figure R4). Third, the projected changes in return levels of this simulation lie within the CMIP5 EURO-CORDEX range (Rybka et al., 2023). The comparison with the ERA5-driven simulation for 1979–2000 (RC3.2) will quantify the effect of the driving data. Since no other convection-permitting projections exist for the domain, a different driving model cannot be chosen; we will state this as a limitation.

Planned changes in the revised manuscript

- Data section: MIROC5 biases and role of the regional model.
- Results: comparison with the ERA5-driven simulation (RC3.2); Limitations: single driving GCM.

Referee comment RC1.8

The ‘second-order effect’ is not clearly understood from this study, and we feel the authors should elaborate on this more. Does this mean the intensification for short and medium duration of rainfall as mentioned in Line 288? Because this is in contrast with the conclusions that indicate a shift towards medium and long duration events.

We agree that the term is well enough described. It does not denote a stronger intensification of short-duration rainfall and does not contradict the shift towards longer events. The duration classes (Figure 5, Conclusions) refer to the duration at which each event reaches its maximum extremeness (WEI). L288 refers to $E_{T,A}$ at all eight durations of the same events, integrated over area (Figure 7): under the adapted dGEV, the WEI durations and areas hardly change, but $E_{T,A}$ at short and medium durations increases. In terms of Figure 3a, the peak of the $E_{T,A}$ curves remains similar while their shoulders rise, which increases the xWEI. We will rename the effect (e.g., “cross-scale broadening of extremeness”), define it in the Methods and quantify it (e.g., by the ratio xWEI/WEI). As the effect appears as a single step after 1971–2000 (also within the MIROC5-driven chain), we will test whether it is related to the dGEV fits or to sampling variability before interpreting it physically (RC1.9, RC3.2).

Planned changes in the revised manuscript

- Methods: definition and metric; rewritten Results, Discussion and Abstract.
- Robustness of the step change (dGEV parameters per period, bootstrap, comparison with the ERA5-driven run for 1979–2000).

Referee comment RC1.9

In estimating the dGEV and adapted dGEV parameters in Section 3.1.2, the authors should also elaborate on any uncertainties when estimating the parameters or sensitivity to outlier events, and whether there was any evaluation on how this could affect the return period.

We agree. Pooling all durations stabilises the dGEV fits (Ulrich et al., 2020; Fauer et al., 2021), but the shape parameter remains uncertain for 30 annual maxima. Lengfeld et al. (2023) showed that excluding the event year from the fit increases WEI and xWEI of the most extreme radar-observed events by 17–28 %. This matters for the fixed dGEV, for which historical events are in-sample and future events out-of-sample. We will (i) show maps and distributions of the dGEV parameters for all periods (Figure R5: preliminary comparison of the shape parameter with RADKLIM), (ii) quantify the in-sample effect with leave-one-year-out fits for the top 10 historical events, (iii) report the share of grid cells at the return period cap of 1000 years (Müller and Kaspar, 2014) and test the sensitivity to the cap, (iv) test a fixed shape parameter (e.g., $\xi = 0.1$), and (v) provide bootstrap confidence intervals for the top-200 statistics. We will report the according results in the supplement if it does not change the main conclusions.

Planned changes in the revised manuscript

- Methods: return period cap and fit uncertainty (Lengfeld et al., 2023); Supplement: dGEV parameter maps, share of capped cells.
- Leave-one-year-out fits (top 10 historical events), fixed-shape test, bootstrap confidence intervals.

Technical corrections

Referee comment RC1.T1

Line 161: Maybe it would be better to rephrase the concluding line as follows instead: “Hence, we refrain from mixing dGEV parameters from COSMO-CLM driven by MIROC5 (1971–2000) with the simulation driven by ERA5.”

We will rephrase the sentence as suggested.

References cited in this response

- Carlo, O. S., Fedele, G., Stocchi, P., Calmanti, S., Raffa, M., Mercogliano, P., and Lionello, P.: A nested time-window framework for evaluating sub-daily precipitation extremes in two convection permitting model datasets, EGU sphere [preprint], <https://doi.org/10.5194/egusphere-2026-4515>, 2026.
- Fauer, F. S., Ulrich, J., Jurado, O. E., and Rust, H. W.: Flexible and consistent quantile estimation for intensity–duration–frequency curves, *Hydrol. Earth Syst. Sci.*, 25, 6479–6494, <https://doi.org/10.5194/hess-25-6479-2021>, 2021.
- Hauser, M., Engelbrecht, F., and Fischer, E.: Global warming levels for CMIP5 and CMIP6 (cmip_warming_levels), Zenodo [data set], <https://doi.org/10.5281/zenodo.3591806>, 2019.
- Lengfeld, K., Walawender, E., Winterrath, T., and Becker, A.: CatRaRE: A Catalogue of radar-based heavy rainfall events in Germany derived from 20 years of data, *Meteorol. Z.*, 30, 469–487, <https://doi.org/10.1127/metz/2021/1088>, 2021.
- Lengfeld, K., Voit, P., Kaspar, F., and Heistermann, M.: Brief communication: On the extremeness of the July 2021 precipitation event in western Germany, *Nat. Hazards Earth Syst. Sci.*, 23, 1227–1232, <https://doi.org/10.5194/nhess-23-1227-2023>, 2023.
- Moon, H., Gudmundsson, L., Guillod, B. P., Venugopal, V., and Seneviratne, S. I.: Intercomparison of daily precipitation persistence in multiple global observations and climate models, *Environ. Res. Lett.*, 14, <https://doi.org/10.1088/1748-9326/ab4169>, 2019.
- Müller, M. and Kaspar, M.: Event-adjusted evaluation of weather and climate extremes, *Nat. Hazards Earth Syst. Sci.*, 14, 473–483, <https://doi.org/10.5194/nhess-14-473-2014>, 2014.
- Rybka, H., Haller, M., Brienens, 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.
- Ulrich, J., Jurado, O. E., Peter, M., Scheibel, M., and Rust, H. W.: Estimating IDF curves consistently over durations with spatial covariates, *Water*, 12, 3119, <https://doi.org/10.3390/w12113119>, 2020.
- Van Vuuren, D. P., O'Neill, B. C., Tebaldi, C., et al.: The Scenario Model Intercomparison Project for CMIP7 (ScenarioMIP-CMIP7), *Geosci. Model Dev.*, 19, 2627–2656, <https://doi.org/10.5194/gmd-19-2627-2026>, 2026.
- Voit, P. and Heistermann, M.: A new index to quantify the extremeness of precipitation across scales, *Nat. Hazards Earth Syst. Sci.*, 22, 2791–2805, <https://doi.org/10.5194/nhess-22-2791-2022>, 2022.
- Zhong, P., Brunner, M., Opitz, T., and Huser, R.: Spatial modeling and future projection of extreme precipitation extents, *J. Am. Stat. Assoc.*, 120, 80–95, <https://doi.org/10.1080/01621459.2024.2408045>, 2025.