

Review of Rosin et al. (2026) – Radar-Derived Intensity–Duration–Area–Frequency Relations for Assessing Hydrological Hazards in Complex Terrain

Review 1 response

General assessment

The manuscript presents an application of the areal Simplified Metastatistical Extreme Value framework to derive radar-based Intensity–Duration–Area–Frequency relationships for summer precipitation extremes over Switzerland. The study uses 9 years of 5 min, 1 km² radar QPE data from the MeteoSwiss radar network, evaluates pixel-scale return levels against long-term rain-gauge estimates, and illustrates the hydrological relevance of the resulting IDAF fields through three recent high-impact flood-producing events. The topic is relevant for HESS because it addresses scale-dependent precipitation extremes, radar-based design rainfall estimation, and hydrological hazard assessment in complex Alpine terrain.

The manuscript has several strengths. It addresses an important problem: the estimation of spatially and temporally distributed rainfall extremes in mountainous terrain where gauge networks are sparse and not fully representative. The use of radar data for IDAF analysis is promising, and the event-based severity diagrams are potentially useful for linking rainfall scale dependence to hydrological impacts. The manuscript is also generally well connected to recent literature on SMEV, radar-based frequency analysis, areal reduction factors, and Swiss precipitation extremes.

However, in its present form, I think the manuscript requires major revision before it can be considered for publication. My main concerns are not with the overall topic, which is relevant and potentially valuable, but with the clarity of presentation, the organization of the results and discussion, and the level of methodological justification provided for several key choices. In particular, the manuscript should better justify the use of a 9-year radar record for estimating return periods up to 100 years, the choice of only 100 bootstrap repetitions, the spatial smoothing of the Weibull shape parameter, and the interpretation of radar–gauge discrepancies. Several figures are introduced without sufficient interpretation in the text, and some sections contain discussion-type material within the results section, while the discussion itself does not sufficiently refer back to the manuscript's own figures and results.

I therefore recommend major revisions. The manuscript could become a strong contribution after improving the structure, tightening the interpretation, and adding clearer methodological support for the main assumptions.

Response: We thank the reviewer for the careful and constructive assessment. We have revised the manuscript to distinguish the application contribution from the previously published IDAF methodology of Rosin et al. (2024) and to clarify the assumptions and limits of extrapolation from the nine-year radar record. We removed the 100-year return-level estimates, limiting the analysis to return periods of up to 50 years. We strengthened the treatment of event definition, censoring, spatial regularisation and bootstrap uncertainty, adding sensitivity analyses of the dry-separation

criterion, the smoothing window and the bootstrap ensemble size, which we increased from 100 to 800 replicates. We also defined the scope of the radar–gauge comparison and reorganised the Results and Discussion. Finally, we expanded the figure descriptions, corrected Figures 1, 6 and 10, and reframed the three event studies as illustrations of scale-dependent rainfall severity rather than as evidence for general hydrological response mechanisms.

Major comments

1) The manuscript applies an existing methodology, namely the areal SMEV framework, to Switzerland. This is a valid and potentially useful contribution, especially because the Swiss Alpine domain is topographically complex and hydrologically relevant. However, the manuscript should more explicitly distinguish between methodological novelty and regional/application novelty. At present, the introduction states that the study “further advances this framework” by applying it to Alpine topography and linking duration–area scaling to observed impacts. This statement would be stronger if the authors clarified exactly what is new relative to Rosin et al. (2024) and other radar-based IDAF or ARF studies. For example, is the main novelty the Swiss radar dataset, the Alpine complexity, the validation against long-term gauges, the event-severity application, or the interpretation of topographic controls? This distinction is important because the statistical framework itself appears to be largely adopted from previous work. I recommend adding a concise paragraph at the end of the introduction explicitly listing the paper’s main contributions.

Response: Thank you for this comment. We agree that the previous wording did not clearly distinguish this study from the earlier methodological work of Rosin et al. (2024). We have revised the last few passages in the Introduction. The first identifies the IDAF methodology of Rosin et al. (2024) as an existing framework and explains its application to Switzerland; the second states the specific contributions of this study.

The first passage now reads:

Here, we apply the SMEV-based IDAF methodology of Rosin et al. (2024) to the complex and heterogeneous Alpine topography of Switzerland. The contribution of this study lies in this application rather than in the statistical method itself: we use the resulting multiscale return levels to examine how topography influences the duration–area scaling of precipitation extremes, how the relevant rainfall scale varies among events with different spatial footprints, and how this relates to recent high-impact hydrological events.

The contribution paragraph now reads:

The specific contributions of this study are to: (i) evaluate radar-derived return levels against independent long-term rain-gauge records, thereby assessing the reliability of estimates obtained from the relatively short radar archive; (ii) produce spatially distributed IDAF estimates across the contrasting topographic regions of Switzerland and examine how the spatial patterns and duration–area scaling of extreme rainfall vary among regional topographic and precipitation regimes; and (iii) illustrate the application of these multiscale estimates through the analysis of three recent high-impact flood events, including the July 2021 regional floods and the June 2024 event. By linking observed event rainfall to duration- and area-dependent return levels, the event analyses demonstrate how the IDAF framework

diagnoses changes in rainfall rarity and hydrological relevance across spatiotemporal scales.

2) A central assumption of the manuscript is that 9 years of summer-only radar data are sufficient to estimate return levels up to 100 years using SMEV. The authors cite previous SMEV work to support this, but the issue deserves a more explicit and critical treatment in this manuscript. The concern is especially important because the record includes only June–August precipitation, i.e. about 27 summer months in total. Even if SMEV uses ordinary events rather than annual maxima, estimation of 50- and 100-year return levels from such a record still depends strongly on the validity of the assumed Weibull tail, event independence, stationarity, radar-QPE stability, and the regionalization/smoothing strategy. The manuscript should therefore better quantify and discuss the extrapolation uncertainty associated with 50- and 100-year estimates. The current presentation risks giving the impression that 100-year estimates from a 9-year radar archive are as reliable as those obtained from much longer records. The manuscript should be more cautious and explicit about this limitation.

It is therefore suggested to consider the following points:

- how many independent ordinary events are available per pixel or region;
- how this number varies spatially and with event definition;
- whether the estimated number of storms per year is stable across the 9-year period;
- how sensitive 50- and 100-year return levels are to the 90 % censoring threshold, the 15×15 km² shape-smoothing window, and the hail-removal procedure;
- whether the 100-year maps should be interpreted as robust design estimates or as exploratory radar-based estimates with substantial uncertainty.

Response: Thank you. We agree that return levels substantially longer than the nine-year radar record must be interpreted cautiously. To reduce the degree of extrapolation, we removed the 100-year return-level estimates throughout the manuscript and now limit the analysis to return periods up to 50 years. The remaining 50-year estimates are presented as radar-based estimates that are conditional on the SMEV assumptions and on the quality of the radar QPE, rather than as stand-alone design values, and the revised Conclusions state this explicitly.

We also clarify that the SMEV framework uses substantially more event information than annual-maxima approaches, which can improve the stability of return-level estimates from short records, conditional on the validity of the SMEV assumptions and the quality of the precipitation data. At this concern, it is important to point out that the used assumption of Weibull tails for the analysed area is supported by several previous studies (e.g., Peleg et al., 2025a, b; Akbary et al., 2026; Marra et al., 2026), which used longer observational records to evaluate such an assumption. It follows that the assumption should not be considered as based on nine years of data; rather, nine years of data are just used for estimating the parameters of an already validated tail model.

To better clarify this in the paper, we extended the details of Sect. 3.5, *SMEV framework*, that includes now the following paragraphs:

The Weibull distribution is intended to represent the upper tail rather than the full ordinary-event distribution. Parameter estimation therefore uses left censoring, with observations below the 90th percentile treated as censored. The censoring level was selected using the longer rain-gauge records and the statistical testing procedure described and implemented

by Marra et al. (2023). This procedure identifies the minimum threshold above which the null hypothesis of Weibull-tail behaviour is not rejected. Although lower thresholds were acceptable at some gauges, 90% was the lowest common censoring level that provided an acceptable fit across all assessed durations and climatic regions. A common threshold was applied across regions, durations, and aggregation areas to facilitate comparison of the resulting uncertainties.

The gauge-derived threshold is applied to the radar data because gauge and radar events originate from the same regional precipitation regime and are expected to have upper tails governed by broadly similar processes. This does not imply that their intensity distributions or Weibull parameters are identical, given their different measurement supports and the effects of spatial averaging.

Previous studies provide regional support for this choice. Dallan et al. (2024) identified suitable censoring levels of 90% for 1 h precipitation and 85% for longer durations in the Alpine-Mediterranean region, while Akbary et al. (2026) adopted a common 85% level for sub-daily areal precipitation over Switzerland after testing the Weibull-tail representation. Together with other analyses based on longer observational records (Peleg et al., 2025a, b; Marra et al., 2026), these studies indicate that the Weibull-tail assumption does not rely solely on the nine-year radar record. The radar record is used to estimate the model parameters, with SMEV drawing on multiple ordinary events per year rather than annual maxima alone. The reliability of the resulting return-level estimates nevertheless remains conditional on model adequacy and precipitation-data quality.

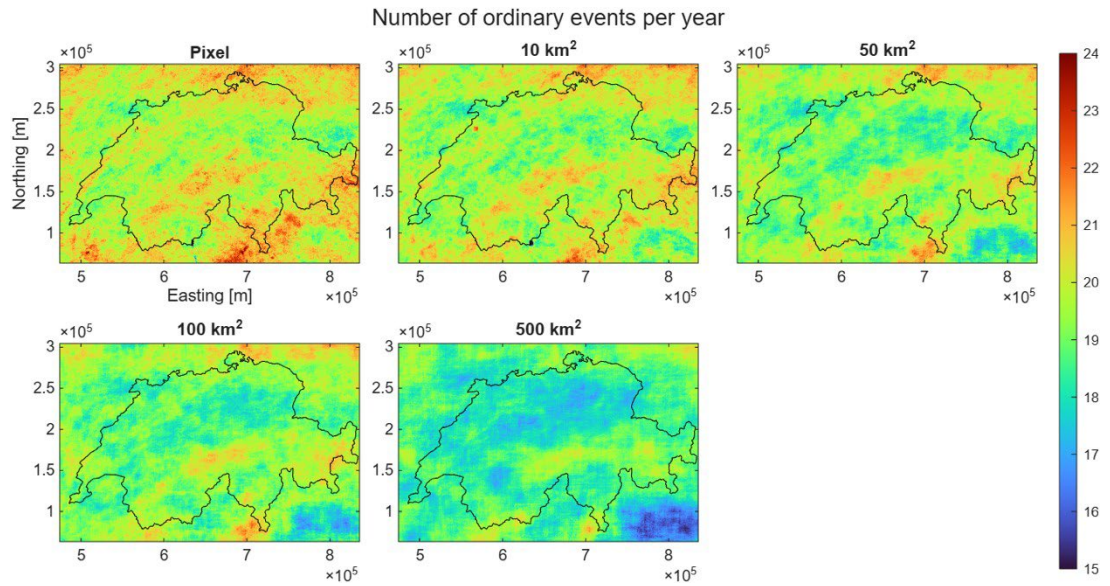
We have also added three complementary analyses to document the information supporting the estimates.

First, we have added Figure S7 (shown below) showing the spatial distribution of the mean annual number of ordinary rainfall events and the total number of upper-tail events used to estimate the SMEV parameters for different areal scales.

The following passage has been added to Sect. 4.2, *Spatial maps of return levels*:

Before examining the derived return levels, we first assess the number of ordinary events available for SMEV parameter estimation (Figure S7). The mean annual number of ordinary events, n , ranges from 17 to 22 per grid cell across all pixels and aggregation areas (Figure S7a), corresponding to approximately 150–200 ordinary events per grid cell over 2016–2024. After 90% left censoring, 15–20 upper-tail events per grid cell remain uncensored over the nine-year record (Figure S7b). The remaining ordinary events are retained in the likelihood as censored non-exceedances, so the information used for SMEV parameter estimation exceeds the number of uncensored events. The domain-mean annual number of ordinary events also remained relatively stable among the nine years, indicating that event frequency did not vary strongly over the study period.

(a)



(b)

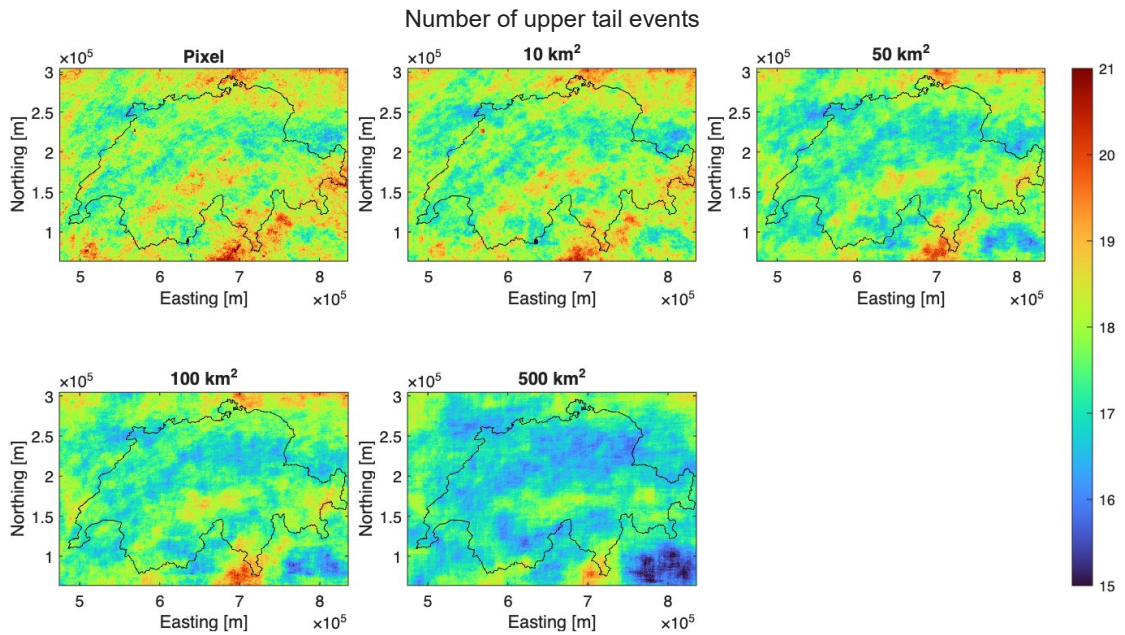


Figure S7. Spatial distributions of (a) the mean annual number of ordinary rainfall events and (b) the number of upper-tail events, defined as ordinary rainfall events retained after 90% left censoring and used for SMEV parameter estimation, for aggregation areas of 1 km² (pixel), 10 km², 50 km², 100 km², and 500 km²

Second, we evaluated the sensitivity to the spatial smoothing window size, testing windows from 5 to 50 km for both 20- and 50-year return levels. See more details in response to comment #5 below that focuses on this point.

Third, we have increased the block-bootstrap ensemble from 100 to 800 repetitions and added a convergence analysis in Supplement Sect. S2 and Figures S5 and S6, showing the convergence of the median 20-year return level and of the 5th–95th percentile interval width with increasing

ensemble size (see the detailed updates in text and figures in the response to major comment #6). The revised manuscript also states explicitly that the bootstrap intervals quantify conditional sampling uncertainty associated with observed interannual variation in event frequency and magnitude (see the detailed text in the response to major comment #4). These intervals do not account for the full uncertainty associated with radar-QPE errors, event definition, Weibull-tail adequacy, stationarity, hail filtering, or spatial regularisation.

It should also be noted that Dallan et al. (2024) showed in an idealised experiment that, for a 10-year record, this block-bootstrap approach captured most (approximately 86%) of the true 90% confidence-interval width. We now note this in Sect. 4.3.

Regarding the censoring threshold, all gauges and durations passed the Weibull-tail test at 90% or at a lower threshold, so 90% is applicable to all of them (see the response to comment #14). Once the threshold lies within the range where the Weibull-tail hypothesis is not rejected, higher thresholds are expected to give statistically equivalent return levels, apart from larger random estimation uncertainty from the smaller retained sample. Regarding the hail-removal procedure, we did not repeat the analysis without hail filtering. The filter removes only timesteps with a high (>90%) probability of hail at the ground and acts as a quality-control step on the radar QPE. The bootstrap intervals are conditional on the radar-QPE product, including this quality control, as now stated in Sect. 4.3.

3) Another ambiguity drives from if the manuscript compares radar-derived SMEV return levels with gauge-derived SMEV return levels, not with GEV-derived return levels. This should be stated consistently and clearly. the comparison is also limited to the 1 km^2 radar-pixel scale, while the main novelty of the manuscript concerns areal aggregation up to 500 km^2 . This limitation is understandable because the gauge network cannot validate areal return levels directly, but it should be emphasized more strongly. The validation supports the radar product at the pixel scale only; it does not fully validate the areal IDAF estimates.

Response: Thank you for highlighting this ambiguity. We have now clarified throughout the manuscript that both radar- and gauge-derived return levels are estimated using the SMEV framework; GEV estimates are not used in the validation or elsewhere in the analysis. The nine-year radar record would provide only nine summer annual maxima per location for a GEV fit, making estimates at the longer return periods considered here highly uncertain. The text in Sect. 4.1 states that return levels for both radar and gauges are computed with SMEV:

To assess the reliability of the radar-based SMEV estimates, we therefore compare them against independent return level estimates calculated from gauge data. SMEV-based return levels are computed for all 60 gauges using their full available summer records, and the corresponding radar-derived return levels are extracted from the 1 km^2 radar pixel containing each gauge. Unlike gauge-blended products such as CombiPrecip, the radar QPE was not adjusted to match the gauge return levels used in this comparison. However, its uniform summer mean-field adjustment was derived from radar–gauge comparisons (Sect. 3.3), so the evaluation is not wholly independent of gauge information.

We also modified caption of Figure 2 to further clarify this point:

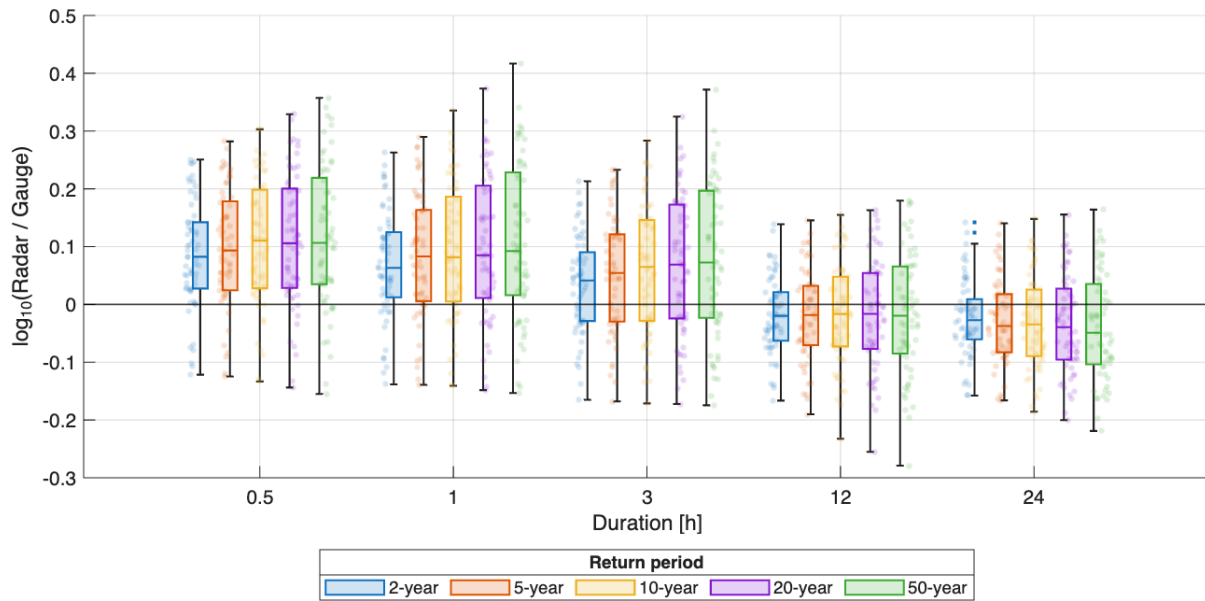


Figure 2. Bias between the radar- (1 km² pixel scale) and gauge-derived return levels for return periods of 2, 5, 10, 20 and 50 years and durations of 30 min, 1, 3, 12 and 24 h. Return levels for radar and gauges are derived with the SMEV framework.

We have also revised the terminology to emphasise that this comparison is restricted to the 1 km² radar-pixel scale. Consequently, it supports the reliability of the radar-based return levels at the pixel scale but does not constitute a direct validation of the areal IDAF estimates for aggregation areas from 10 up to 500 km². Direct validation at these larger spatial scales is not possible using the available point-gauge network, and we now acknowledge this limitation more explicitly. We state in Sect. 4.1, *Validation against rain-gauge-derived return levels*:

As the gauge network is not sufficiently dense to support a robust evaluation at larger areal aggregations, the validation is restricted to the pixel scale and does not directly validate the areal IDAF estimates derived for aggregation areas of 10–500 km².

In the Conclusions we now state that the comparison with gauges is limited to the pixel scale:

Comparison of the pixel-scale return levels with 60 long-term gauges shows broad agreement, with duration- and region-dependent differences. The areal estimates (10–500 km²) are supported methodologically but cannot be validated directly with the available gauge network.

We also added the following limitation regarding sub-kilometre rainfall peaks:

This highlights the necessity of evaluating extremes across multiple dimensions, rather than relying solely on point-scale data. However, we also note that even the smallest radar estimate represents an average over a 1 km² pixel and thus cannot characterise sub-kilometre rainfall peaks within an urban catchment. The Lausanne event illustrates this limitation, as the highly localised peak recorded by the city-centre gauge was substantially smoothed within the corresponding radar pixel.

In addition, Rosin et al. (2024) compared 24 h, 100 km² radar-derived return levels with estimates from spatially averaged rain-gauge observations in six sufficiently gauged areas of Israel and found generally good agreement. We now refer to this in Sect. 4.1:

Areal comparisons are possible where gauges are sufficiently dense, as demonstrated for 100 km², 24 h return levels by Rosin et al. (2024). This provides methodological support for areal-scale radar–gauge evaluation, although differences in precipitation regime and dataset mean that their results do not constitute a direct validation of the 10–500 km² estimates presented here.

4) Third, the discussion of radar–gauge disagreement should separate different sources of discrepancy more explicitly: radar-QPE errors, gauge measurement errors, point-area representativeness differences, temporal aggregation effects, and sampling uncertainty. These processes have different implications. For example, a disagreement caused by sub-pixel variability does not necessarily imply radar bias, whereas disagreement caused by Z–R uncertainty or hail contamination does.

I suggest adding a short conceptual paragraph explaining what the validation can and cannot establish.

Response: Thank you for this suggestion. We have clarified what the bootstrap intervals measure and what the radar–gauge comparison can establish and have revised the validation discussion to distinguish radar-QPE errors, gauge measurement errors, point-to-pixel representativeness, temporal effects, and sampling uncertainty. Section 4.3 now states:

Uncertainty is shown as 90% confidence intervals derived from 800 bootstrap repetitions, computed by first calculating the spatial mean return level within each box for every bootstrap realisation and then taking the corresponding percentile range. These intervals quantify sampling variability conditional on the selected event definition, censoring threshold, Weibull-tail model, spatial-smoothing procedure, and radar-QPE product. They do not quantify structural uncertainty associated with possible departures from the independence or stationarity assumptions, variation in precipitation-generating processes, systematic radar-QPE errors, or the choice of smoothing window. In an idealised experiment, Dallan et al. (2024) showed that, for a 10-year record, this block-bootstrap approach captured most (approximately 86%) of the true 90% confidence-interval width.

We have also added a short conceptual paragraph to Sect. 4.1 that separates the sources of disagreement and states what the comparison can and cannot establish:

The radar–gauge differences in Figures 2 and 3 combine several sources: (i) residual radar-QPE errors, including beam shielding and beam height, vertical variability of reflectivity, uncertainty in the Z–R relationship, residual hail contamination, and the spatially uniform 0.87 adjustment; (ii) gauge measurement errors, notably wind-induced undercatch, which is largest at short durations and at exposed sites; (iii) the point-to-pixel representativeness gap, which is largest at short durations and long return periods; and (iv) sampling uncertainty in both SMEV fits, together with the different record periods of the gauges (full records) and the radar (2016–2024). Only (i) constitutes radar bias; (ii)–(iv) produce disagreement even with an error-free radar. Recomputing the gauge return levels for 2016–2024 alone preserved the bias pattern, indicating that the difference in record period is not the main driver.

The related text in Sect. 4.1 now reads:

Radar-QPE errors can arise from beam geometry, uncertainty in the reflectivity–rain-rate relation, or hail contamination. Additionally, though gauge data are often used as ground truth, they also contain errors and biases, most significantly wind-induced undercatch (typically 5–10% for ordinary events but exceeding 30% during extremes) which disproportionately affects short durations (Pollock et al., 2018; Sieck et al., 2007). Radar may capture small sharp peaks that the gauges miss when the gauge is offset from the highest intensity rainfall within the 1 km² pixel.

....

It should also be noted that exact agreement is not expected between the gauge and radar estimates, even with perfect measurements, as gauges capture point-scale measurements, whereas radar QPE represents a spatially averaged value over a 1 km² pixel and therefore contains small-scale variability (Gires et al., 2014). This 'representativeness gap' is most pronounced at very short accumulation times (under 5 min according to Villarini et al., 2008) and longer return periods (Peleg et al., 2018). The comparison therefore shows the magnitude and spatial and duration dependence of disagreement, but cannot by itself isolate radar-QPE bias or establish unbiased return levels at individual sites.

5) The authors smooth the SMEV Weibull shape parameter using a 15 × 15 km² moving window and then re-estimate the scale parameter. This is an important methodological choice because the shape parameter controls tail behaviour and therefore strongly affects high return periods.

The manuscript states that the window size was selected by sensitivity analysis, but the main text should provide more information on this sensitivity analysis. For example:

- What window sizes were tested?
- What metric was used to choose 15 × 15 km²?
- Was the choice evaluated against gauge-derived return levels?
- Does the optimal window size vary between the Plateau, Jura, Southern Alps, and inner-Alpine valleys?
- Could smoothing across sharp topographic gradients mix physically distinct precipitation regimes?
- How sensitive are 50- and 100-year return levels to this smoothing choice?

Because this smoothing step regularizes the results and narrows the uncertainty intervals, it should be described and justified more rigorously.

Response: Thank you for this important comment. We agree that the smoothing scale requires explicit evaluation, especially in complex terrain, where excessive spatial averaging could suppress physically meaningful local gradients.

We clarify that this is a common approach to reduce spatial noise in extreme value analyses, and several previous studies have similarly smoothed the rainfall-distribution parameters: Deidda et al. (2021) applied a 9 x 9 km moving average to interpolated GEV parameter fields, while Srivastava and

Mascaro (2023) smoothed radar-derived GEV parameter estimates using a 40-km-radius moving-average window. Lompi et al. (2025) also applied a 3 x 3 spatial filter to results from grid-cell SMEV analyses. Other studies apply post-estimation smoothing to the rainfall return levels. Brajkovic et al. (2025) applied a 3x3 averaging kernel to climate-model return-level maps, Kašpar et al. (2026) applied a weak Gaussian filter to radar-derived design-precipitation fields.

We extended the relevant text in Sect. 3.5 and added a Supplement Figure S2 showing the window size analysis:

The scale and shape parameters are initially estimated independently at each pixel for all duration–area combinations using least-squares regression in Weibull-transformed coordinates (Marani and Ignaccolo, 2015). To reduce sampling variability associated with the nine-year record, the shape parameter is then smoothed using a 15 km × 15 km moving window. The window size was selected through a sensitivity analysis designed to balance uncertainty reduction against preservation of spatial rainfall variability. Square windows with side lengths of 5–50 km were evaluated for 20- and 50-year return levels using paired bootstrap samples across all durations and the six regional analysis boxes (Figure S2). Increasing the window size generally narrowed the 90% bootstrap intervals but increased the difference from the unsmoothed estimates. The selected 15 km window reduced the interval widths while changing the regional median return levels only modestly in most regions, with larger changes in regions with strong precipitation gradients (Figure S2). It was therefore retained as a compromise across regions and return periods.

After smoothing, the scale parameter is re-estimated at each pixel using the original ordinary-event sample, conditional on the smoothed shape parameter. Smoothing the shape parameter rather than the observations avoids modifying the event distribution before fitting and potentially distorting high-quantile return levels. As shown in Figure S2, the procedure reduces estimation uncertainty while only modestly affecting the estimated return levels. Comparable spatial smoothing has previously been applied to GEV and SMEV parameter fields and return-level maps (Deidda et al., 2021; Srivastava and Mascaro, 2023; Lompi et al., 2025; Brajkovic et al., 2025; Kašpar et al., 2026).

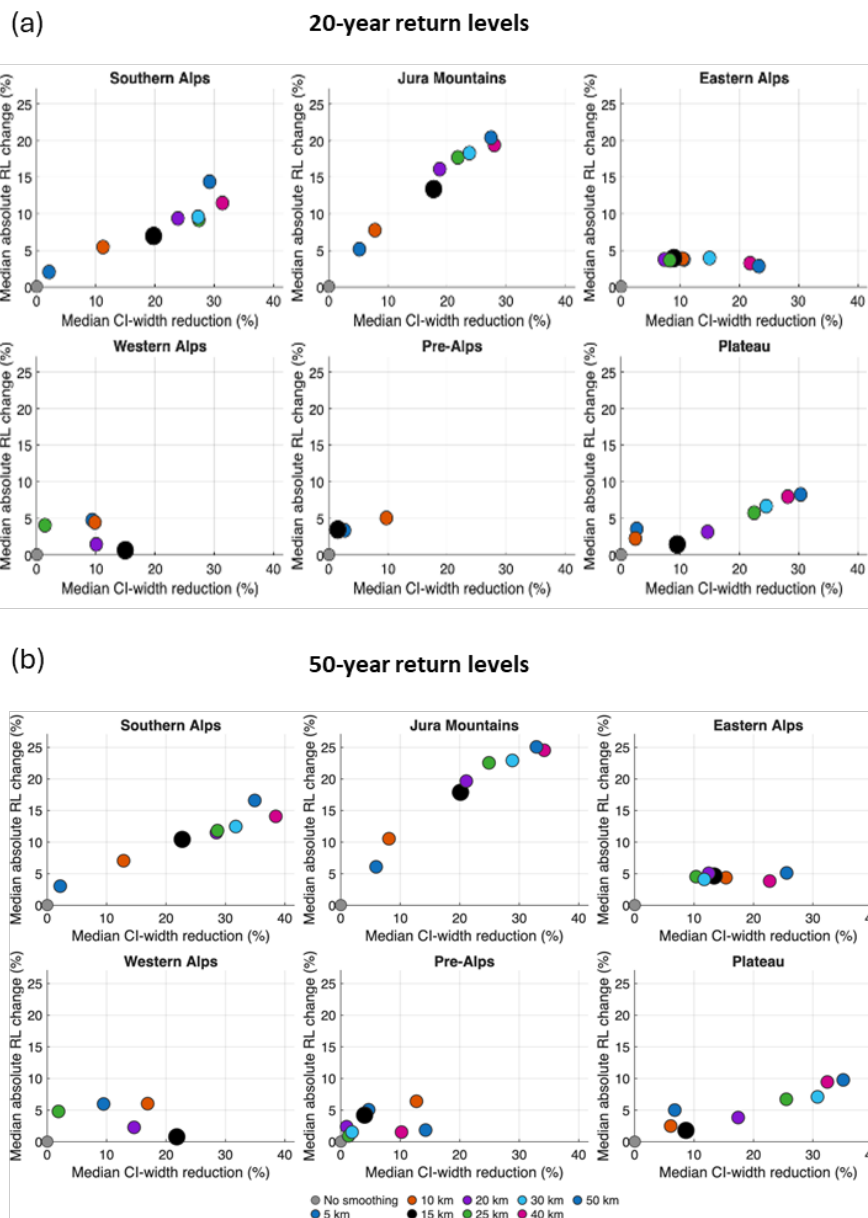


Figure S2. Median reduction in the relative 5–95% bootstrap confidence-interval width versus the median absolute change in (a) 20-year and (b) 50-year return levels for smoothing windows of 5–50 km, relative to the unsmoothed analysis, across the six study regions. Colours indicate smoothing-window size.

The trade-off differs among regions (Figure S2). Return levels change only modestly in most regions, whereas larger changes occur in regions with strong precipitation gradients, notably the Jura. Because the 15 km window is small relative to the regional divisions, smoothing mainly affects locations close to sharp transitions. A single window was retained for consistency across regions and return periods. We did not optimise the window against gauge-derived return levels, to avoid tuning the radar estimates to the gauge network; Figure S4 compares the radar–gauge agreement obtained with the smoothed and unsmoothed estimates.

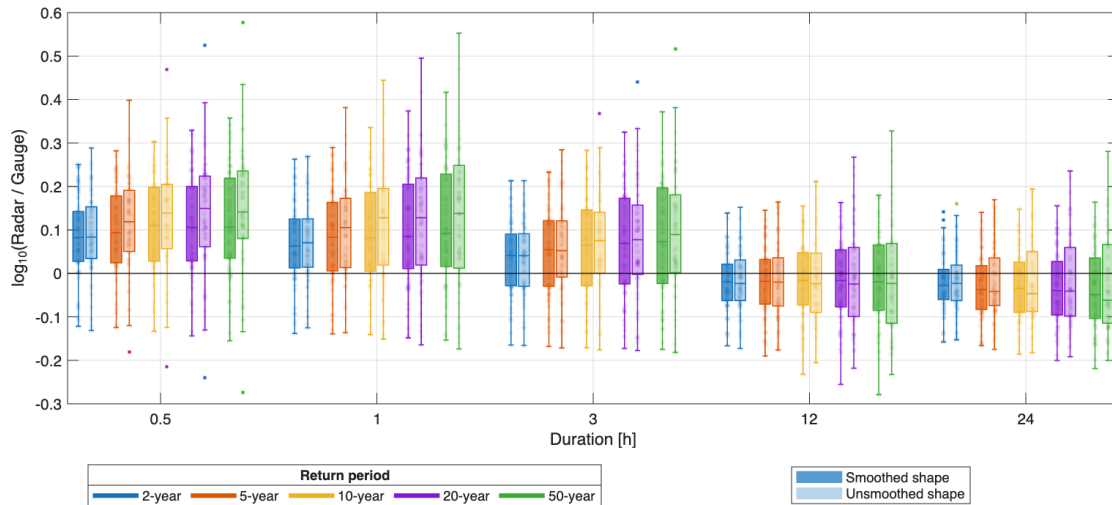


Figure S4. Bias between radar- (1 km² pixel scale) and gauge-derived return levels, expressed as $\text{Log}_{10}\left(\frac{\text{Radar}}{\text{Gauge}}\right)$, for estimates obtained using smoothed and unsmoothed Weibull shape parameters.

6) The manuscript uses 100 block-bootstrap repetitions to estimate uncertainty ranges. It is unclear whether 100 repetitions are sufficient, especially for high return periods and for spatially distributed estimates. With only 100 bootstrap samples, the empirical percentiles are based on the tails of a small bootstrap ensemble, which may be unstable.

The authors should justify the choice of 100 repetitions or increase the number substantially, for example to 500 or 1000 if computationally feasible. At minimum, they should include a sensitivity test showing that the uncertainty intervals are not materially affected by increasing the number of bootstrap repetitions.

This point is important because the uncertainty intervals in Fig. 6 appear narrow, and the authors attribute this partly to shape-parameter smoothing. The reader needs to know whether the narrow intervals reflect genuine stability or are partly an artefact of the bootstrap design and smoothing procedure.

Response: Thank you for raising this important point. We find that published SMEV applications have used bootstrap ensembles ranging from 100 (e.g. Rosin et al., 2024) to 1,000 repetitions (e.g. Marra et al., 2022). Thus, 100 repetitions fall within published practice but represent the lower end of the range. We have therefore performed a bootstrap-convergence test to investigate how the number of bootstraps affects the results and identify the optimum number. This analysis shows that 100 repetitions are sufficient for stable median return levels (within about $\pm 2\%$ of the 1000-repetition estimates), but not for stable interval widths, which differed by more than 10% in some regions. We therefore increased the ensemble to 800 repetitions, at which interval widths are within about $\pm 2\%$ of the 1000-repetition reference, and recomputed all reported bootstrap intervals and figures. Regarding the narrow intervals in Figure 6, these are conditional on the shape smoothing, which pools information from neighbouring pixels; intervals obtained without smoothing are substantially wider (Figure S2). Sect. 4.3 now states explicitly that the intervals do not include structural uncertainty, including that associated with the choice of smoothing window.

The following text explaining the bootstrap test has been added to Supplement Sect. S2 together with corresponding Figures S5 and S6:

The sensitivity of the estimated return levels and their uncertainty intervals to the number of block-bootstrap repetitions was assessed for the six regional analysis boxes using the 10 km² aggregation area and 60 min duration. Figure S5 shows the sensitivity of the median 20-year return levels and the corresponding 5th–95th percentile interval widths to ensemble sizes ranging from 1 to 1000 repetitions. Figure S6 presents their percentage differences relative to the estimates obtained using 1000 repetitions.

We find that the median return-level estimates converged relatively rapidly across all six regions. For small numbers of bootstrap iterations, the estimates fluctuate, but beyond approximately 100 repetitions, the median estimates are stable. Relative to the results obtained with 1000 repetitions, the median estimates are generally within approximately $\pm 2\%$ after around 100 repetitions.

The estimated uncertainty-interval widths are more sensitive to the number of bootstrap repetitions, and continue to fluctuate beyond 100 repetitions, particularly in regions with wider return-level distributions (Southern Alps and Jura Mountains). However, beyond approximately 200 repetitions the interval widths become progressively more stable. At 100 repetitions, the interval widths differ from the corresponding 1000-repetition estimates by more than 10% in some regions. By 600 repetitions, these differences are generally within approximately $\pm 5\%$, and by 800 repetitions they approach $\pm 2\%$ across the six regions.

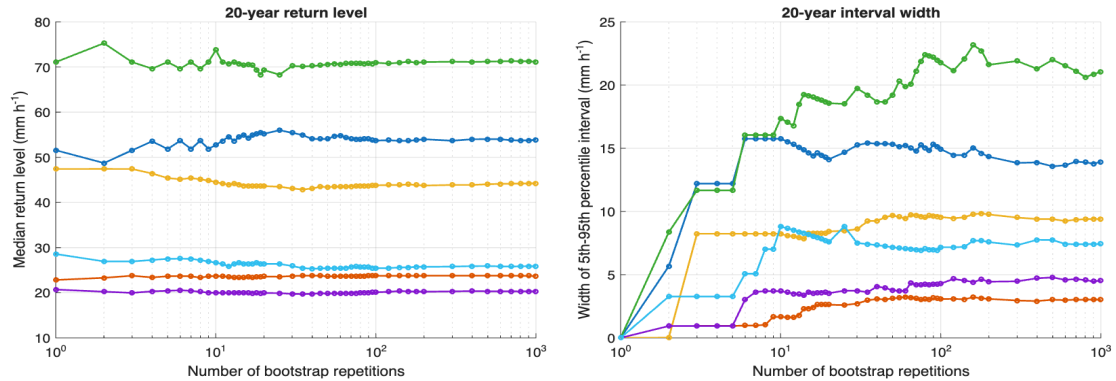


Figure S5. Sensitivity of regional rainfall return-level estimates to the number of block bootstrap repetitions. Results are shown for the six regional analysis boxes using a 10 km² area and 60 min duration. The figure shows the median 20-year return level and the corresponding 5th–95th percentile bootstrap interval width for ensemble sizes ranging from 1 to 1000 repetitions.

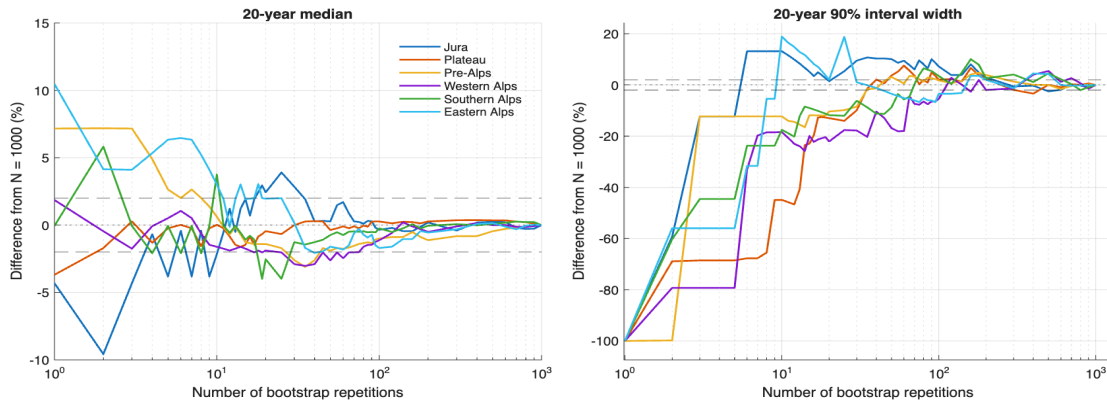


Figure S6. Relative sensitivity of regional 20-year rainfall return-level estimates to the number of block bootstrap repetitions. Percentage differences in the median 20-year return level and in the corresponding 5th–95th percentile interval width are shown for the six regional analysis boxes, using the estimates based on 1000 bootstrap repetitions as the reference. Results are for a 10 km² aggregation area and 60 min duration.

7) Several parts of the results section contain interpretation and literature comparison that would fit better in the discussion. For example, the comparison of spatial patterns with previous Swiss radar climatologies and the physical explanation of orographic mechanisms could be moved from Sect. 4 to the discussion.

At the same time, the discussion section currently does not refer sufficiently to the manuscript's own results and figures. It contains many useful ideas, but the connection to the figures is sometimes weak. The discussion should more consistently refer back to manuscript's figures so that the reader can verify the claims.

I recommend reorganizing the manuscript as follows:

1. Present the main return-level maps and IDAF/ARF curves in the results section with direct description of what each figure shows.
2. Move broader physical explanations and comparisons with previous studies to the discussion.
3. In the discussion, explicitly refer back to the relevant figures when interpreting regional differences, ARF behaviour, and simple-scaling behaviour.

Figures 6, 7, and 8 are central to the manuscript, but the text does not provide enough detailed interpretation of them. Figure 8 in particular is introduced but not adequately discussed in the results section. This makes it unclear why the figure is included and what the reader should learn from it.

Response: Thank you for this helpful suggestion. We have reorganised the relevant material to distinguish more clearly between the direct presentation of the results and their broader physical interpretation. Section 4.2 retains the return-level maps and a direct description of the spatial patterns they show, while the comparisons with previous Swiss climatologies and the explanations involving orographic lifting, moisture availability and regional precipitation regimes have been moved to the Discussion.

The former Figures 6 and 7 have been combined into two clearly separated parts of revised Figure 6. Figure 6(a) presents IDAF relationships by region, whereas Figure 6(b) reorganises the results by fixed aggregation area to facilitate comparisons among regions and to identify departures from log-log linearity. The regional panels have also been ordered geographically, and the sub-panels of Figure 6(b) are now labelled (1)–(5). The Results text (Section 4.3) now explains the distinct purpose of each panel. In addition, Figure 7 (old Figure 8) is further discussed:

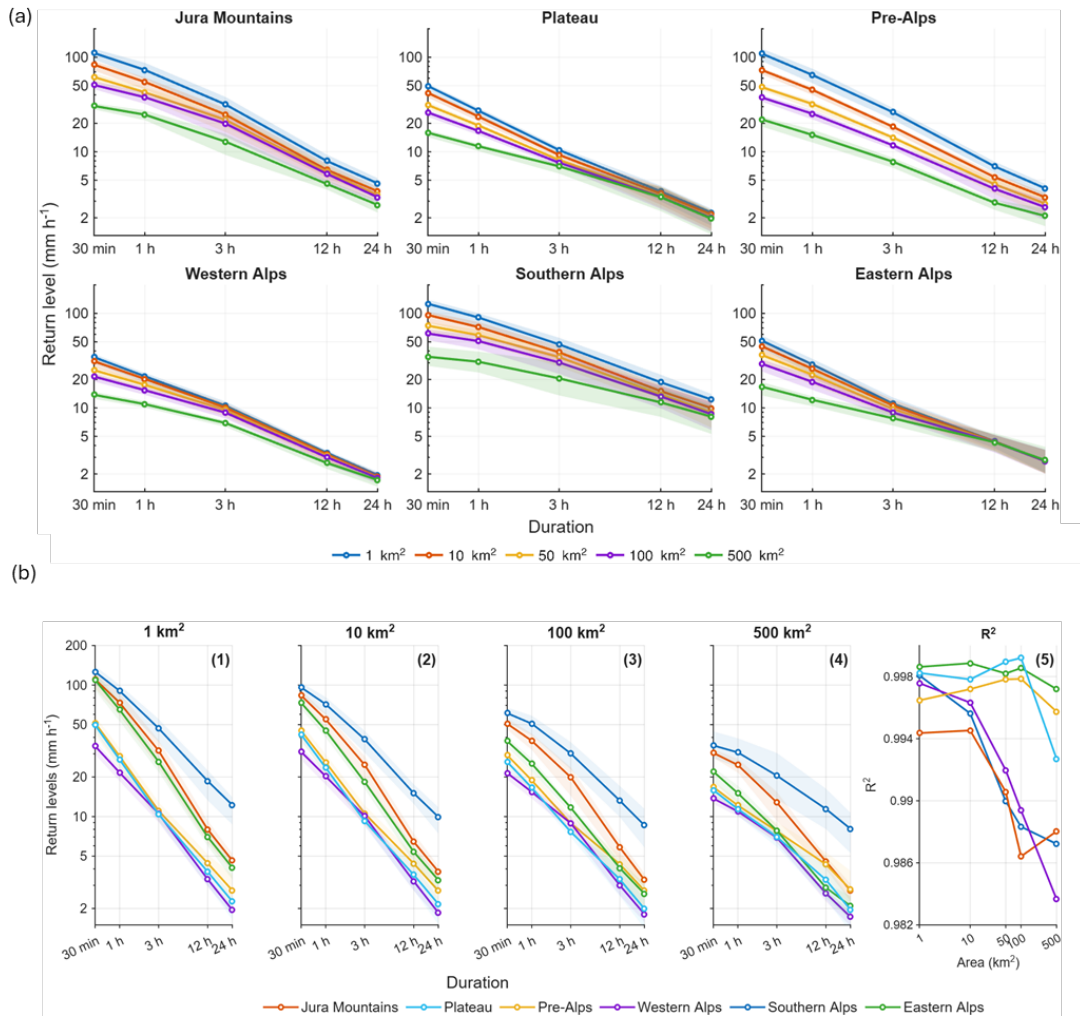


Figure 6. Twenty-year rainfall Intensity–Duration–Area–Frequency (IDAF) relationships for the six analysis regions. (a) Regional IDAF curves, with each panel showing one region and individual curves representing different aggregation areas. (b) Panels (1)–(4) show return levels as a function of duration for aggregation areas of 1, 10, 100 and 500 km², and panel (5) shows the coefficient of determination (R^2) of the linear regression between log-transformed return level and duration as a function of aggregation area. Shaded bands indicate approximate 90% sampling-uncertainty intervals derived from 800 bootstrap replicates; they do not include radar-QPE uncertainty. The locations of the regional analysis boxes are shown and labelled in Figure 1.

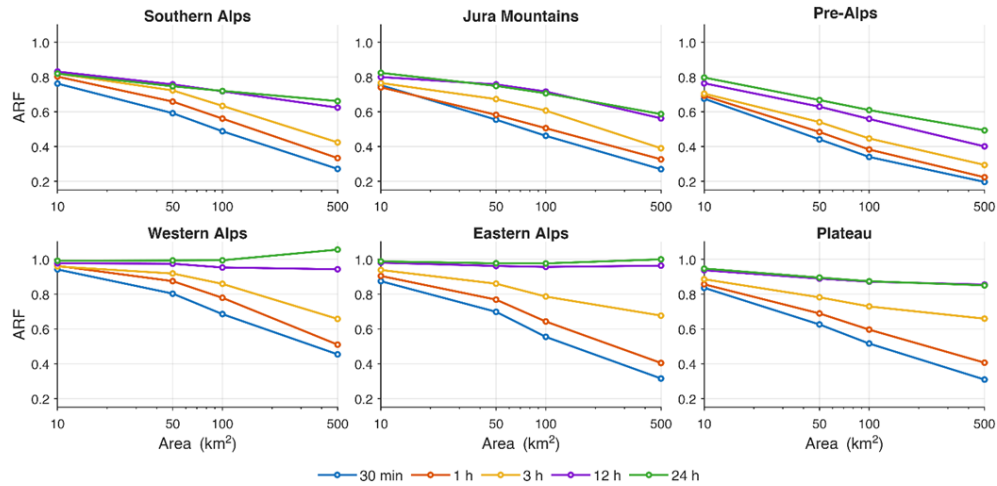


Figure 7. ARFs estimated for 20-year return periods as a function of aggregation area for the six analysis regions. Regional analysis-box locations are shown in Figure 1.

Figure 6 presents the IDAF relationships for 20-year return periods in two forms. Figure 6(a) presents 20-year IDAF curves for each region (additional return periods are shown in Figure S11). Figure 6(b) presents the same curves, organised by aggregation area, and includes the R^2 of the linear regression between the log-transformed intensity and duration. Figure 7 then presents the corresponding ARF values.

These two panels show that return levels decrease substantially with both increasing spatial and temporal scale, with clear regional contrasts. The highest return levels occur over the Southern Alps, Jura, and Pre-Alps, as expected. The 90% uncertainty bounds are generally narrow because spatial smoothing pools information across neighbouring pixels and reduces sampling variability. In line with this interpretation, IDAF curves derived without shape smoothing exhibit substantially larger uncertainty (see Figure S2).

Figure 7 shows substantial regional and duration-dependent differences in the 20-year, 1 km^2 -reference ARFs. ARFs generally decrease with increasing aggregation area, with the strongest reductions occurring at shorter durations. The Pre-Alps consistently exhibit the lowest ARFs, while the Southern Alps and Jura also show pronounced decreases with increasing area. ARFs generally increase with duration, and the decline with aggregation area becomes progressively weaker, particularly in the Eastern and Western Alps and on the Plateau. Properties of the derived IDAF curves, parameters, and ARFs are discussed further in Section 6.

8) The event-based severity diagrams are interesting and potentially valuable. However, some statements are written too generally given that they are based on individual case studies. For example, the statement that the 21 June 2024 event implies highest hazard for small to medium-sized catchments with sub-hourly to multi-hour response times is physically plausible, but the authors should clarify whether this is a new finding, a confirmation of established hydrological understanding, or an illustration of how the IDAF framework diagnoses event severity. The event

analyses should be framed as illustrative applications rather than as independent proof of general hydrological rules. The authors should also avoid wording that may appear contradictory, such as describing an event that caused fatalities and destructive impacts as not producing “major flooding.” If the intended meaning is that the event did not produce major flooding in larger river basins, this should be stated explicitly.

Response: We agree and have reframed the three event analyses as illustrative applications of the radar-based IDAF framework rather than independent demonstrations of universal hydrological rules. The following statement was added to Sect. 5:

The three case studies are intended as illustrative applications of the IDAF framework: their scale-dependent severity patterns are interpreted in light of established hydrological understanding, rather than as independent evidence for general relationships between rainfall scale and catchment response.

The contradictory wording has also been revised to avoid minimising an event that caused severe damage and fatalities. For the 21 June 2024 event we now state that:

By contrast, the event did not produce flooding in large rivers, which generally depends on rainfall sustained over longer durations and broader areas.

Specific and technical comments

9) The introduction is generally well motivated. However, the novelty statement should be sharpened. The authors should explicitly distinguish between the existing SMEV/areal-SMEV methodology and the new contribution of applying it to Swiss radar data and hydrological events.

Response: Thank you. We have revised the introduction to distinguish more explicitly between the previously developed methodology and the contribution of the present study.

The Introduction first identifies Rosin et al. (2024) as the source of the IDAF methodology:

Recently, Rosin et al. (2024) extended the SMEV framework to the areal scale, using 12 years of radar-derived precipitation data over the Eastern Mediterranean to estimate return levels across multiple spatiotemporal scales. The study showed that short radar records can be used to reliably incorporate spatial aggregation into extreme rainfall analysis and to drive Intensity–Duration–Area–Frequency (IDAF) relationships at high spatial resolution.

The revised text then clarifies that the present study applies this existing framework to Switzerland and uses its multiscale estimates to examine recent hydrological events:

Here, we apply the SMEV-based IDAF methodology of Rosin et al. (2024) to the complex and heterogeneous Alpine topography of Switzerland. The contribution of this study lies in this application rather than in the statistical method itself: we use the resulting multiscale return levels to examine how topography influences the duration–area scaling of precipitation extremes, how the relevant rainfall scale varies among events with different spatial footprints, and how this relates to recent high-impact hydrological events.

Finally, we have added an explicit statement of the study's three contributions at the end of the Introduction:

The specific contributions of this study are to: (i) evaluate radar-derived return levels against independent long-term rain-gauge records, thereby assessing the reliability of estimates obtained from the relatively short radar archive; (ii) produce spatially distributed IDAF estimates across the contrasting topographic regions of Switzerland and examine how the spatial patterns and duration–area scaling of extreme rainfall vary among regional topographic and precipitation regimes; and (iii) illustrate the application of these multiscale estimates through the analysis of three recent high-impact flood events, including the July 2021 regional floods and the June 2024 event. By linking observed event rainfall to duration- and area-dependent return levels, the event analyses demonstrate how the IDAF framework diagnoses changes in rainfall rarity and hydrological relevance across spatiotemporal scales.

10) Line 113: The term “inner-Alpine valleys” should be defined more clearly. Please refer explicitly to the relevant regions in Fig. 1, for example Western Alps and Eastern Alps, or mark representative valleys if the term refers to specific subregions.

Response: Thank you for this comment. We have revised the terminology throughout the manuscript. Where the text refers to sheltered inner valleys, it now specifies that these are the inner valleys of the Western and Eastern Alps, corresponding to the regional divisions shown and labelled in Figure 1. For example, Sect. 2 now reads:

Conversely, the inner valleys of the Western and Eastern Alps lie in persistent rain shadows, experiencing the driest conditions. [...] Annual precipitation exceeds 2,000 mm in the Pre-Alps, Jura, and Southern Alps, while the inner valleys of the Western and Eastern Alps receive 400–800 mm (Isotta et al., 2014).

and Sect. 6.1 reads:

The Eastern and Western Alps exhibit the weakest areal reduction (highest ARFs) and near-full convergence of IDAF curves at long durations, indicating reduced spatial variability and more homogeneous storm structures. These inner-Alpine valleys lie in pronounced rain shadows where incoming air masses are partially depleted on the outer slopes (Frei and Schär, 1998).

11) Figure 1: The figure is important for the rest of the manuscript, but some graphical elements are unclear. What do the thin blue lines represent? Please define all lines and symbols in the caption. In addition, the six regional analysis boxes should be labelled directly on the map or in an inset. Since these boxes are used later for the IDAF curves, the reader should not have to infer which box corresponds to Jura, Plateau, Pre-Alps, Western Alps, Eastern Alps, or Southern Alps.

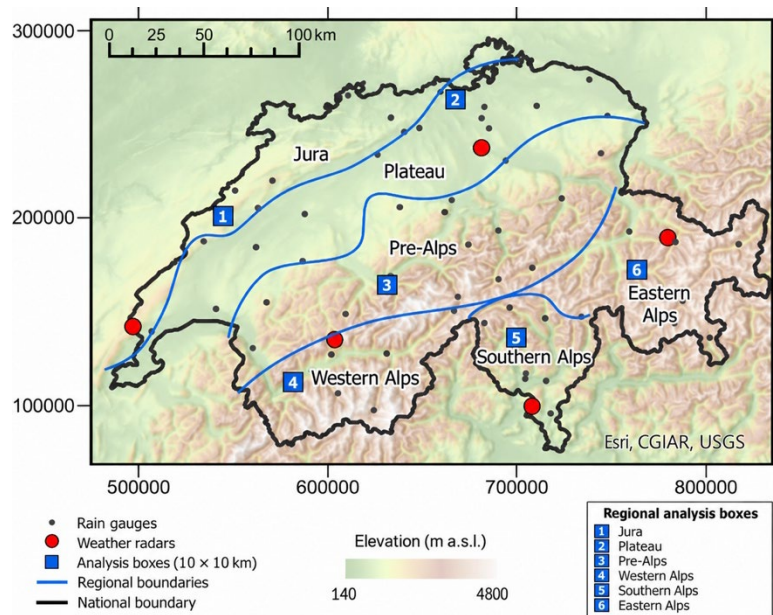


Figure 1. Geographical map of the study area showing the six topographical regions (Jura Mountains, Plateau, Pre-Alps, Western Alps, Southern Alps and Eastern Alps). Thin blue lines delineate the approximate regional boundaries; numbered blue squares mark the six $10 \times 10 \text{ km}^2$ regional analysis boxes (1 Jura, 2 Plateau, 3 Pre-Alps, 4 Western Alps, 5 Southern Alps, 6 Eastern Alps) used in Figures 6 and 7; black dots mark rain gauges; red circles mark weather radars; and the black outline marks the Swiss national boundary. Shading shows elevation. Basemap sources: Esri, CGIAR, USGS.

Response: We thank the reviewer for identifying these ambiguities. We have revised Figure 1 and expanded its caption so that every line and symbol is explicitly defined. The thin blue lines represent the approximate boundaries of the six geographical regions, the black outline denotes the Swiss national boundary, black dots indicate rain gauges, red circles indicate weather-radar locations, and blue squares identify the six $10 \times 10 \text{ km}^2$ analysis boxes. Each box is now numbered directly on the map and identified in the legend. These revisions make the relationship between the analysis boxes and the regional IDAF and ARF results explicit.

12) In section 3.3, please provide more detail on how the multiplicative summer bias adjustment factor was estimated. Was it calculated using all 60 gauges? Was this factor spatially uniform across all regions? If a single national correction factor is used, the authors should discuss whether this may introduce regional biases in complex terrain.

Response: We have now expanded Sect. 3.3, Radar preprocessing and quality control, to specify how the factor was estimated, confirm it was applied uniformly across the study domain, and clarify its limitations:

We apply a multiplicative mean-field bias adjustment of 0.87 to all radar pixels and summer timesteps. The factor was estimated from pooled summer radar–gauge comparisons at all 60 gauges. It corrects domain-mean bias only; regional and duration-dependent errors related to terrain, radar visibility, beam geometry, gauge exposure, and precipitation regime

may remain. We did not apply regional or duration-specific corrections because these would use the gauge network to tune the local spatial and temporal structure of the radar estimates, increasing the risk of overfitting. Residual radar–gauge differences are therefore interpreted as region- and duration-dependent discrepancies, and further bias correction is beyond the scope of this study.

13) In section 3.4, the definition of independent storms as wet periods separated by at least 24 h of dry weather should be further justified for summer convective precipitation in Switzerland. Is 24 h appropriate for both short-lived convective events and longer synoptic systems? How sensitive are the results to this choice?

Response: We thank the reviewer for raising this important point. We have expanded Sect. 3.4 and added a sensitivity analysis of the ordinary-event definition to the Supplement (Figure S1), showing the sensitivity of the storm catalogue to the dry-separation criterion at the 10 km² spatial scale across the six study regions.

The dry-separation period used to identify independent rainfall events should be selected with consideration of the regional climatology and the temporal scales under investigation (Marra et al., 2020). We selected a 24 h dry-separation period as a conservative compromise intended to limit the artificial subdivision of long rainfall episodes while avoiding the more extensive merging associated with longer separation periods.

Previous sub-daily SMEV applications have used separation periods of 6, 12, or 24 h, depending on the maximum accumulation duration considered and the regional precipitation climatology. A 24 h criterion has been used extensively in SMEV studies, including applications in Alpine regions (Dallan et al., 2023) and specifically for Swiss summer precipitation (Peleg et al., 2025a,b). Dallan et al. (2023) investigated the effect of changing the dry–wet threshold and found that reasonable changes affected event frequency but produced no appreciable change in hourly return levels. However, their sensitivity analysis did not examine the dry-separation interval itself.

More broadly, event-based precipitation studies in Switzerland have adopted inter-event times ranging from 2 h (Gaál et al., 2014) to 24 h (Leonarduzzi and Molnar, 2020), depending on the intended application. The 24 h threshold is therefore a relatively conservative choice within the range used in previous Swiss studies. It is also consistent with the maximum accumulation duration considered in the present analysis, which extends to 24 h.

There is also a theoretical reason why changes in the storm catalogue need not translate into comparable changes in return levels. In SMEV, the distribution of annual maxima depends jointly on the magnitude distribution of independent ordinary events and their mean annual number. Once the separation threshold yields statistical independence of the storms, reasonable alternative choices are therefore expected to give similar return levels, provided that the relevant duration-specific maxima are retained and the event-magnitude tail is fitted appropriately.

A complete investigation of the optimised spatial aggregation and SMEV return-level analysis for different separation thresholds would be computationally prohibitive. We therefore performed a targeted sensitivity test to determine how the separation criterion affects the rainfall events and the duration-specific ordinary event sample. We repeated the event identification using dry-separation periods of 6, 12, 18, 24, 36, and 48 h at the six regional 10 km² analysis boxes.

Increasing the separation period systematically reduced the number of identified storms in all six regions. Relative to a 6 h threshold, the adopted 24 h threshold reduced the mean annual storm

count by approximately 45–60%, demonstrating that shorter thresholds frequently divide rainfall episodes containing extended intra-storm dry intervals. Increasing the threshold from 24 to 48 h produced a further reduction of approximately 33–45%, showing that the events do not become completely insensitive at 24 h and that longer thresholds increasingly merge rainfall periods separated by more than one dry day.

As expected, increasing the separation period also increased event duration, wet duration, total rainfall, and the maximum rainfall associated with individual events (not shown). Conversely, the mean rainfall per event timestep generally decreased because the longer events included more intervening low-rainfall or dry periods. These responses were broadly consistent among the six analysed regions, with no indication that a fundamentally different separation threshold was required for any individual region.

The analysis therefore confirms that event characteristics and the short-duration ordinary-event sample depend systematically on the separation definition. Although the analysis did not identify a uniquely optimal threshold, it indicates that 18–24 h provides a reasonable balance between splitting meteorological episodes and merging distinct storms. We retain 24 h because it is the more conservative value within this range and is consistent with the maximum accumulation duration considered. We note however that this analysis assesses the sensitivity of the ordinary-event sample rather than the complete SMEV return-level estimates.

We have added Figure S1 to the Supplement Sect. S1, Methodological sensitivity and validation, showing the sensitivity of the storm catalogue to the dry-separation criterion, and the following text to Sect. 3.4:

Its suitability was assessed by repeating the event identification using separation periods of 6, 12, 18, 24, 36, and 48 h at representative locations in the six regional analysis boxes. Figure S1 presents the sensitivity of the storm catalogue to the dry-separation criterion. Although the analysis did not identify a uniquely optimal threshold, it indicated that 18–24 h provides a reasonable balance: shorter thresholds substantially increased the number of identified storms, whereas longer thresholds increasingly merged rainfall episodes. We therefore retained the more conservative 24 h threshold, which is also consistent with the maximum accumulation duration considered in this study.

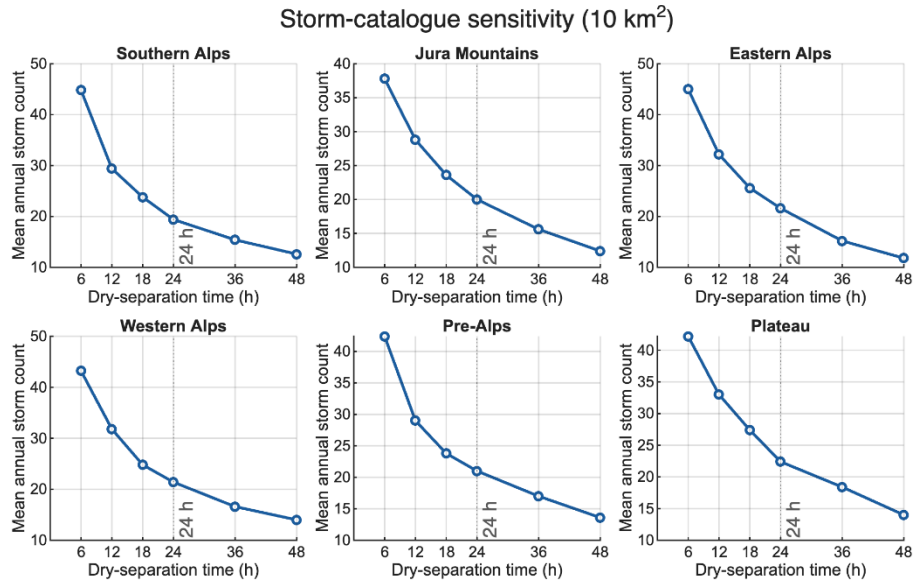


Figure S1. Sensitivity of the storm catalogue to the dry-separation criterion at the 10 km^2 spatial scale across the six study regions.

14) In section 3.5, the 90 % left-censoring threshold should be better justified in the main text. The manuscript states that this was selected from a sensitivity analysis using gauge data. Please summarize the main outcome of that sensitivity analysis and indicate whether the selected threshold is valid across all durations and regions.

Response: We thank the reviewer for this comment. We agree that the explanation for the 90% left-censoring threshold should be stated more clearly. The gauge-based analysis does not demonstrate that 90% is the optimal threshold for every radar aggregation scale; rather, it provides a common threshold within the stable Weibull-tail range that can be applied consistently across the analysis.

We have therefore revised Sect. 3.5 to explain the selection procedure and summarize its principal outcome. The revised text reads:

The Weibull distribution is intended to represent the upper tail rather than the full ordinary-event distribution. Parameter estimation therefore uses left censoring, with observations below the 90th percentile treated as censored. The censoring level was selected using the longer rain-gauge records and the statistical testing procedure described and implemented by Marra et al. (2023). This procedure identifies the minimum threshold above which the null hypothesis of Weibull-tail behaviour is not rejected. Although lower thresholds were acceptable at some gauges, 90% was the lowest common censoring level that provided an acceptable fit across all assessed durations and climatic regions. A common threshold was applied across regions, durations, and aggregation areas to facilitate comparison of the resulting uncertainties.

The gauge-derived threshold is applied to the radar data because gauge and radar events originate from the same regional precipitation regime and are expected to have upper tails governed by broadly similar processes. This does not imply that their intensity distributions

or Weibull parameters are identical, given their different measurement supports and the effects of spatial averaging.

Previous studies provide regional support for this choice. Dallan et al. (2024) identified suitable censoring levels of 90% for 1 h precipitation and 85% for longer durations in the Alpine-Mediterranean region, while Akbary et al. (2026) adopted a common 85% level for sub-daily areal precipitation over Switzerland after testing the Weibull-tail representation. Together with other analyses based on longer observational records (Peleg et al., 2025a, b; Marra et al., 2026), these studies indicate that the Weibull-tail assumption does not rely solely on the nine-year radar record. The radar record is used to estimate the model parameters, with SMEV drawing on multiple ordinary events per year rather than annual maxima alone. The reliability of the resulting return-level estimates nevertheless remains conditional on model adequacy and precipitation-data quality.

The complete statistical testing procedure and its implementation are described by Marra et al. (2023), with code available in Marra (2022); we therefore summarise its application and principal outcome here rather than reproducing the full methodological description.

15) The authors should also report the typical number of ordinary events used for fitting. This would help readers assess the effective sample size of the SMEV estimation.

Response: Thank you for this suggestion. We now report the number of ordinary events used for fitting the SMEV parameters and have added Figure S7 showing its spatial distribution for each aggregation area. The mean annual number of ordinary events ranges from 17 to 22 per grid cell, i.e. approximately 150–200 ordinary events per grid cell over the nine years. Following the 90% left censoring, the number of upper-tail events ranges from 15 to 20. Note that the lower 90% of ordinary events remains represented in the likelihood as censored non-exceedances. See our detailed response to comment #2 above, showing the figure and the revised text.

16) In lines 221 – 224, The manuscript uses 100 bootstrap repetitions. Please justify this number or increase it. With only 100 repetitions, the 5th and 95th percentile estimates may be unstable, particularly for 100-year return levels. A sensitivity test using more repetitions would strengthen the uncertainty analysis.

Response: As detailed in our response to comment #6 above, we added a convergence assessment and increased the production ensemble from 100 to 800 block-bootstrap repetitions. All uncertainty intervals and corresponding figures are now based on the 800-member ensemble.

17) In section 3.6, The ARF definition relative to a 1 km^2 radar pixel is reasonable for gridded data, but the terminology should be handled carefully. Since the denominator is not a true point

measurement, the authors should consistently refer to “1 km²-reference ARFs” or “radar-pixel-reference ARFs” rather than classical point-to-area ARFs.

Response: Thank you for highlighting this distinction. We agree that the denominator represents a 1 km² radar pixel rather than a true point measurement. We have therefore revised Sect. 3.6 and Eq. (2) to define the ARFs explicitly relative to the 1 km² radar scale. We now describe these quantities as 1 km²-reference ARFs at their first definition and clarify that, for brevity, they are referred to simply as ARFs throughout the remainder of the manuscript.

Areal reduction factors (ARFs) quantify the decrease in extreme rainfall intensity with increasing spatial aggregation. In the fixed-area formulation, ARFs are defined as the ratio between areal-average return level and the corresponding point location rainfall return level for the same duration and exceedance probability.

In classical point-to-area applications the reference area (A_0) represents a gauge measurement and approaches 0. However, as the Swiss gauge network is not sufficiently dense to provide robust point-to-area scaling, ARFs are computed here relative to the 1 km² radar pixel, consistent with gridded-data ARF studies where the smallest resolved spatial support is used as the reference scale (e.g. Lutz et al., 2024).

Following the modified Bell method (Bell, 1976) commonly used in design-rainfall studies (e.g. Lutz et al., 2024; Svensson and Jones, 2010), we define the 1 km²-reference ARFs as:

$$\text{ARF}(D, T, A) = \frac{I_{(D,T,A)}}{I_{(D,T,A_0)}} \quad (2)$$

where I is the return-level rainfall intensity for a specific area A , duration D and return period T , and $A_0=1 \text{ km}^2$. All ARFs reported hereafter are 1 km²-reference ARFs, rather than point-to-area ARFs.

18) In Figure 2, the validation is useful, but the caption and text should specify clearly that both radar and gauge return levels are estimated using SMEV, unless a different method is actually used. If GEV estimates are also used anywhere, this needs to be stated explicitly and methodologically described.

Response: We thank the reviewer for pointing out this ambiguity. Both the radar- and rain-gauge-derived return levels shown in Figure 2 were estimated using the SMEV framework. No GEV estimates are used in this validation or elsewhere in the analysis. We have revised the figure caption and the corresponding text to state this explicitly and avoid any possible confusion regarding the estimation method. See the detailed response to comment #3 above, that presents the revised text.

19) In Figure 3, the authors should discuss whether the regional differences are related to radar beam geometry, altitude, gauge exposure, precipitation regime, or the uniform bias-adjustment factor. The current interpretation remains somewhat general.

Response: We agree that the regional differences in radar–gauge bias require a more specific interpretation. The single domain-wide factor cannot remove regional errors related to beam visibility, shielding, vertical reflectivity structure, elevation, gauge exposure or precipitation regime.

We have expanded the interpretation of Figure 3 to discuss those mechanisms as plausible explanations for the observed duration- and region-dependent pattern. We note that the present analysis does not quantify the individual contribution of these mechanisms. They are discussed as plausible explanations for the observed spatial pattern rather than as definitive causal attributions. The text in Sect. 4.1, *Validation against rain-gauge-derived return levels* now reads:

Figure 3 presents the spatial structure of the radar-gauge bias. Radar overestimation is most evident over the Jura and Eastern and Western Alps, whereas underestimation is more prominent over the Plateau, Pre-Alps, and Southern Alps, with the magnitude and sign varying with duration. The broad spatial variability is consistent with previous evaluations of the Swiss radar QPE. Wolfensberger et al. (2021) showed that radar performance varies with radar visibility and reported a tendency towards underestimation in Alpine regions with poor visibility. Germann et al. (2022) identified the combined effects of beam shielding and vertical variations in radar reflectivity as likely dominant sources of QPE uncertainty in mountainous regions once ground clutter has been adequately removed.

20) In line 290 and Figure 4, when stating “across Switzerland,” please make clear in the caption that the boundary shown in Fig. 4 is the Swiss national boundary. This will help readers unfamiliar with the map.

Response: We have revised the caption of Figure 4 to clarify that the outline shown in the figure represents the Swiss national boundary. We have also slightly revised the corresponding text to avoid ambiguity.

Figure 4 caption now reads:

Figure 4. Spatial distribution of 20-year rainfall return levels across durations and aggregation areas. Return levels are shown in mm h^{-1} for durations of 30 min, 3, and 24 h and for areas of 1, 10, and 500 km^2 . Red circles in the top row indicate the equivalent spatial extent of each areal scale. The black outline indicates the Swiss national boundary.

21) In lines 292-295 and Figure 4, when referring to geographic regions such as the Jura Mountains, Alpine barrier, Southern Alps, Plateau, and inner-Alpine valleys, these should be visible or labelled in the figure, or the reader should be directed clearly to Fig. 1. Otherwise, the spatial interpretation is difficult to follow.

Response: We agree that the regional interpretation of Figure 4 is difficult to follow without a clear geographical reference and have therefore revised the text to direct readers explicitly to Figure 1. The revised text in Sect. 4.2 now reads:

For short durations and small areas (30 min-1 h, 1-10 km²) extremes are more localised, with the highest rainfall intensities occurring over the Southern Alps, and secondary maxima along the transition zones between the plains and foothills of the Alpine barrier and the Jura Mountains (geographical regions referred to here are defined in Figure 1). In contrast, the inner valleys of the Eastern and Western Alps and the Plateau show significantly lower values. As duration increases, the Alpine/Jura enhancement weakens, whereas the Southern Alps hotspot persists across all temporal scales.

22) In lines 298 – 303, this paragraph compares the results with previous studies and explains physical mechanisms. It is relevant, but it fits better in the discussion section. The results section should first describe the manuscript's own findings; comparison with previous work should then be used in the discussion to interpret them.

Response: We moved the comparison with previous Swiss studies and the broader physical interpretation from the Results to the Discussion. Sect. 6.1 now reads:

The duration-dependent return-level spatial pattern observed in Figure 4 is consistent with the station-based IDF analyses of Haruna et al. (2023) for 1 and 24-h 100-year return levels across Switzerland, and the climatology of warm season 1 km² extremes reported by Panziera et al. (2016, 2018), using CombiPrecip data (MeteoSwiss, 2018a; Sideris et al., 2014). Panziera et al. attribute the observed distribution to instability along the Alpine slopes which can be easily released through orographic forcing, a mechanism absent over the Plateau. Conversely, at higher altitudes near the Alpine crest convection is inhibited due to less moisture availability.

23) In line 316, the statement that there is “a narrowing of the Southern Alps maximum” is difficult to appreciate visually from Fig. 5. The spatial structure appears broadly similar across return periods. If the authors want to make this point, they should quantify it, for example by showing the area exceeding a given percentile or threshold as a function of return period. The same figure also raises a broader question: are 9 years of data sufficient to resolve the spatial differences among 50- and 100-year return levels? The authors should discuss this more explicitly.

Response: We agree that the claimed narrowing was not demonstrated clearly and have removed that statement. We also removed the 100-year maps and limit the analysis to 50 years. Regarding whether nine years suffice to resolve spatial differences at long return periods, we added the following to Sect. 4.2:

Differences between the 20- and 50-year maps arise largely from extrapolation of the fitted tails and inherit the associated model uncertainty. Regional contrasts that exceed the bootstrap interval widths (Figure S11b) are considered robust, whereas finer-scale differences at 50 years should be interpreted cautiously.

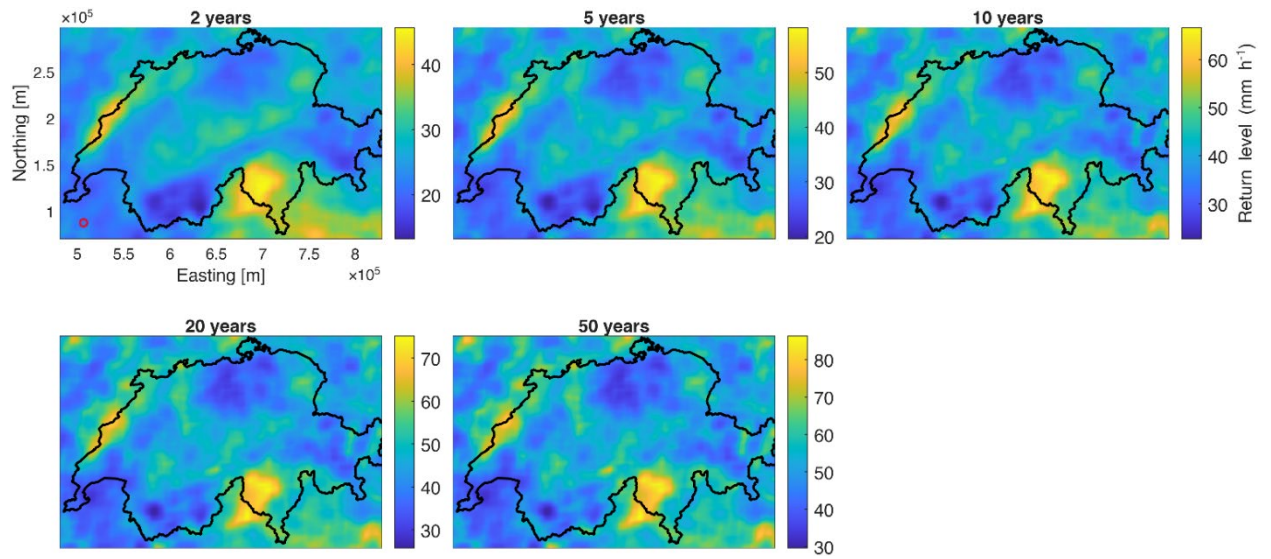


Figure 5. Spatial distribution of 30-min rainfall return levels (mm h^{-1}) for a 50 km^2 aggregation area and return periods from 2 to 50 years. Colour scales differ between panels to emphasise spatial structure at each return period. Red circle (first panel) indicates the equivalent spatial extent of the 50 km^2 aggregation area. The black outline indicates the Swiss national boundary.

24) Figures 6 and 7 convey partly overlapping information. Figure 6 shows IDAF curves by region, while Figure 7 reorganizes similar information by duration and adds the R^2 analysis. The authors should consider merging these figures or moving part of the information to the supplement. If both figures are retained, their distinct purposes should be explained more clearly. In addition, the panels in Fig. 6 could be reorganized geographically. For example, the top row could show Jura, Plateau, and Pre-Alps, and the bottom row could show Western Alps, Southern Alps, and Eastern Alps. This would make the regional comparison more intuitive.

Response: We thank the reviewer for this helpful suggestion. We agree that the former Figures 6 and 7 contain related information, and we have therefore combined them as two clearly separated parts of a revised Figure 6 (see the revised figure in the response to comment #7).

Although both parts are based on the same return-level estimates, their different organisation highlights complementary aspects of the results. Figure 6(a) presents the IDAF relationships separately for each region, allowing the effects of duration and spatial scale to be clearly examined within individual climatic regions. Figure 6(b), in contrast, reorganises the results by fixed aggregation area, making it easier to compare regions and to identify departures from linearity in log-log space across spatial scales. These departures are not so apparent in Figure 6(a) and are quantified in Figure 6(b) using the coefficient of determination, R^2 .

We have revised the figure caption and the corresponding text to clarify these distinct purposes. We have also reorganised the regional panels geographically, as suggested by the reviewer (top row: Jura Mountains, Plateau, Pre-Alps; bottom row: Western Alps, Southern Alps, Eastern Alps). The same regional order is used in Figure 6(b), whose sub-panels are now labelled (1)–(5).

The relevant revised text and Figure 6 in Sect. 4.3 are as follows:

Figure 6 presents the IDAF relationships for 20-year return periods in two forms. Figure 6(a) presents 20-year IDAF curves for each region (additional return periods are shown in Figure S11). Figure 6(b) presents the same curves, organised by aggregation area, and includes the R^2 of the linear regression between the log-transformed intensity and duration. Figure 7 then presents the corresponding ARF values.

These two panels show that return levels decrease substantially with both increasing spatial and temporal scale, with clear regional contrasts. The highest return levels occur over the Southern Alps, Jura, and Pre-Alps, as expected. The 90% uncertainty bounds are generally narrow because spatial smoothing pools information across neighbouring pixels and reduces sampling variability. In line with this interpretation, IDAF curves derived without shape smoothing exhibit substantially larger uncertainty (see Figure S2).

25) The regional names should be made clearer. The reference to Fig. 1 is not sufficient because the boxes in Fig. 1 are not labelled directly. Please label the regional boxes in Fig. 1 or add a small map inset to Fig. 6.

Response: The six analysis boxes are now labelled directly in Figure 1 (see the revised figure in response to comment #11) with their regional names, and the revised Figure 6 caption refers to those labels (response to comment #24).

26) The text should explicitly describe Fig. 7. In particular, the authors should explain what the high R^2 values imply.

Response: Thank you. We have revised the text to explain in more detail the high R^2 values, qualify their interpretation, and describe their modest regional variation with aggregation area.

6.2 Simple versus multiscaling behaviour, now reads:

Against this background, the Swiss regional IDAF curves (Figure 6(b)) remain near linear in log–log space across all the analysed spatial scales. This is quantified in the R^2 values from the linear regression between the log-transformed 20-year intensity and duration, which exceed 0.98 for all area sizes. The high values indicate that a single power law closely describes the duration dependence of the estimated return levels over the analysed range. This is consistent with approximately simple scaling over the analysed range, despite the physical complexity of summer convective processes. Because R^2 is computed from five durations only, it is a lenient diagnostic that is insensitive to moderate curvature, and simple scaling would additionally require the scaling exponents to be invariant across return periods, which is not tested here. This near linear behaviour is notable as summer Alpine precipitation is typically associated with high intermittency and convection, conditions often linked to multiscaling behaviour (Molnar and Burlando, 2008).

The area dependence of the R^2 values is modest but regionally variable. The clearest decrease occurs in the Western Alps, consistent with the location of the analysis box within an inner-Alpine rain-shadow valley, where short-duration extremes are dominated by local convective rainfall, whereas larger aggregation areas and longer accumulation periods increasingly incorporate precipitation over adjacent, more exposed slopes. The Southern Alps and Jura Mountains also display a small decrease in R^2 with area; strong orographic forcing produces localised convective extremes at short durations, while longer-duration and larger-area accumulations increasingly include multiple cells and more spatially coherent rainfall. By contrast, the Eastern Alps and Pre-Alps remain close to linear across all analysed areas, indicating that spatial aggregation changes their return levels more smoothly and approximately proportionally across durations. The Plateau shows a noticeable decrease in R^2 only at 500 km².

Nevertheless, all R^2 values exceed 0.98. The differences therefore indicate only small departures from a single power law, rather than distinct scaling regimes or direct evidence of different rainfall-generating mechanisms. Overall, the area dependence is weaker than that reported by Rosin et al. (2024) for eastern Mediterranean desert, coastal, and mountainous regions, where R^2 decreased consistently with increasing aggregation area.

27) Figure 8 is not sufficiently discussed in the results section. Since ARFs are a central component of IDAF analysis, the authors should describe the main patterns in detail. For example, which regions show the strongest areal reduction? How does ARF vary with duration? Are any ARFs greater than 1? What does this imply physically? How does this compare with previous studies?

Response: We have expanded the Results description of Figure 7 (former Figure 8) presenting the 1 km²-reference ARFs (see the revised figure in response to comment #7). The following was added to Sect. 4.3 *IDAF curves*:

Figure 7 shows substantial regional and duration-dependent differences in the 20-year, 1 km²-reference ARFs. ARFs generally decrease with increasing aggregation area, with the strongest reductions occurring at shorter durations. The Pre-Alps consistently exhibit the lowest ARFs, while the Southern Alps and Jura also show pronounced decreases with increasing area. ARFs generally increase with duration, and the decline with aggregation area becomes progressively weaker, particularly in the Eastern and Western Alps and on the Plateau. Properties of the derived IDAF curves, parameters, and ARFs are discussed further in Section 6.

The following was added to Sect. 6.1, *Spatial patterns and spatiotemporal scaling of extreme precipitation*:

The Western Alps exhibit a 24 h ARF slightly exceeding 1 at the 500 km² scale, indicating that the estimated return level of area-averaged rainfall is slightly higher than the corresponding return level at the 1 km² reference scale. Although conventional engineering ARFs are generally expected to decrease with area and remain at or below 1, the ARF used here is defined as the ratio between the return level of area-averaged rainfall and the corresponding

return level at the 1 km² radar reference scale, evaluated at the same return period. This location-specific ratio therefore differs from a conventional regional engineering ARF and is not inherently constrained to values at or below 1.

Values exceeding 1 have occasionally been reported empirically for fixed-area ARFs, particularly in regions with strong spatial precipitation gradients (e.g. Catchlove and Ball, 2003; Haruna et al., 2024). In the Western Alps, the 1 km² reference-scale return levels are relatively low within the sheltered inner-Alpine valleys, whereas adjacent, more exposed slopes exhibit higher 24 h return levels. As the aggregation footprint expands to include these higher-return-level areas, the area-averaged return level can slightly exceed that at the 1 km² reference scale. Thus, an ARF greater than 1 does not imply that spatial aggregation increases rainfall; rather, it reflects spatial heterogeneity in extreme precipitation and the comparatively low return level at the reference scale. A similar behaviour is observed in the Eastern Alps, where the 24 h ARF approaches 1 (0.9997) at 500 km².

We also compare the ARF behaviour with previous studies in Sect. 6.1:

The magnitude and shape of the ARF curves quantify regional differences in the spatial organisation of extreme precipitation across Switzerland. Many studies have demonstrated that ARFs tend to decay faster with area for short durations than for long durations (Ramos et al., 2005; Breinl et al., 2020). This is attributed generally to the predominantly convective nature and small spatial extent of short-duration events (Sivapalan and Blöschl, 1998). We similarly find that, in all regions, short-duration ARFs decrease more strongly with increasing area.

28) In lines 334 – 338, the opening paragraph of Section 5 explains the general applicability of IDAF relations and severity diagrams. This material may fit better in the methodology section, where the event-severity analysis is introduced. Section 5 could then begin directly with the selected events and the rationale for choosing them.

Response: Thank you for this suggestion. We prefer to retain the introductory paragraph in Sect. 5 because its purpose is interpretive rather than methodological, and is used to transition from the IDAF climatology presented in Sect. 4 to its application to individual events in Sect. 5.

29) Figure 9 may be redundant with Fig. 1. The authors could consider merging the event-location information with Fig. 1 or providing Fig. 9 as an inset or supplementary figure. If Fig. 9 is retained, the caption should explain why a separate location figure is needed.

Response: We retained the separate event-location figure (revised Figure 8) because it shows the precise domains used for the three case studies, which are much smaller than the national-scale regions and cannot be shown legibly in Figure 1. The caption now states this purpose and defines the event-domain rectangles, gauges and other symbols. To reduce redundancy, the figure may alternatively be moved to the Supplement if required.

Revised caption of Figure 8 now reads:

Figure 8. Locations of the three extreme hydrological events. Coloured rectangles indicate the analysis domains for each event. Red circles mark the locations of the five radar stations, and black circles indicate rain-gauge locations, and the black outline indicates the Swiss national boundary. The event domains are shown separately from Figure 1 because they are much smaller than the regional divisions and cannot be displayed legibly at national scale. The corresponding multiscale severity diagrams for these events are shown in Figures 9–11. Basemap sources: Esri, CGIAR, USGS | Powered by Esri.

30) In lines 360 – 361, the statement that the patterns imply the highest hazard for small to medium-sized catchments with sub-hourly to multi-hour response times is hydrologically reasonable. However, the authors should clarify that this is an interpretation of this event’s scale-dependent rainfall structure, not a general conclusion derived from a single event.

Response: We revised the wording in Sect. 5 to make clear that this is an event-specific interpretation:

For this event, the concentration of the largest return periods at short durations (30 min to 1 h), for both small and large aggregation areas, is consistent with high rainfall forcing of the steep, rapidly responding catchments affected during the event (Borga et al., 2014). In the steep Mesolcina catchments, intense short-duration rainfall rapidly exceeded infiltration and channel capacity, producing rapid runoff, sharp rises in streamflow, and debris flows. By contrast, the event did not produce flooding in large rivers, which generally depends on rainfall sustained over longer durations and broader areas.

We do not present this case as a general hydrological rule, as we now clarify in the same section, before the event descriptions:

The three case studies are intended as illustrative applications of the IDAF framework: their scale-dependent severity patterns are interpreted in light of established hydrological understanding, rather than as independent evidence for general relationships between rainfall scale and catchment response.

31) In page 16, and regarding Mesolcina event, the manuscript states that the event lacked the duration to drive major flooding in larger river basins. This should be clarified, because the event itself caused severe impacts and fatalities. The current wording could be misread as minimizing the severity of the event.

Response: We agree and revised the passage to avoid minimising the event. It now acknowledges the rapid runoff and debris flows in the steep Mesolcina catchments, and states only that the event did not produce flooding in large rivers, which generally requires rainfall sustained over longer durations and broader areas. The revised text is as follows:

For this event, the concentration of the largest return periods at short durations (30 min to 1 h), for both small and large aggregation areas, is consistent with high rainfall forcing of the steep, rapidly responding catchments affected during the event (Borga et al., 2014). In the steep Mesolcina catchments, intense short-duration rainfall rapidly exceeded infiltration and channel capacity, producing rapid runoff, sharp rises in streamflow, and debris flows. By contrast, the event did not produce flooding in large rivers, which generally depends on rainfall sustained over longer durations and broader areas.

32) In line 365, the phrase “significant discrepancy” between gauge- and radar-derived return periods needs clarification. From Fig. 10, some gauges located near affected regions appear to show return periods broadly consistent with nearby radar estimates. If the discrepancy refers to gauges missing the storm core because of their spatial distribution, this should be stated more precisely. The issue may be sparse sampling rather than disagreement where radar and gauges are colocated.

Response: Thank you for this clarification. We agree that the original phrase “significant discrepancy” could be interpreted as disagreement between radar and gauge estimates at co-located locations. We therefore changed the text to read (the text also clarifies that the Grono gauge, mentioned earlier in the section, is not among the 60 long-term gauges):

A discrepancy is evident between the gauge- and radar-derived return periods. The long-term gauge network, which does not include the Grono gauge, was not sufficiently dense to fully capture the spatial distribution of rainfall during the event. Only two of these gauges recorded return periods exceeding 2 years; both were located outside the highest-intensity part of the rainfall band and recorded return periods below 5 years across all durations. Radar, by contrast, captured the narrow but intense rainfall band and therefore provided a more complete representation of the event’s spatial structure and scale-dependent severity.

33) In lines 383 – 384, the statement that the maximum radar-derived return period is only 33 years, compared with more than 3000 years for the downtown gauge, is important but not sufficiently supported by Fig. 11 as currently presented. Visually, the gauge dot and adjacent radar colours appear relatively similar. The authors should provide the numerical values explicitly, state which gauge is used, and explain how the gauge return period was calculated. If the difference is mainly due to sub-kilometre spatial variability and 10 min versus 30 min accumulation differences, this should be clearly stated. Otherwise, the reader may not understand how the manuscript supports the 3000-year claim.

Response: Thank you for identifying this inconsistency. On re-examining the figure, we found that the downtown Lausanne gauge had inadvertently been omitted from the original version, and the visible gauge marker was the nearby Pully station. This explains why the gauge marker and adjacent radar colours appeared relatively similar and why the figure did not support the stated >3000-year estimate. We have corrected the revised figure (now Figure 10) to include and label both stations. Because all return periods exceeding 100 years are grouped within the 100+ colour class, the exact >3000-year value is provided in the text. The corrected figure is shown below.

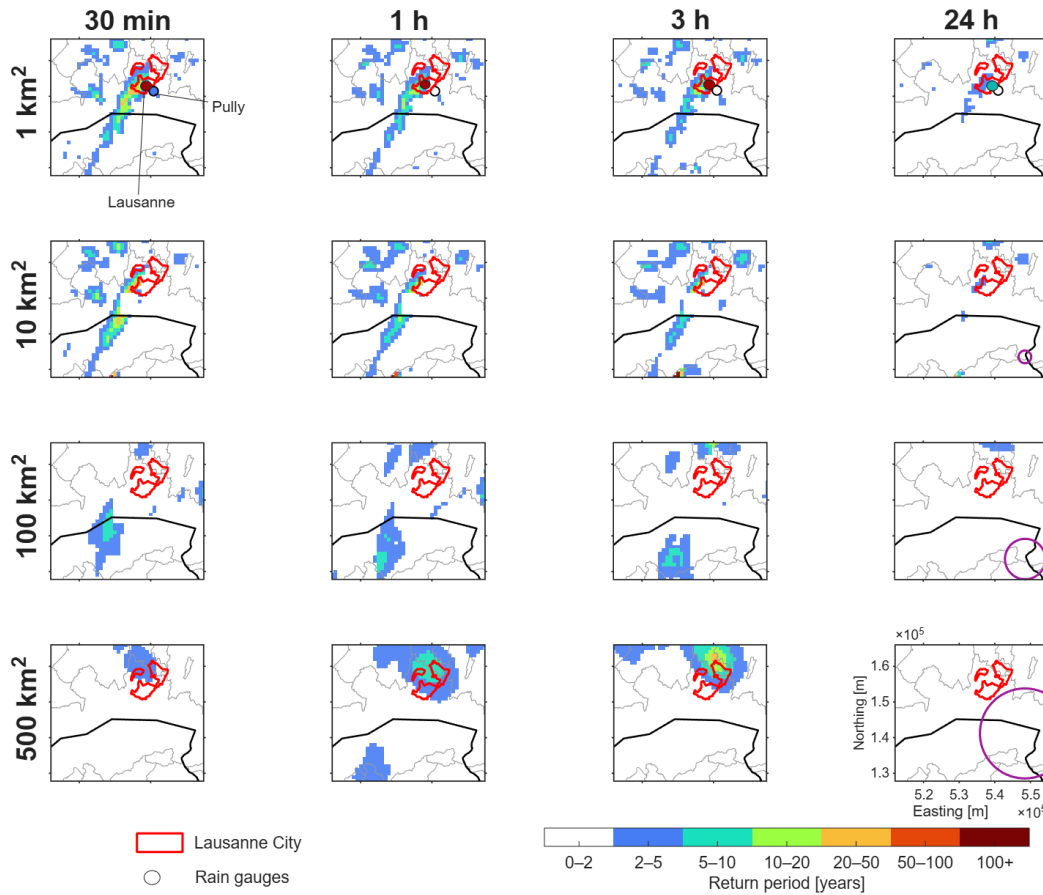


Figure 10. Same as Figure 9, but for the 11 June 2018 event.

We agree that the manuscript did not adequately support the estimated return period exceeding 3000 years or explain why it differs so substantially from the radar-derived estimate. We have revised the text to provide the numerical values explicitly, identify the gauge and reference station, explain how the gauge return period was calculated, and clarify the temporal and spatial bases of the comparison. Both estimates correspond to a 30 min accumulation, but they differ in spatial support. The revised text reads:

However, even at the smallest aggregation considered (1 km² and 30 min), the maximum radar-derived return period within the Lausanne analysis domain is only 33 years. In contrast, the observed 30 min accumulation at the downtown Lausanne gauge corresponds to an estimated return period exceeding 3000 years, although it was computed using the SMEV parameters of the nearby Pully station (44 years of data) due to the short length of the Lausanne series (nine years of data). The estimated return period exceeding 3000 years should consequently be interpreted as evidence that the observed accumulation lies far beyond the fitted Pully distribution, rather than as a precise recurrence-interval estimate.

The discrepancy probably reflects both spatial representativeness and event-specific radar QPE errors. The gauge measures rainfall at a point, whereas a radar pixel represents the mean areal rainfall over a 1 km² area and therefore attenuates highly localised peaks that can occur over scales of only a few hundred metres (Peleg et al., 2018). The IDAF estimates are consequently applicable only at spatial supports of 1 km² and larger and should not be

used to reconstruct sub-kilometre urban rainfall peaks. In addition, Gabella et al. (2018) assert that deviations from the assumed exponential drop size distribution (DSD) also led to a severe underestimation of the 1 km² radar QPE peak rainfall during the event, when using the standard operational Swiss-tuned Z–R relationship. Under this relationship, the instantaneous radar-derived rate near the core is measured at 124 mm h⁻¹, whereas adopting a maritime-tropical Z-R relation yields 265 mm h⁻¹, in closer agreement with the gauge.

34) In line 395, the term “northern Alpine slopes” is difficult to interpret from Fig. 12 alone. Please either mark this region in the figure or use terminology that directly corresponds to the regional divisions shown in Fig. 1.

Response: Thank you for this comment. We have revised the text to clarify how the term “northern Alpine slopes” relates to the regional divisions used in Figure 1. The revised text specifies that this area encompasses parts of the Pre-Alps and the northern sectors of the Western and Eastern Alps shown in Figure 1:

In contrast to the convective bursts described above, July 2021 saw prolonged heavy rainfall driven by cool moist air masses rising from the Atlantic. Precipitation totals exceeded 100 mm across extensive areas, with 150 mm recorded along the northern side of the Alpine region, encompassing parts of the Pre-Alps and the northern sectors of the Western and Eastern Alps (FOEN, 2023; Figure 1).

35) In lines 426 – 431, this paragraph discusses Pre-Alps, Jura, Southern Alps, convective storms, stratiform precipitation, severe deep convection, and ARF behaviour. The ideas are relevant, but the paragraph is difficult to follow and should be rewritten. The authors should refer explicitly to their own results, especially Fig. 8, before invoking broader physical explanations.

Response: We have revised the paragraph to distinguish more clearly between our results and their physical interpretation. The revised text first describes the regional ARF patterns shown in Figure 7 (former Figure 8) and then draws on previous studies to provide a plausible explanation. We have also avoided implying that precipitation type was classified directly in our analysis. The revised text in Sect. 6.1 now reads:

Our results show that the Pre-Alps, Jura, and Southern Alps have the smallest ARFs and therefore the strongest decrease in return levels with increasing aggregation area across all durations (Figure 7). This regional pattern is consistent with the occurrence of intense, spatially localised precipitation in these regions. Previous studies have shown that convective extremes are more strongly affected by spatial aggregation than stratiform extremes (by approximately 20–30%; Eggert et al., 2015) and have documented frequent severe convection in the Southern Alps and enhanced convective activity over the elevated terrain of the Pre-Alps and Jura (Feldmann et al., 2023). These findings provide a plausible physical explanation for the stronger areal reduction observed in these regions, although precipitation type is not classified directly in the present analysis.

36) The discussion section should be substantially revised. At present, it contains many relevant physical interpretations, but it does not sufficiently guide the reader through the manuscript's own evidence. The authors should refer frequently to manuscript figures when discussing:

- regional contrasts in return levels;
- duration dependence;
- area dependence;
- ARF behaviour;
- the role of orography;
- differences between convective and longer-duration events;
- the practical hydrological relevance of IDAF curves.

The discussion would be much clearer if each interpretive claim were linked to a specific figure or result.

Response: Thank you, we have reorganised the Discussion around the manuscript evidence. Regional return-level contrasts are linked to Figures 4–6, ARF behaviour to revised Figure 7, parameter patterns to the corresponding supplementary figures. Broader explanations involving orography, convective organisation and longer-duration precipitation follow the observed results and are phrased as interpretations rather than direct causal tests. For example, Sect. 6.1 now opens:

The results reveal that summer rainfall extremes across Switzerland are strongly organised by topography and vary systematically with both duration and aggregation area (Figures 4, 5 and 6).

Minor comments

37) “simplified-variant Metastatistical Exrteme Value” should be corrected to “simplified variant of the Metastatistical Extreme Value.”

Response: We thank the reviewer; we could not find this exact phrase. The closest wording is: “...and its simplified variant, the SMEV, ...”.

38) “Intensity-duration-area-frequency” should be consistently written as “Intensity–Duration–Area–Frequency.”

Response: Corrected and standardised as “Intensity–Duration–Area–Frequency” throughout, including the title.

39) Geographic terms should be used consistently: Jura Mountains, Plateau, Pre-Alps, Western Alps, Eastern Alps, Southern Alps, inner-Alpine valleys.

- The manuscript should avoid introducing figures without substantive discussion in the text.
- Figure captions should define all symbols, lines, regions, and boundaries

Response: Thank you. We have standardised the geographic terminology throughout the manuscript, figures and Supplement, consistently using Jura Mountains, Plateau, Pre-Alps, Western

Alps, Eastern Alps and Southern Alps, corresponding to the regions shown and labelled in Figure 1. Where the text refers to the sheltered inner-Alpine valleys, it now specifies that these are the inner valleys of the Western and Eastern Alps (see also the response to comment #10). Short forms such as “Jura” are used only where the region has already been defined in the same passage.

We also expanded the Results to describe the main message of every figure before it is used in the Discussion, with particular attention to the revised Figures 6 and 7 (the two-part IDAF figure, including the R^2 diagnostic, and the ARF figure; see the responses to comments #7, #24, #26 and #27).

Finally, we revised all figure captions to define the plotted variables, line types, colours, symbols, regions, uncertainty bands and map boundaries, and to state the reference scale used for the ARFs.

Relevant references

Akbary, R., Dallan, E., Astagneau, P. C., Wood, R. R., Marra, F., Brunner, M. I., and Borga, M.: Scale-dependent biases in Alpine sub-daily areal precipitation extremes: added value of convection permitting models, *Hydrol. Earth Syst. Sci.*, 30, 4117–4139, <https://doi.org/10.5194/hess-30-4117-2026>, 2026.

Bell, F. C.: The areal reduction factor in rainfall frequency estimation, Report No. 35, Institute of Hydrology, Wallingford, UK, http://nora.nerc.ac.uk/id/eprint/5751/1/IH_035.pdf, 1976.

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.

Brajkovic, J., Fettweis, X., Noël, B., Van De Vyver, H., Ghilain, N., Archambeau, P., Pirotton, M., and Doutreloup, S.: Increased intensity and frequency of extreme precipitation events in Belgium as simulated by the regional climate model MAR, *J. Hydrol. Reg. Stud.*, 59, 102399, <https://doi.org/10.1016/j.ejrh.2025.102399>, 2025.

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.

Catchlove, R. H. and Ball, J. E.: A hydroinformatic approach to the development of areal reduction factors, in: *Proceedings of the 28th International Hydrology and Water Resources Symposium*, 1.9–1.15, 2003.

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., Borga, M., Fosser, G., Canale, A., Roghani, B., Marani, M., and Marra, F.: A method to assess and explain changes in sub-daily precipitation return levels from convection-permitting simulations, *Water Resour. Res.*, 60, e2023WR035969, <https://doi.org/10.1029/2023WR035969>, 2024.

Deidda, R., Hellies, M., and Langousis, A.: A critical analysis of the shortcomings in spatial frequency analysis of rainfall extremes based on homogeneous regions and a comparison with a hierarchical

boundaryless approach, *Stoch. Env. Res. Risk A.*, 35, 2605–2628, <https://doi.org/10.1007/s00477-021-02008-x>, 2021.

Eggert, B., Berg, P., Haerter, J. O., Jacob, D., and Moseley, C.: Temporal and spatial scaling impacts on extreme precipitation, *Atmos. Chem. Phys.*, 15, 5957–5971, <https://doi.org/10.5194/acp-15-5957-2015>, 2015.

Feldmann, M., Hering, A., Gabella, M., and Berne, A.: Hailstorms and rainstorms versus supercells – a regional analysis of convective storm types in the Alpine region, *npj Clim. Atmos. Sci.*, 6, 19, <https://doi.org/10.1038/s41612-023-00352-z>, 2023.

FOEN: Faktenblatt 1: Ausmass und Ursachen der Hochwasser im Juli 2021, Federal Office for the Environment, Bern, Switzerland, <https://www.bafu.admin.ch/de/hochwasser-an-den-jurarandseen> (last access: 12 October 2026), 2023.

Frei, C. and Schär, C.: A precipitation climatology of the Alps from high-resolution rain-gauge observations, *Int. J. Climatol.*, 18, 873–900, [https://doi.org/10.1002/\(SICI\)1097-0088\(19980630\)18:8<873::AID-JOC255>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-0088(19980630)18:8<873::AID-JOC255>3.0.CO;2-9), 1998.

Gaál, L., Molnar, P., and Szolgay, J.: Selection of intense rainfall events based on intensity thresholds and lightning data in Switzerland, *Hydrol. Earth Syst. Sci.*, 18, 1561–1573, <https://doi.org/10.5194/hess-18-1561-2014>, 2014.

Gabella, M., Frassl, M., Germann, U., Joerg, P. C., and Wild, E.: UrbanRain18: High-resolution rainfall observations in urban environments, in: *Proceedings of UrbanRain18, 11th International Workshop on Precipitation in Urban Areas*, 2018.

Germann, U., Boscacci, M., Clementi, L., Gabella, M., Hering, A., Sartori, M., Sideris, I. V., and Calpini, B.: Weather radar in complex orography, *Remote Sens.*, 14, 503, <https://doi.org/10.3390/rs14030503>, 2022.

Gires, A., Tchiguirinskaia, I., Schertzer, D., Schellart, A., Berne, A., and Lovejoy, S.: Influence of small scale rainfall variability on standard comparison tools between radar and rain gauge data, *Atmos. Res.*, 138, 125–138, <https://doi.org/10.1016/j.atmosres.2013.11.008>, 2014.

Haruna, A., Blanchet, J., and Favre, A.-C.: Modeling intensity–duration–frequency curves for the whole range of non-zero precipitation: a comparison of models, *Water Resour. Res.*, 59, e2022WR033362, <https://doi.org/10.1029/2022WR033362>, 2023.

Haruna, A., Blanchet, J., and Favre, A.-C.: Estimation of intensity–duration–area–frequency relationships based on the full range of non-zero precipitation from radar-reanalysis data, *Water Resour. Res.*, 60, e2023WR035902, <https://doi.org/10.1029/2023WR035902>, 2024.

Isotta, F. A., Frei, C., Weilguni, V., Perčec Tadić, M., Lassègues, P., Rudolf, B., Pavan, V., Cacciamani, C., Antolini, G., Ratto, S. M., Munari, M., Micheletti, S., Bonati, V., Lussana, C., Ronchi, C., Panettieri, E., Marigo, G., and Vertačnik, G.: The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data, *Int. J. Climatol.*, 34, 1657–1675, <https://doi.org/10.1002/joc.3794>, 2014.

Kašpar, M., Hulec, F., Müller, M., and Crhová, L.: Improving design precipitation estimates by combining estimates from high-resolution adjusted radar data and long-term ombrographic measurements, *Atmos. Res.*, 330, 108557, <https://doi.org/10.1016/j.atmosres.2025.108557>, 2026.

Leonarduzzi, E. and Molnar, P.: Deriving rainfall thresholds for landsliding at the regional scale: daily and hourly resolutions, normalisation, and antecedent rainfall, *Nat. Hazards Earth Syst. Sci.*, 20, 2905–2919, <https://doi.org/10.5194/nhess-20-2905-2020>, 2020.

Lompi, M., Marra, F., Deidda, R., Caporali, E., Borga, M., and Dallan, E.: Non-stationary frequency analysis of long-term convection permitting simulations reveals sub-daily extreme precipitation changes in central-southern Europe, *Adv. Water Resour.*, 205, 105071, <https://doi.org/10.1016/j.advwatres.2025.105071>, 2025.

Lutz, J., Roksvåg, T., Dyrddal, A. V., Lussana, C., and Thorarinsdottir, T. L.: Areal reduction factors from gridded data products, *J. Hydrol.*, 635, 131177, <https://doi.org/10.1016/j.jhydrol.2024.131177>, 2024.

Marani, M. and Ignaccolo, M.: A metastatistical approach to rainfall extremes, *Adv. Water Resour.*, 79, 121–126, <https://doi.org/10.1016/j.advwatres.2015.03.001>, 2015.

Marra, F.: A test for the hypothesis: block maxima are samples from a parent distribution with Weibull tail, Zenodo [code], <https://doi.org/10.5281/zenodo.7234707>, 2022.

Marra, F., Borga, M., and Morin, E.: A unified framework for extreme subdaily precipitation frequency analyses based on ordinary events, *Geophys. Res. Lett.*, 47, e2020GL090209, <https://doi.org/10.1029/2020GL090209>, 2020.

Marra, F., Armon, M., and Morin, E.: Coastal and orographic effects on extreme precipitation revealed by weather radar observations, *Hydrol. Earth Syst. Sci.*, 26, 1439–1458, <https://doi.org/10.5194/hess-26-1439-2022>, 2022.

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.

Marra, F., Dallan, E., Canale, A., Prosdocimi, I., Papacharalampous, G., Borga, M., and Papalexiou, S. M.: Apparent heavy tails of sub-daily precipitation explained by the coexistence of lighter-tailed processes, *Geophys. Res. Lett.*, 53, e2025GL119705, <https://doi.org/10.1029/2025GL119705>, 2026.

MeteoSwiss: Documentation of MeteoSwiss grid-data products: hourly precipitation estimation through rain-gauge and radar: CombiPrecip, Federal Office of Meteorology and Climatology MeteoSwiss, https://www.meteoswiss.admin.ch/dam/jcr:2691db4e-7253-41c6-a413-2c75c9de11e3/ProdDoc_CPC.pdf (last access: 12 October 2026), 2018a.

Molnar, P. and Burlando, P.: Variability in the scale properties of high-resolution precipitation data in the Alpine climate of Switzerland, *Water Resour. Res.*, 44, W10404, <https://doi.org/10.1029/2007WR006142>, 2008.

Panziera, L., Gabella, M., Zanini, S., Hering, A., Germann, U., and Berne, A.: A radar-based regional extreme rainfall analysis to derive the thresholds for a novel automatic alert system in Switzerland, *Hydrol. Earth Syst. Sci.*, 20, 2317–2332, <https://doi.org/10.5194/hess-20-2317-2016>, 2016.

Panziera, L., Gabella, M., and Germann, U.: 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.

Peleg, N., Marra, F., Fatichi, S., Paschalis, A., Molnar, P., and Burlando, P.: Spatial variability of extreme rainfall at radar subpixel scale, *J. Hydrol.*, 556, 922–933, <https://doi.org/10.1016/j.jhydrol.2016.05.033>, 2018.

Peleg, N., Koukoulou, M., and Marra, F.: A 2 °C warming can double the frequency of extreme summer downpours in the Alps, *npj Clim. Atmos. Sci.*, 8, 216, <https://doi.org/10.1038/s41612-025-01081-1>, 2025a.

Peleg, N., Koukoulou, M., Rajczak, J., Kotlarski, S., Dallan, E., and Marra, F.: Hotter summers, heavier showers: global warming and its impact on Swiss short-duration rainfall extremes, *Weather Clim. Extremes*, 50, 100829, <https://doi.org/10.1016/j.wace.2025.100829>, 2025b.

Pollock, M. D., O'Donnell, G., Quinn, P., Dutton, M., Black, A., Wilkinson, M. E., Colli, M., Stagnaro, M., Lanza, L. G., Lewis, E., Kilsby, C. G., and O'Connell, P. E.: Quantifying and mitigating wind-induced undercatch in rainfall measurements, *Water Resour. Res.*, 54, 3863–3875, <https://doi.org/10.1029/2017WR022421>, 2018.

Ramos, M. H., Creutin, J.-D., and Leblois, E.: Visualization of storm severity, *J. Hydrol.*, 315, 295–307, <https://doi.org/10.1016/j.jhydrol.2005.04.007>, 2005.

Rosin, T., Marra, F., and Morin, E.: Exploring patterns in precipitation intensity–duration–area–frequency relationships using weather radar data, *Hydrol. Earth Syst. Sci.*, 28, 3549–3566, <https://doi.org/10.5194/hess-28-3549-2024>, 2024.

Sideris, I. V., Gabella, M., Erdin, R., and Germann, U.: Real-time radar–rain-gauge merging using spatio-temporal co-kriging with external drift in the alpine terrain of Switzerland, *Q. J. Roy. Meteor. Soc.*, 140, 1097–1111, <https://doi.org/10.1002/qj.2188>, 2014.

Sieck, L. C., Burges, S. J., and Steiner, M.: Challenges in obtaining reliable measurements of point rainfall, *Water Resour. Res.*, 43, W01420, <https://doi.org/10.1029/2005WR004519>, 2007.

Sivapalan, M. and Blöschl, G.: Transformation of point rainfall to areal rainfall: intensity–duration–frequency curves, *J. Hydrol.*, 204, 150–167, [https://doi.org/10.1016/S0022-1694\(97\)00117-0](https://doi.org/10.1016/S0022-1694(97)00117-0), 1998.

Srivastava, A. N. and Mascaro, G.: Improving the utility of weather radar for the spatial frequency analysis of extreme precipitation, *J. Hydrol.*, 624, 129902, <https://doi.org/10.1016/j.jhydrol.2023.129902>, 2023.

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.

Villarini, G., Mandapaka, P. V., Krajewski, W. F., and Moore, R. J.: Rainfall and sampling uncertainties: a rain gauge perspective, *J. Geophys. Res.-Atmos.*, 113, D11102, <https://doi.org/10.1029/2007JD009214>, 2008.

Wolfensberger, D., Gabella, M., Boscacci, M., Germann, U., and Berne, A.: RainForest: a random forest algorithm for quantitative precipitation estimation over Switzerland, *Atmos. Meas. Tech.*, 14, 3169–3193, <https://doi.org/10.5194/amt-14-3169-2021>, 2021.