Response to anonymous referee #1

The manuscript is about how precipitation lapse rates (PLRs) affect runoff simulation and flood frequency in Swiss mountain catchments, as authors used weather generator outputs to drive HBV hydrological model(s) with differential PLRs from 0 to 10 % per 100m in both components and also isolated the effect of hydrological model PLRs itself by using specific 5% PLRs in WG. The topic is interesting and methodologically is conceptually sound. However, some sections raise questions that should be answered during the revision.

We thank referee #1 for reading the manuscript carefully and providing thoughtful and constructive comments. Below, our responses are presented in blue and the referee comments in black (bold).

General Comments:

GC1: The PLRs range from 0 to 10 % is presented as a reasonable choice, yet 10% seems a bit on the extreme side. The introduction itself (later also discussion) mentioned that short duration extremes can have negative lapse rates (reverse orographic effect). As the WG produces hourly data, the question is why not also explore the reverse orographic effect and/or if both effects could be somehow combined in WG (different based on seasonality? winter is more prone to daily extremes, summer to hourly). This could be a bit further discussed I believe (maybe as suggestion for future work?), no matter that it is already mentioned in the limitations.

Indeed, the 10% lapse rate might seem to be on the extreme side. However, this value has been widely used, for example in Sweden and the Czech Republic and large sample studies for catchments with similar characteristics. Additionally, Finger et al., (2015) used the HBV model in a multiple data set calibration approach to estimate runoff composition in different catchments in Switzerland and showed that the lapse rate parameter varied from 5% to 15% per 100m. Therefore, we included this value both for comparison purposes, but also because it might be indeed representative for some of our higher elevation catchments.

Regarding the reverse orographic effect and the differences based on seasonality, we have thought of assigning different lapse rates based on season and region. However, the lapse rates are highly variable in space and time and the identification of a clear pattern can be very challenging. This is highlighted in the study of Benoit et al. (2024) where the authors show that almost 38% of the cases tested in Switzerland show no clear pattern of precipitation elevation relationship while the reverse orographic effect might not play a very critical role. Additionally, the lapse rates assigned here represent long-term average values per calendar day whereas negative values apply to short-term extremes. Therefore, to assign a negative value seemed not straightforward as a choice and requires additional research based on historical records and even better on different spatially higher resolved precipitation products.

In line 583 of the submitted manuscript we mentioned that "*Future work should therefore aim to implement more flexible and regionally varying precipitation-elevation relationships in both the weather generator and the hydrological model, allowing for local gradients and small-scale variability*" and also in line 677: "*We thus suggest that moving towards more flexible precipitation-elevation relationships may have important implications for understanding the water balance, in particular for high-elevation catchments.*" However, we acknowledge that this point could be further discussed and we have now elaborated in line 583 of the submitted version of the manuscript that "*Future work should therefore aim to identify regional and seasonal variation of precipitation-elevation relationship based on historical records or more sophisticated precipitation products. That would allow different lapse rates to be applied*

to each catchment in both the weather generator and the hydrological, thereby allowing for small-scale variability."

GC2: The baseline of 5% PLR is justified by referencing previous work of the same group (Line 144), while in the introduction they introduced some work with PLR 5%. This raises a question: is the 5% typical across Switzerland only or also used somewhere else? What is the mean value in Switzerland? Might be nice to support the baseline at Line 144 with references out of group too.

Thank you for asking this question. Our lapse rate of 5% per 100 m is supported by recent literature. For example, Ruelland, (2020) used an inverse snow-hydrological modelling approach by calibrating local gradients for two different hydrological models and found mean rates of 30% km^{-1}. Ménégoz et al. (2020) found rates for the European Alps that are on average 33% and 38% km^{-1} for winter and summer respectively. Farinotti et al. (2012) found rates of 0.00–0.07% m^{-1} over 9 high-alpine Swiss catchments, with a mean value of 0.035% m^{-1}. This applies to long-term climatology and not to single events and for this reason can be compared to the parameter value in the HBV model. Lastly as mentioned in GC1, Finger et al. (2015) used the HBV in a multiple data set calibration approach to estimate runoff composition in different catchments in Switzerland and showed that the lapse rate parameter varied from 5% to 15% per 100 m. We have added this information to section 3.1.

GC3: The WG is parameterized on 120 stations (daily, precipitation) across Switzerland and output is later computed via Thiessen polygons to catchment scale. Fig. 3b tries to explain representativeness of stations per basin, which is good, but the manuscript is missing the map showing these stations. Fig. 1 also makes it a bit confusing. -> What does "fully modeled with HBV" mean and why such basins are highlighted? Does it mean that such basins were modeled as semi distributed HBV and the rest as lumped? Lastly, the Sarine and Maggia basins are repeatedly used as example catchments; it would be nice to show them on the map and maybe provide some basic characteristics of these two catchments.

Thank you for this suggestion. First, we made a typographical error: the correct number of stations is 102, not 120 (also given as the sum of the stations in Table 2). We have corrected this in the manuscript. In Figure 1, we show the catchments and the sub-catchments that are used in this study. The highlighted areas represent larger catchments that encompass their respective sub-catchments. We highlighted them to show that we modeled the sub-catchments and the whole catchment independently. We have updated the map to show the sub-catchments more distinctly and the large catchments are outlined. We have also updated the map to include the stations and we have changed the "*fully modelled with HBV*" to "*modelled with HBV*" to avoid any confusion. The names of the two exemplary catchments are shown and some information about them is now provided in Section 2.1.

GC4: The daily to hourly disaggregation by analogue method is referenced extensively but is not adequately described in the manuscript. Authors included five references at Line 161 for details, yet only Viviroli et al. 2022 actually documents this step (Evin et al., 2019, 2018 and also GWEX R code include only 3d to daily). More targeted citations, together with a brief description in the manuscript, would help the reader follow the method. Now, I have these questions in mind: Can the method extrapolate beyond the envelope of observed days? Is the method validated and sound to maintain the daily to hourly extremes? Is there randomness in the selection or just RMSE (This might influence the sampling of extremes as a single extreme day can be selected more times, especially on the tail). Last, did you actually check if the WG produces close enough extreme statistics, for example GEV daily and hourly scale for some return period (20y?) in comparison to observation station data?

We have revised this section of the manuscript so that it includes the relevant information about the disaggregation method. Specifically, we have added the following text: *"Analog days are selected within the same intensity and seasonality classes by comparing the daily precipitation field of each simulated day with that of the period for which hourly observations are available (1990–2019). Several stations exist within a specific radius in the precipitation field, preserving the spatiotemporal structures. The best analogue day is then selected from a pool of 100 analogues that are retained for each day. The selection is performed based on the root mean square error."* We kept only the citation of Viviroli et al. 2022 in the manuscript. However, we provide all the necessary information in the next questions which are part of internal reports.

Now, I have these questions in mind: Can the method extrapolate beyond the envelope days? Is the method validated and sound to maintain the daily to hourly extremes? Is there randomness in the selection or just RMSE (This might influence the sampling of extremes as a single extreme day can be selected more times, especially on the tail)

Simulated daily intensities can exceed observed daily intensities because they are generated from the EGPD distribution. As a consequence, simulated hourly intensities can also exceed observed hourly intensities because the disaggregation only affects the sub-daily distribution of hourly intensities, not the daily amount.

The methods of course have some limitations. The first is related to the inadequacy between simulated and observed precipitation. As explained, the analog days are defined from the similarity between precipitation fields. Therefore, although there may be a positive simulated precipitation, the precipitation at a specific gauge on the analog day could be 0. In this case, the precipitation is disaggregated using a local analog, i.e. the day in the archive period with the point precipitation intensity closest to the simulated one. The downside of this solution is that it may deteriorate the temporal correlation of hourly precipitation series between gauges on a given day.
The second limitation arises when large to very large precipitation amounts are simulated. Beyond a certain threshold, extreme daily precipitation is expected to happen due to long-duration events. Nevertheless, due to the statistical nature of the disaggregation approach, it cannot guarantee such criteria. If either the highest observed precipitations at the gauge are much smaller than the simulated one or if the chosen analog day does not exhibit heavy precipitation at the gauge, it can happen that large to very large daily precipitation amounts are distributed across a few hours only. To limit the occurrence of such situations, a collection of "rescue" analog days was built that distinguish several intensity classes based on high quantiles.
However, those patches cannot prevent all the situations encountered, and specifically, the disaggregation of extremes precipitation can lead to hourly intensities whose physical realism is hard to judge. These limitations are expected to affect the simulated hydrology, specifically in the smallest catchments, where very high flows could be obtained. The impact of such situations should however remain limited at the scale of the large catchments.
As mentioned in the manuscript, for each sub-catchment, areal estimates are computed with the Thiessen Polygon method. For comparison purposes, the areal estimates directly from observed data were computed, using all the available stations during the period 2009-2013. Figure 10 shows that the 90th percentile of absolute deviations between areal estimates directly computed from hourly observations P_{1h} and areal estimates computed from daily observations disaggregated at an hourly scale $\widetilde{P_{1h}}$. We observed that the differences do not exceed 0.2mm/h for most of the catchments. For a few catchments, either daily or hourly gauges are distant and explain these discrepancies.

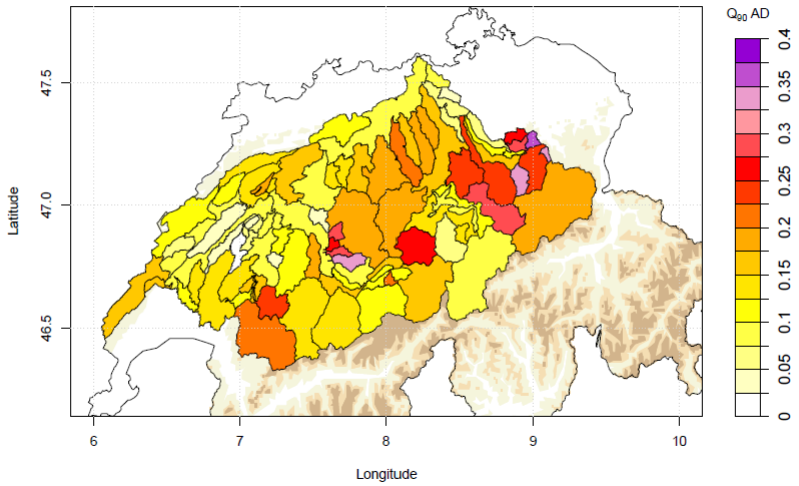

Figure 1 90th percentile of absolute deviations between hourly areal precipitation observations computed from the daily gauges (disaggregated at an hourly scale) and the hourly gauges.

Last, did you actually check if the WG produces close enough extreme statistics, for example GEV daily and hourly scale for some return period (20y?) in comparison to observation station data?

Yes, the method has been extensively validated and tested, e.g., in Evin et al. (2019). GWEX is designed to robustly reproduce the most critical properties of precipitation at different spatial and temporal scales with an emphasis on extreme precipitation amounts. We have added this paragraph at the end of Section 3.2: "*For the evaluation of the weather generator, Evin et al. (2019) used a multi-scale framework in space and time. The authors show that the estimation of all EGPD distribution parameters was performed robustly. They also conclude that the statistics of dry spell lengths, daily precipitation amounts, extreme precipitation amounts estimated from observations are, at all scales, very well reproduced by the generated scenarios*."

Additionally, we show exceedance plots where the annual maximum simulated and disaggregated observed precipitation for different stations are illustrated at an hourly, daily and 3-daily scale (Fig. 2, 3 and 4). We can see that small differences may occur but overall, for most of the stations, the extremes are captured well. However, there were expected over and underestimations especially at stations with uncertain observational records.

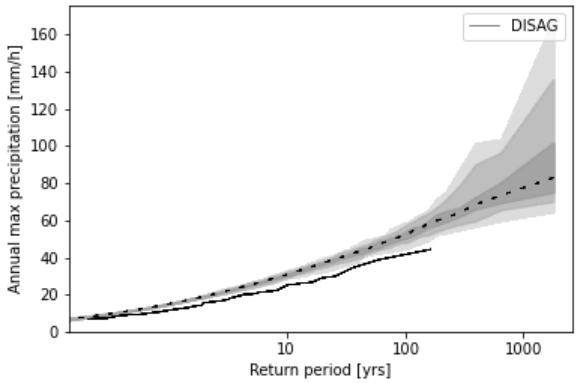

Figure 2 hourly precipitation sums at station in Frauenfeld in the Thur in Andelfingen River catchment. The black line shows observations, the gray areas show confidence intervals from the simulations (dark grey: 50%; medium grey: 90%; light grey: 99%). The x axis is return period [yrs], the y axis is precipitation [mm].

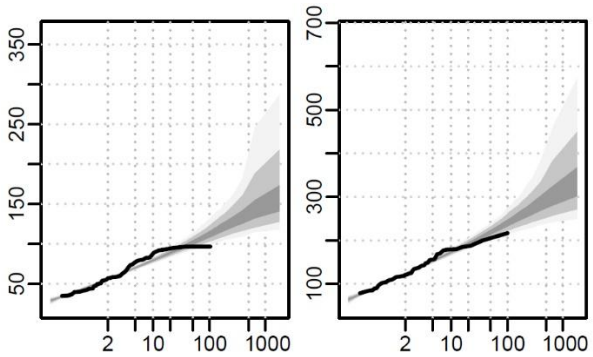

Figure 3 1-day (left) and 3-day (right) precipitation sums at station Kleine Schediegg in the Aare River catchment. The black line shows observations, the gray areas show confidence intervals from the simulations (dark grey: 50%; medium grey: 90%; light grey: 99%). The x axis is return period [yrs], the y axis is precipitation [mm].

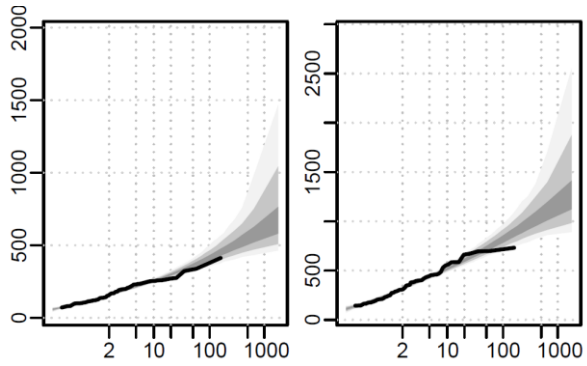

Figure 4 1-day (left) and 3-day (right) precipitation sums at station Mosogno in the Maggia River catchment. The black line shows observations, the gray areas show confidence intervals from the simulations (dark grey: 50%; medium grey: 90%; light grey: 99%). The x axis is return period [yrs], the y axis is precipitation [mm].

GC5: I somehow expected a clearer cross-analysis comparing whether PLRs matter more in the WG or in the Hydrological Model. The dual experiment presented is methodologically sound but the narrative in the manuscript requires the reader to connect the dots themselves. Starting section 4.1, the sensitivity

of mean annual precipitation to PLR_WG is essentially an expected outcome. Later sections, when authors focus on mixed signals of PLR_WG and PLR_Hydro, make it harder to disentangle which one actually drives the final results, especially when the hydrological model is calibrated with 5% PLR (Line 180). After that, section 4.7 shows that seasonality of floods is highly controlled by PLR_HBV and magnitude by PLR_WG. This is again a bit of an expected outcome, but overall I feel that the manuscript downplays the WG contribution. I would suggest reorder the results section because now it could raise a question of why they even bothered to do PLR_WG if the main factor is PLR_HBV (as it sounds like that in 4.7 after excessive analysis in 4.3, 4.4, 4.5), but the final decision is of course on the authors' shoulders.

We thank the reviewer for this thoughtful and constructive comment. We appreciate the suggestion to strengthen the cross-analysis of the relative contributions of PLR_WG and PLR_HBV, as well as the observation that the role of the weather generator may not be sufficiently emphasized in the current presentation.

Nevertheless, we would refrain from changing the structure of the Results section, as the current order provides a logical progression from the individual effects of PLR_WG to the combined effects of PLR_WG and PLR_HBV on the hydrological response and reveals the relative importance of each. We experimented with changing both PLRs because these two should be consistent in a modelling chain, although they represent two different things. The main focus of the paper is PLR_HBV as its effect is overlooked in most of the modeling studies.

GC6: To me, the flood frequency analysis is not clear. The authors didn't mention if AMFs are extracted from daily or hourly discharge series. This matters because each can have different behavior (seasonality too). All analysis is later performed as empirical and on medians I assume. But it raises the question, how certain are you with 30 scenarios for a 1000y return period? Does it follow some expected distribution? Do you have larger uncertainty CI intervals for higher return periods? This should be acknowledged when discussing the uncertainty results.

Section 3.4 “Flood frequency analysis” has been now moved to section 3.1 “Simulation workflow” to make the methodological steps clear from the beginning. Thus, in the first line of the last paragraph of Section 3.1 we clarify that “*Overall, hydrological simulations were performed at an hourly time scale for…*“ and later on the same paragraph that “*Then, for each scenario of 1000 years annual maximum floods (AMFs) were extracted from the hourly discharge series and a return period assigned*.” Lastly we mention that “*That led to 30 hourly estimates per return period up to 1000 years, which allows us to comprehensively compare flood estimates…*”

Indeed, the medians are based on the AMFs of the 30 scenarios. Not all analyses are performed using only the medians. As explained in the paper, we do compare the medians, but also the uncertainty range and the CIA (Fig. 10). We did not include the confidence bands because they would add visual clutter without providing substantial value to the interpretation of the results. In the uncertainty section, we are trying to explain the differences compared to the baseline. Indeed, this might not cover the questions of the reviewer. However, these questions were covered in detail in previous work (Kritidou et al., 2025, 2026). In these papers, we show that the uncertainty of extreme flood estimates is always increasing with return period. Additionally, when different sources of uncertainty are considered, their relative contribution is highly variable among the test catchments. We revised the manuscript so that in the second line of section 4.6 it is mentioned that “*Although the uncertainty of floods is generally increasing with return period, as discussed in previous work (Kritidou et al., 2025, 2026), this study focuses on the differences in medians, uncertainty ranges and confidence interval amplitudes (CIAs) between the AMFs….*”

GC7: The temperature lapse rates are calculated as daily climatology — does it include some uncertainty bounds and year by year fluctuations, or is it assuming that each day has this particular lapse rate based on mean? It would be nice to include a brief explanation in the manuscript. Additionally, it would be beneficial to mention that the temperature lapse rate can vary due to the diurnal cycle, although I understand this was outside the scope of your study.

The temperature lapse rates were estimated for 7 different regions covering all of Switzerland. For each region we have:

• Calendar lapse rates are estimated for each day of the year and each hour of the day (i.e., 24×366 calendar lapse rates).

• Hourly lapse rates are estimated for each single time step (hour) of each day of each year of the considered period (i.e. about 24x365x39 for the period 1981-01-01 to 2019-12-31). The hourly lapse rate for a given time step is then estimated from observations available for this time step (e.g. 01/01/2020 20:00).

The lapse rate is estimated for each hourly time step based on the temperature observations from 01.01.1981 to 31.12.2019, available around the region. For each catchment, the hourly time series of the lapse rate is based on either the hourly or calendar lapse rates.

For each hourly time step the typical linear regression with a 5-hour moving window based on hourly lapse rates was used. If the R^2 of the hourly regression is above 0.6 (the elevation/temperature relationship can be modelled adequately with a linear relationship) (Hingray et al., 2006; Winter et al., 2019) then the hourly lapse rate is retained. If not, the calendar lapse rate is considered. The calendar lapse rates are estimated for each hour and each day of the year using a 5-hour moving window and a 31-day moving window.

For all regions, the percentage of hourly lapse rates retained (i.e., the percentage of hourly R^2 greater than 0.6) is stronger during spring and summer than in autumn and winter, and during the day (12:00 and 18:00) than during the night (0:00 and 6:00) (Figure 5). The percentage of hourly lapse rates which have been retained is close to 100% during the day in Spring and Summer, whereas it is close to 30% or 40% (depending on the region) during the night in winter.

We added a brief explanation in section 3.2. Additionally, we slightly revised the sentence in section 3.3 referring to temperature lapse rates. The revised sentence reads "*For temperature, we always used the generated lapse rate time series (TCALT [°C/100 m]) for each calendar day, computed from temperature observations as explained in Section 3.2.*"

Specific Comments:

L34: Missing comma after citations in parentheses and missing full stop (dot) at the end of the sentence.

We revised accordingly.

L263: "ΔElevation is illustrated in Fig. 3b" — based on the figure, this should refer to Fig. 3a.

We changed "Fig. 3b" to "Fig. 3a".

L327–328: "While small but positive sensitivity coefficients occur in winter and spring, they are close to zero in summer for almost all catchments. That suggests that precipitation change does not drive summer snowmelt." This interpretation needs revision. Winter sensitivity coefficients range from roughly 0 to 0.75 and spring from 0 to 0.5; these are not "small." For summer, the trend appears to be pulled higher by a single point on the left side; without it, the signal would look similar to spring. The current wording risks overstating the contrast between seasons.

Thank you for this suggestion. The way it is explained might overstate the contrast between seasons and we have therefore revised the wording to *"Small but positive sensitivity coefficients occur in winter, spring and summer for almost all catchments. This could indicate that precipitation changes alone are not the main driver of summer snowmelt."*

Fig. 7e: The figure is somewhat confusing. Each catchment is represented by four points, colored by elevation, and the conclusion is drawn from the full sample of 108 points. Does the relationship hold for each catchment individually, or is it an artifact of aggregation? A supplementary version of the same figure with each catchment colored by PLR would help.

We agree that the original figure made it difficult to distinguish the relationship for each catchment. We also identified an inadvertent duplication of data points in the original plot, which resulted in more points being displayed than intended. This plotting error has been corrected and Figure 7e updated accordingly. This correction does not affect the correlation results or the main conclusions of the study.

In response to the reviewer's suggestion, we made a new version of the figure in which each catchment is colored differently. Instead of points, we plotted a line for each catchment. From the plot we can see that the relationship holds for the majority of the catchments while some catchments show different behavior. That is expected as shown in the manuscript: when a different lapse rate is applied in GWEX, the impact can be either very large or minimal.

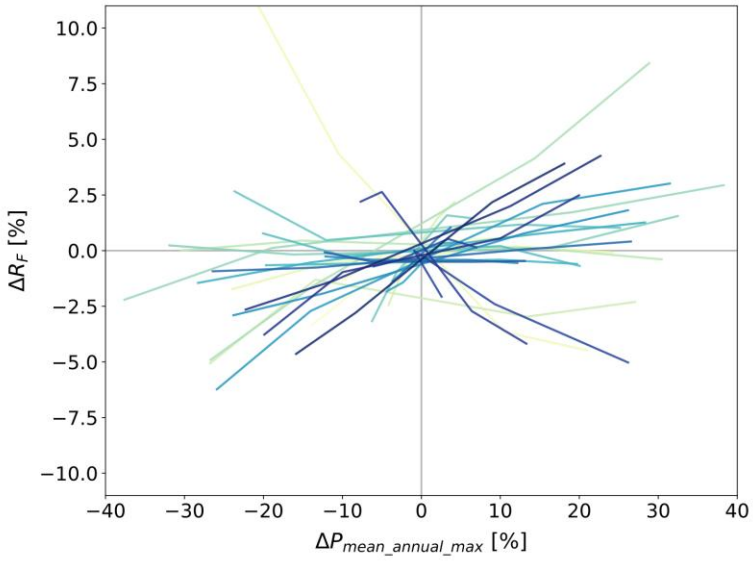

Figure 5 Relationship between the change in the flood seasonality ratio and the change in mean annual precipitation relative to the baseline. Each catchment is represented by one line consisting of four points and a different color.

Fig. 10: The title of subplot (b) should be moved slightly higher; it currently overlaps with the panel.

We updated the plot so that the title does not overlap.

L417: "range ts compared" — is "ts" meant to stand for something, or is it a typo?

We corrected the text so that it reads “is compared”.

L519: "This suggests that floods occurring in winter and early spring result from a combination of liquid precipitation and potential earlier snowmelt." The mechanism behind "earlier snowmelt" could be explained more explicitly. I guess lower PLRs lead to more snow falling at lower elevations, where it melts earlier; alternatively, wetter snow and a thinner snowpack at higher elevations could degrade more quickly. A short clarification of which mechanism the authors invoke would help.

Thank you for this comment. With higher lapse rates, we have more snow accumulation at higher elevations, which is retained and melts in spring. That is why the snowmelt contribution to floods is larger during summer. In contrast, smaller lapse rates lead to larger snowmelt contribution in winter. Precipitation amounts are smaller in higher elevations and the amounts falling in lower elevations leads to direct snowmelt as also explained by the reviewer. So, the earlier snowmelt can arrive already in winter because, for example with a 0% PLR, there is no redistribution of precipitation. The term “earlier snowmelt” is used to describe the snowmelt that occurs earlier in the year for smaller PLRs compared to larger PLRs.

L545: "PLR_GWEX" should be "PLR_WG" for consistency with the rest of the manuscript.

Thank you for pointing this out. We corrected PLR_GWEX to PLR_WG.

L616: Typo: "Czeck Republic" → "Czech Republic".

We corrected the typo in the manuscript.

Bibliography: Bertoncini and Pomeroy (2025) appears twice as two identical entries (2025a and 2025b). Please further check if this is the only one.

We updated the entries so that only one reference appears in the text and in the references.

Note that there might be more typos, so please double check those as well.

We read carefully through the manuscript and corrected typos.

References

Benoit, L., Koch, E., Peleg, N., and Mariethoz, G.: Precipitation-elevation relationship: Non-linearity and space–time variability prevail in the Swiss Alps, Journal of Hydrology X, 25, 100186, https://doi.org/10.1016/j.hydroa.2024.100186, 2024.

Farinotti, D., Usselmann, S., Huss, M., Bauder, A., and Funk, M.: Runoff evolution in the Swiss Alps: projections for selected high-alpine catchments based on *ENSEMBLES* scenarios, Hydrological Processes, 26, 1909–1924, https://doi.org/10.1002/hyp.8276, 2012.

Finger, D., Vis, M., Huss, M., and Seibert, J.: The value of multiple data set calibration versus model complexity for improving the performance of hydrological models in mountain catchments, Water Resources Research, 51, 1939–1958, https://doi.org/10.1002/2014WR015712, 2015.

Hingray, B., Mezghani, A., Schaefli, B., Niggli, M., Faivre, G., Guex, F., Hamdi, Y., Musy, A. (2006). Estimation des débits de crue du Rhône à Porte du Scex et autres points amont caractéristiques. Rapport final du projet CONSECRU 2. Laboratoire Hydrologie et Aménagements. EPFL, Lausanne, 83p. + 86p. Annexes.

Kritidou, E., Kauzlaric, M., Staudinger, M., Evin, G., Hingray, B., Vis, M., Seibert, J., and Viviroli, D.: Impact of different weather generator scenarios on extreme flood estimates in Switzerland, Stoch Environ Res Risk Assess, 39, 847–866, https://doi.org/10.1007/s00477-024-02843-8, 2025.

Kritidou, E., Kauzlaric, M., Staudinger, M., Evin, G., Hingray, B., Vis, M., and Viviroli, D.: Partitioning uncertainties of extreme flood estimates using long continuous simulations, Journal of Hydrology, 668, 134804, https://doi.org/10.1016/j.jhydrol.2025.134804, 2026.

Ménégoz, M., Valla, E., Jourdain, N. C., Blanchet, J., Beaumet, J., Wilhelm, B., Gallée, H., Fettweis, X., Morin, S., and Anquetin, S.: Contrasting seasonal changes in total and intense precipitation in the European Alps from 1903 to 2010, Hydrol. Earth Syst. Sci., 24, 5355–5377, https://doi.org/10.5194/hess-24-5355-2020, 2020.

Ruelland, D.: Should altitudinal gradients of temperature and precipitation inputs be inferred from key parameters in snow-hydrological models?, Hydrol. Earth Syst. Sci., 24, 2609–2632, https://doi.org/10.5194/hess-24-2609-2020, 2020.

Winter, B., Schneeberger, K., Dung, N. V., Huttenlau, M., Achleitner, S., Stötter, J., Merz, B., and Vorogushyn, S.: A continuous modelling approach for design flood estimation on sub-daily time scale, Hydrological Sciences Journal, 64, 539–554, https://doi.org/10.1080/02626667.2019.1593419, 2019.