Author Response to Reviews of

Climate model spread outweighs glacier model spread in 21st-century drought buffering projections

Ultee et al.
The Cryosphere

RC: *Reviewer Comment*,   AR: *Author Response*,   □ Manuscript text

The authors thank both reviewers for their consideration of our manuscript. Each reviewer provided insightful comments that will make the work stronger. We particularly appreciated several great suggestions on communicating the CMIP6 ensemble results.

In response to review comments, we have made changes including:

- Adding context on the SPEI method and the choice of drought metric

- Addressing the "hot model problem" with new analysis and practical suggestions for practitioners

- Adding supplementary tables with details on each glacier model and on the GCM ensembles

- Updating Figure 2 with grid, revised axes labels, and corrected color scheme to align with Figure 3

- Updating Figures 4 and 5 to show GEM-forcing ensemble spread and CMIP6 archive spread on the same panel, colored by basin glacier area fraction

- Correcting small code discrepancy, in which the 11 members of the glacier forcing ensemble were not included in CMIP6 ensemble statistics. Figures and text now reflect correct results for a full ensemble of 123 members (the 112 previously mentioned, plus the 11 used to force the glacier models now included as part of the full ensemble).

1. Reviewer 1

1.1. Summary

RC: *The manuscript evaluates how uncertainty in climate forcing (GCMs) and uncertainty across three global glacier models (GloGEM, PyGEM, OGGM) propagate to projections of glacial drought buffering for 75 large glacierized river basins worldwide. The authors compute two versions of a 3-month SPEI (one that includes glacier runoff) using CMIP6 forcings and monthly glacier runoff outputs. They summarize buffering as the difference in drought severity between SPEIs in three periods and compare the spread arising from the GCMs versus the spread across the three glacier models. Main findings: Glacial buffering generally increases through the 21st century and correlates with basin glacier fraction. Qualitative buffering trends are similar across the three glacier models. Inter-GCM spread in buffering substantially exceeds inter-glacier-model spread, and the commonly used 11-member glacier forcing ensemble undersamples the broader CMIP6 spread.*

This paper addresses an important and timely research question, building on previous work (Ultee et al., 2022; Wimberly et al., 2025). The manuscript is generally well-conceived, clearly written, and presents a

valuable approach, particularly in its use of a large multi-GCM and multi-realization ensemble. Before acceptance, some methodological clarifications, additional contextual details, and sensitivity analyses would further strengthen the manuscript. Below, I outline detailed comments and specific suggestions to guide the authors in refining their work.

1.2. Major comments

RC: *[Inclusion of "hot" models.] While reading the manuscript, I wondered whether sampling the entire CMIP6 spread is necessarily the most informative approach. Recent studies (e.g., Hausfather et al., 2022) have highlighted that several CMIP6 models exhibit unrealistically high climate sensitivities, leading to warming projections that exceed observationally constrained ranges. Consequently, many multi-model analyses now either exclude these "hot" models or reframe their results by grouping or averaging simulations according to warming levels. It would therefore be valuable for the authors to discuss how the inclusion of such "hot" models may influence their conclusions. Specifically, would the finding that inter-GCM spread outweighs inter-glacier-model spread remain robust if the analysis were repeated using a more balanced subset of GCMs?*

Hausfather, Z., Marvel, K., Schmidt, G. A., Nielsen-Gammon, J. W., & Zelinka, M. (2022). Climate simulations: recognize the 'hot model' problem. Nature, 605(7908), 26–29

AR: *The reviewer makes a great point. We have identified which models fall within the "likely" range of assessed climate sensitivity, as suggested by Hausfather et al. (2022), produced new versions of Figures 4 and 5 limited to those models, and added that information to the supplement. The results as originally reported are robust to this filtering. We address this in a new paragraph in the discussion section. Thank you for this suggestion!*

RC: *[GCM selection and ensemble representativeness.] The paper's central claim depends on comparing the 11-member ensemble used to force glacier models against the wider CMIP6 ensemble. The manuscript should explicitly list the 11 GCMs and briefly explain how these GCMs were selected (if there is a reason). In addition, provide a table of the 24 GCMs used for comparison/reproducibility purposes (including the number of realizations per GCM).*

AR: *Yes, absolutely. We have added these tables in the Appendix. Pursuant to the above point about "hot models", we have also indicated which models are within the central estimates of climate sensitivity. The 11 GCMs used to force the glacier models were chosen to be consistent with the current "standard" set used in glacier modelling, which we note in the Methods (line 72) and Discussion (line 224-227). To our knowledge, the regional paper that first used that ensemble is Compagno et al. 2022, which lists the GCMs only in tags to the supplementary data, and does not explain their selection criteria.*

RC: *[Details about downscaling / bias correction.] The Methods state that the glacier models were forced by a single continuous historical (2000–2014) + SSP (2015–2100) simulation per GCM. Please provide details on any bias correction/downscaling applied to GCM variables before feeding glacier models and computing SPEI.*

AR: *We have added a brief description of the method to the Methods section and to the new table of glacier models, Table B1. The three models all use similar methods but with slightly different choices of reanalysis product, baseline period, etc. – all detailed in the table.*

RC: *[Influence of precipitation factors.] Several glacier models apply precipitation correction factors to compensate for known underestimation biases in precipitation datasets. While such factors primarily affect mass balance calibration, they can significantly influence the magnitude of simulated runoff. Consequently,*

these corrections may amplify the apparent glacier buffering capacity. It was not entirely clear from the manuscript whether such precipitation correction factors were applied within the glacier model simulations used here, and if so, whether they differ across the three models. I recommend that the authors clarify this aspect in the Methods section. If corrections were applied, please specify their magnitude. If not, a brief discussion on the potential implications of uncorrected precipitation biases for the glacier runoff component would strengthen the paper.

AR: *Yes, the reviewer is correct. All of the glacier runoff simulations in our analysis include precipitation correction factors and the factors may well differ by glacier model. In fact, Wimberly et al 2025 deduced that the precipitation factor is the primary source of glacier model differences in simulated runoff. Unfortunately, those precipitation correction factors are not saved with the glacier model runs, which makes them difficult to analyse directly. We have revised the Methods section as follows:*

> *For the GCM grid point closest to the glacier, the method calculates an additive (for temperature) and multiplicative (for precipitation) monthly bias between the GCM data and a reanalysis product, which is assumed to be constant in time. The interannual variability of the GCM temperature series is further adjusted to be consistent with the reanalysis data. **The local precipitation factor is further calibrated in each glacier model's mass balance module to produce mass change consistent with satellite observations over the period 2000-2020. All models calibrate to the same observed mass change dataset of Hugonnet et al. (2021)**.*
>
> *The three glacier models provide glacier-by-glacier projections of glacier volume and fixed-gauge water runoff, with monthly resolution, for the period 2000-2100. **Glacier model differences, especially in calibration of the precipitation factor, can lead to significant discrepancies in simulated runoff. We have provided a more complete analysis of the factors affecting glacier runoff output in Wimberly et al. (2025)**.*

The effect of precipitation scaling on SPEI was addressed in our earlier work, Ultee, Coats & Mackay 2022, and found to be distinct from the effect of including runoff from a glacier evolution model. We have made note of this in the discussion.

RC: *[Glacial runoff definition.] In L90, glacial runoff is defined as the sum of ice and rain runoff from glacierized areas. However, because the three glacier models use fixed-gauge water runoff, this definition likely also includes seasonal snowmelt. This is critical, as the study's core argument attributes all runoff from glacierized regions to glacier change, while part of it may originate from seasonal snowmelt. This distinction has been highlighted in recent discussions (Gascoin, 2024), stressing the need for precise terminology when describing "glacial runoff." If possible, the analysis should isolate ice melt to maintain conceptual consistency with drought buffering through glacier mass loss. Otherwise, the authors should clearly acknowledge this limitation and the potential overestimation of glacier influence.*

AR: *Yes, good point. For our large-scale drought analysis, with modification to SPEI moisture source term, we need fixed gauge runoff. The three glacier models we analyse do include snowmelt in their "melt" term, and we have revised Methods section 2.2 to make that explicit. Unfortunately, none of the three provides partitioned output that tracks ice and snow melt separately, so it is not possible to isolate ice melt. This is an area of development in glacier modelling. We have updated the Discussion to clarify that our approach may overestimate the role of glaciers rather than snow in the future basin configuration.*

RC: *[Use of multiple realizations per GCM.] To my understanding, this is the first study to use such a large archive of future simulations (>100), incorporating multiple models and realizations. This is a valuable*

and novel aspect, as most studies rely on single realizations per GCM. While perhaps beyond the current scope (L127–128), it would be interesting to include some discussion on how representative ensembles with only one realization per GCM compare to this larger multi-realization approach. Providing brief insights on this point would offer a useful lesson for modelers and practitioners regarding ensemble design.

AR: *This is an interesting point, though we agree that fully addressing it is beyond the current scope of the manuscript. We have added to the Discussion to clarify that single-realization simulations will under-sample internal variability, and we explain why that is of interest for glacier modelling. In the new Table C2, summarizing the CMIP6 archive, we now note how many realizations are included for each model.*

The most novel advance would be to run all of these realizations through the glacier models and compare results forced by a single realization against those forced by multiple realizations of the same model. This would quantify to what extent under-sampled internal variability affects runoff outputs. However, as we note in the manuscript, this task would be a far greater computational demand than is currently practical for global glacier modelling. We choose not to pursue it further here, but we hope that we or other groups will have the opportunity to try it in follow-on studies.

1.3. Specific comments

RC: *Title: Consider adding "CMIP6" to make the forcing context explicit.*

AR: *Yes, absolutely. Revised to "CMIP6 climate model spread outweighs glacier model spread in 21st-century drought buffering projections".*

RC: *Abstract: Good summary. Suggest explicitly stating the number of GCMs (11 vs. 24) and glacier models (3). Please also consider briefly mentioning the scenario used.*

AR: *Thank you. We have added mention of the number of GCMs in the glacier model ensemble and in the broader CMIP6 sample, and we added mention of the SSP scenario. The number of glacier models is already listed.*

RC: *Abstract: "...likely under-sampled." Under-sampled compared to what? The reader currently needs to read the Methods to understand this.*

AR: *Revised to specify:*

> *...under-sampled compared with the full archive of suitable CMIP6 simulations (123 simulations from 28 models for the SSP2-4.5 scenario).*

RC: *L41: Clarify that the initial increase is only expected in some cases, as the "peak water" has already been reached in many regions*

AR: *Revised to*

> *In some basins, the phase of increasing glacial runoff has already passed, with no further increase projected in the 21st century.*

We avoid saying that runoff is projected to strictly "decline", because precipitation changes may compensate for the declining glacial runoff.

RC: *Study areas: This section reads more like a description of the figures rather than an introduction to the*

4

study areas. Include more information about the basins themselves, such as their geographic distribution, climatic or hydrological diversity, and relevance to the research question.

AR: *Yes, good point. We have added the requested information to section 2.1. We have also kept the description of the figure organization, which was requested by the editor in pre-review.*

RC: ***Study areas: Please briefly justify the >3000 km² and >30 km² thresholds.***

AR: *We have added this information to the text. These thresholds were applied in Holmgren (2022) to compute peak water and glacial drought buffering from OGGM simulations. They select large-scale basins that can be studied with GCM output (too coarse for smaller basins) and have enough glacier cover to produce non-trivial runoff. The thresholds are similar to those for "macro-scale glacierized basins" in Huss & Hock (2018) and Ultee, Coats, and Mackay (2022), with a lower minimum basin area. Our 75 study basins include all those studied in the earlier works, as well as 19 slightly smaller ones. The added basins include some that are significant for hydropower potential (e.g. Palena, Pascua), irrigation (Huasco), ecological importance (Har Us Nuur), or cross-border water management questions (Chuy).*

RC: ***Glacier model descriptions: Include a concise summary (or supplementary table) outlining key information for each glacier model, such as version used, variables used, main structural differences, and calibration approach. While the manuscript references Wimberly et al. (2025), providing this overview here would greatly improve readability and help interpret potential differences in the results.***

AR: *We have added brief context in the Methods section and added Table B1 to summarize glacier model details. Good suggestion, thanks.*

RC: ***Historical climate: Does the single continuous historical run correspond to the 'historical' experiment from each GCM, or is it derived from a reanalysis dataset? Please clarify.***

AR: *PyGEM and OGGM simulations are forced by the historical experiment from each GCM. GloGEM simulations use ERA5 forcing. We have added this detail to the table in the appendix and revised the Discussion to comment on it. We retain these differences in historical forcing because we have chosen to use the standard setup of each model, as a user would find in off-the-shelf data products. The philosophy is elaborated in the first paragraph of the discussion.*

RC: ***SPEI selection: The argument for using SPEI could be moderated to better reflect its scope. SPEI only accounts for precipitation and potential evapotranspiration, but does not explicitly represent other catchment hydrological processes.***

AR: *We have addressed the limitations of SPEI in the discussion.*

RC: ***SPEI computation: Potential evapotranspiration is calculated using the Penman-Monteith equation. Please indicate the variables used here rather than in Section 2.4 Model Spread.***

AR: *We have updated the description of the SPEI computation and necessary input variables as requested.*

RC: ***SPEI time window: The authors use a 3-month SPEI and define droughts as SPEI ≤ -1. Please justify: (a) the choice of a 3-month accumulation window – why not 6 or 12 months, given that droughts typically develop over longer periods? (b) the use of the 1900–1979 period for standardization. Shouldn't this align with the period used for historical glacier simulations or with the baseline applied in bias correction of the climate projections?***

AR: *(a) We used a 3-month integration to reflect streamflow drought, in line with our own previous work on SPEI buffering and with other previous applications of SPEI (e.g. López-Moreno et al., 2013). In Ultee, Coats*

& Mackay 2022 we tested integration periods of 3-27 months and found qualitatively similar results. The interquartile range of the GCM ensemble was larger for longer integration times.

(b) The standardization period should be as long as practical to ensure that the relevant modes of climate variability are captured. The 20-year "historical" period of the glacier simulations is too short for this standardization. We maintained the same period chosen in our earlier glacial drought buffering study to facilitate comparison; the "historical" glacier simulations in that study started in 1980, hence the 1979 cutoff of the standardizaton period. The bias correction period for GCM data differs per glacier model, and is generally also shorter than desired for the SPEI standardization period.

RC: *Glacier area for SPEI: Please clarify how glacier area is treated. By "initial glacierized area of the basin," do you refer to the year 2000, 2014, or the year reported in the RGI dataset?*

AR: *The initially glacierized area is the RGIv6 area for PyGEM and GloGEM, and the year-2000 area for OGGM. We have added this information to the supplementary table requested.*

RC: *Figures 2 vs. 3: Please check that the color schemes for the different models are consistent. Also, use consistent names/units for glacial drought buffering.*

AR: *Wow, good catch, thank you! The figures are all scripted, but the order of the color scheme was inadvertently changed between the files generating Figure 2 versus Figure 3. Corrected color scheme and updated axes labels in Figure 2 to match Figure 3.*

RC: *Section 3.1: "Expressed in terms of reduced number of droughts"—this is only shown in terms of severity in Figure 2.*

AR: *Another good catch. Removed the mention of number of droughts. The results do hold for number of droughts, but we elected to focus on drought severity to reduce excess detail in the manuscript.*

RC: *Figure 2: Why is it Δ^2 in the axis label?*

AR: *We have changed the axes labels to be more clear. Figure 2a is drought buffering, expressed as a change in drought severity (Δ SPEI deficit). Figure 2b is the change in that metric between two time periods ($\Delta\Delta$ SPEI deficit, or Δ^2 SPEI deficit). We updated the axes label in 2b to "Δ buffering, end C. vs. early C." and we explain further in the caption.*

RC: *Figures 2–3: Please remind the reader that the values correspond to the difference between SPEI(N) and SPEI(g). Also, indicate that the whiskers correspond to the min–max GCM range.*

AR: *We have added these details to the captions.*

RC: *Figures 4–5: Consider replacing points/circles with a bar plot, adding whiskers to represent the spread among GCMs. Colors could indicate glacier area (as in Fig. 1) or mean annual precipitation/temperature. Combining panels a and b to extend horizontal space may help, while panel c could show aggregated results across all basins.*

AR: *Thank you for these excellent suggestions. We have produced new versions of Figure 4 and 5 with previous panels a/b combined into a single panel, illustrating GCM spread with whiskers and GEM forcing ensemble spread with thicker bars. These are now colored according to glacier area fraction, consistent with Fig 1.*

RC: *Sentences like "results demonstrate conclusively that inter-GCM spread outweighs inter-glacier-model spread" should be tempered to reflect the conditional nature of the findings, acknowledging potential limitations (parameter or initialization uncertainty, calibration choices, data sources, etc.).*

AR: *Yes, thanks. We've done our best to provide this context throughout, and we hope the new appendix tables of model ensemble details will help interpretation too.*

RC: ***Year ranges: Ensure consistency—some text uses 2080–2100, some captions 2081–2100.***

AR: *All have now been corrected to 2080-2100, and we have confirmed this is reflected in the code. Thanks!*

RC: ***Figure 5: Please add "relative to 2000–2020" as in Figure 4 (axis label).***

AR: *Added.*

RC: ***L205–206: I don't fully agree with this statement. For hydrological applications, absolute runoff is important for potential model coupling. Therefore, differences among glacier models may be directly relevant, especially when no normalization is applied, as is the case for many metrics.***

AR: *Yes, good point. We have revised to specify that single glacier model output can be safely used for normalized drought metrics, but that studies requiring absolute runoff should either use multiple models or seek observational constraints on a single model.*

RC: ***Discussion: The section is thoughtful and balanced. Please add a short, practical recommendation for users. For example, if only one glacier model can be used (very common), which steps should be prioritized—choosing a glacier model suited to the region, increasing the sample size of climate models, or selecting GCMs based on their skill in representing regional teleconnections?***

AR: *This is a very reasonable suggestion, but we don't know the order of priority, and we suspect it varies by region. Clearly prioritizing those choices would require uncertainty quantification we have not been able to do in the current work. We have provided brief guidance that specifies the key considerations in simulation design, and how users can work with these unquantified uncertainties.*

2. Reviewer 2

2.1. Summary

RC: *The authors investigate the global drought buffering capacity of glaciers throughout the 21st century as simulated by an ensemble of glacier evolution and global circulation models (GEM resp. GCM). Their goal is not to obtain actual drought buffering estimates, but rather to analyze how drought buffering estimates differ among different GEMs and GCMs, and which of these two model types represents the largest source of uncertainty (expressed in the width of the ensemble spread). Across 75 basins, 3 GEMs and 11 GCMs, they compute a 3-month drought index (SPEI) and define the drought buffering capacity as the difference in drought index between GCM projections with and without GEM coupling. Their conclusions are three-fold: 1) in accordance with previous work, drought buffering increases with increasing basin glacier fraction, 2) drought buffering estimates are much more sensitive to GCM model choice than to GEM model choice, and 3) the ensemble spread of drought buffering estimates would likely be larger had the authors expanded the GCM ensemble to include the full 112-member CMIP6 climate product. The paper is well-structured, well-written and concise.*

The methodology and scope are similar to a previous study by some of the same authors (Ultee et al., 2022). In that study, they looked at a single GEM forced by 8 GCMs and analyzed the actual drought buffering capacity of glaciers worldwide. The current paper does not present large novelty in methodology compared to the previous paper, it essentially repeats the previous study with more GEMs and GCMs and across a larger number of basins, and compares the results from the different model combinations. However, the topic and the results of the work are highly relevant for both the scientific community and the general public. While the previous study presented the community with evidence of the end-of-century drought buffering capacity of glaciers worldwide, the present paper shows the relative importance of GEMs and GCMs in making global drought projections, a highly relevant matter for which a large number of stakeholders count on the scientific community. Despite the limited methodological novelty, I would therefore recommend publication after incorporating the following comments.

2.2. Major comment (potentially)

RC: *The SPEI is computed using PET estimates, which are said to be computed using the Penman-Monteith equation (L83). It is not clear if the computations are done by the authors themselves or if they originate from each of the GCMs. In case they are done by the authors, further justification of this particular PET model is necessary. Many different models exist, and their estimates can vary such that it is often recommended using multiple PET models to account for their uncertainty (Vremec et al., 2024). If there is no solid evidence in the literature that the Penman-Monteith model can safely be used as a single, global PET model, then the paper would highly benefit from a multi-model PET ensemble alongside the multi-model GEM and GCM ensembles. This will allow quantification of the uncertainty coming from PET model choice. If the uncertainty ends up being small, then a small addition to the discussion chapter could be sufficient.*

AR: *We have added clarification stating that we computed the PET using a method standard within the large-scale drought community. Penman-Monteith is the method most commonly used by the large-scale drought community (e.g. Cook, Ault & Smerdon, 2015), and leverages variables that are available from most GCMs. Importantly for our purposes, Penman-Monteith PET also uses the widest range of input variables and should be among the most physically realistic of the PET estimates (see Table 1 of Vremec et al., 2024). The paper defining SPEI explains that the method used to calculate PET is not critical (Vicente-Serrano et al., 2010), and they further refer to a study demonstrating that the choice of different PET methods does not make much*

difference in statistical drought indices such as PDSI and SPEI. We have seen this ourselves in earlier work, when we were encouraged to modify the PET calculation for varying stomatal conductance in a changing climate; we did so, and that correction is now standard in our method, but it made very little difference to the glacial drought buffering results.

2.3. Minor comments

RC: *A minor point of revision concerns the differences in glacier runoff between the different GEMs. The result that GEM uncertainty is much smaller than GCM uncertainty is presented as surprising because the authors knew the runoff differences between GEMs to be considerable. However, the reader is not presented with any evidence of this, except a reference to Wimberley et al. 2025. It would be good for the reader to be at least given an idea of how large these differences are.*

AR: Good point! We have revised to specify the magnitude of differences in runoff between GEMs, and we have added more context about the GCM uncertainty so that the result can be better interpreted on its own.

RC: *A brief discussion on the choice of the SPEI as a drought index is missing. In case other drought indices exist, the reader might also wonder how the choice of a different drought index would change the results and the conclusions of the manuscript.*

AR: Thank you for pointing this out. We have added context on the SPEI and the (lack of) influence of drought metric choice on our conclusions.

2.4. Specific comments

RC: *L2-4: I would argue there is inherent value in global projections (of glacier drought or other), independent of local-scale relevance. In any case, this fits better in the introduction than as the main motivation for the manuscript in the abstract.*

AR: No action taken for now. We have stated our motivation, prioritizing a specific application (informing local water management) rather than resting on an assertion of inherent value that readers might not agree is there.

RC: *L5: omit "latest generation", and potentially "CMIP6".*

AR: Kept these in. We agree with Reviewer 1 that it's helpful to specifically note CMIP6 as the source of forcing in our analysis. CMIP6 is currently the latest generation available.

RC: *L6 "despite differences in absolute glacier runoff": the authors don't show this in the manuscript. I advise the authors to find a way to explain the results without mentioning this explicitly.*

AR: Removed this from the abstract. Inter-model differences in runoff are now noted with more concrete detail in the Discussion.

RC: *L18: isn't this the case for all classes of climate models?*

AR: No, not really. For example, a handful of regional climate models including RACMO and MAR specialize in ice-atmosphere dynamics in glaciated regions (e.g. Noël et al., 2022), but these are regional models, not focused on global simulations. There are also global-scale models with components that have been used in glaciated regions, such as the Joint UK Land Environment Simulator (JULES), which has been coupled with an OGGM to simulate glacier runoff at regional scale (Mackay et al. 2025). However, that configuration does not participate in the Coupled Model Intercomparison Project experiments, i.e. it is not a CMIP-class model.

For now, no action taken on this point.

RC: *L21: Not a good reference for this statement. See (Hanus et al., 2024) : "Large-scale hydrological models mostly lack glacier representation. Only one out of 16 models used to simulate water availability in the Inter-Sectoral Impact Model Intercomparison Project phase 2 (ISIMIP2) had some representation of mountain glaciers, namely CWatM (Telteu et al., 2021). However, even this model has a very simplistic glacier representation resembling a snow redistribution method to avoid snow accumulation (Burek et al., 2020; Telteu et al., 2021). "*

AR: *Added citation to Hanus et al. (2024), as suggested by editor.*

RC: *L26: what about global scale analyses?*

AR: *Not to our knowledge, no. These two sentences are correct as written.*

RC: *L28: replace "possible" with "feasible"*

AR: *Done.*

RC: *L32: specify "settings"*

AR: *Replaced with "places". The Stein et al (2024) paper reports on disparities in country-level hazard exposure (population affected) versus research attention and resources.*

RC: *L33-35: link with previous sentences not sufficiently smooth*

AR: *We believe that the previous change, replacing "settings" with "places", makes the transition clear: we mention what is not currently possible, we highlight where it's a concern, and we offer an alternative.*

RC: *L44: SPEI requires better introduction*

AR: *Yes, good point. Added short paragraph to better introduce SPEI and its use to quantify glacier drought buffering.*

RC: *L45: add "state-of-the-art" in front of "climate models"*

AR: *Added.*

RC: *Figure 1: North American, Icelandic, Patagonian and New Zealand basins are not distinguishable, perhaps zoom-boxes would increase their visibility. Caption: Which Appendix?*

AR: *We have added a reference in the caption to Appendix A, where all regions are presented in a more zoomed-in view.*

RC: *L55-60: This paragraph does not benefit the storyline. These instructions can simply be stated in the figure captions.*

AR: *We have retained the information here as it was requested by the editor before review.*

RC: *L65: Explain which scenario this is*

AR: *Added context as requested.*

RC: *L69: questionable use of "we"*

AR: *Revised to:*

> *Wimberly et al. (2025) provides a more complete analysis of the factors affecting glacier runoff output in these simulations.*

RC: *L76: there is no section 2.3.2, so there is perhaps no need for a sub-section*

AR: *Hah, good point! Removed the subsection header.*

RC: *L82: Why the letter D?*

AR: *Standard in the definition of SPEI (Vicente-Serrano et al 2010; 2025), and we don't see a reason to propose new notation. It could stand for "difference" as the SPEI was introduced using a difference $(P - PET)$ as the metric to standardize, as contrasted with the Standardized Precipitation Index that standardized precipitation values only.*

RC: *L97-98: omit quotation marks*

AR: *Kept quotation marks here, in line with elsewhere in the Methods section where we define quantities analysed in the Results.*

RC: *L98-L99: Unclear sentence, rephrase*

AR: *Rephrased to:*

> *Following standard operational practice (e.g. Danandeh Mehr and Vaheddoost, 2020), we identify a "drought" as any period of continuously negative SPEI during which the SPEI reaches a threshold value of -1 or lower for at least one month. The "severity" of a given drought is the cumulative sum of SPEI during that period.*

RC: *L107: complicated sentence*

AR: *Revised to remove "in the [metric] we analyse", which could have been read as an extra subject and verb. Now the sentence should read more clearly as a list.*

RC: *L107-109: is there no parameter uncertainty in GCMs?*

AR: *Parameter uncertainty in GCMs certainly exists, but would be analysed with a process-oriented study in a single GCM. When analysing a GCM ensemble, we include parameter uncertainty in GCMs within the structural uncertainty among GCMs. We write that the sources of uncertainty "include" the list items. The list is not meant to be exhaustive; as the reviewer noted above, the sentence is complicated already.*

RC: *L138-139: elaborate on the theoretical understanding of hydrological trade-offs*

AR: *Revised to specify the glacier compensation effect, which is optimized at moderate glacier cover. As glacier cover increases beyond the optimum, the basin hydrologic regime becomes melt-dominated rather than balancing melt and precipitation, and the coefficient of variation increases again.*

RC: *Figure 2: increase figure width and apply a (semi-transparent) grid to both sub-figures. In the top figure it is difficult to see which points represent the same glacier*

AR: *Updated figure as requested. We have also clarified in the caption that the markers represent river basins, rather than individual glaciers.*

RC: *L193: unclear*

AR: *Revised to:*

> *Our objective here is rather to provide interpretive guidance for use cases where model uncertainty is important but may not be directly quantified.*

RC: *L206: "that appear large to glaciologists" needs reference or elaboration*

AR: *Revised this paragraph and the previous to be more concrete. Text now reads:*

> *Our results demonstrate explicitly that **large absolute differences in runoff (up to a factor of three; see Wimberly et al. 2025) among glacier models may be relatively unimportant for downstream drought metrics.** These findings imply that any of the three glacier models chosen would be appropriate for regional-scale drought analysis. In a regional-scale drought study conducted with limited computational resources, it may not be necessary to sample an ensemble of different glacier models. However, hydrological studies that require absolute glacier runoff should either sample multiple glacier models or seek observational constraints on runoff to adjust a single glacier model.*

RC: *L216-217: "climate scenario" not the right terminology*

AR: *No action taken here. Our original wording was "emissions scenario", as used by Riahi et al. (2017) in the definition of the Shared Socioeconomic Pathways. But, as discussed with the editor prior to review, the SSPs are not actually emissions scenarios; rather, they are socioeconomic pathways that could produce different combinations of emissions and resulting radiative forcing on the climate system, i.e. "climate [change] scenarios". We changed it to "climate scenarios" to be precise.*

RC: *L236: are there other global glaciers models? They have not been mentioned in the manuscript as far as I am aware*

AR: *We have updated the Methods, section 2.2, to clarify that there are no other global glacier models providing per-glacier runoff projections. Among the models participating in GlacierMIP3 (Zekollari et al. 2025, Supplementary Table S2), OGGM, PyGEM, and GloGEM are the only ones simulating all regions on a per-glacier basis and providing runoff output. Other large-scale models simulate a limited regional domain, do not resolve individual glaciers, or simulate only glacier mass change without simulating runoff.*

RC: *L236: replace "most-up-to-date" with "state-of-the-art"*

AR: *Done.*

RC: *L240-241: That depends on which downstream impact studies. This conclusion is specific to drought buffering in large glacierized basins.*

AR: *Yes, good point. We have revised the Discussion (see above) and the Conclusions to be more precise on this point.*

References

[1] Compagno, L., Huss, M., Miles, E. S., McCarthy, M. J., Zekollari, H., Dehecq, A., Pellicciotti, F., and Farinotti, D.: Modelling supraglacial debris-cover evolution from the single-glacier to the regional scale: an application to High Mountain Asia, The Cryosphere, 16, 1697–1718, doi:10.5194/tc-16-1697-2022, 2022

[2] Benjamin I. Cook et al., Unprecedented 21st century drought risk in the American Southwest and Central Plains. Sci. Adv.1, e1400082 (2015). DOI:10.1126/sciadv.1400082

[3] Huss, M., Hock, R. Global-scale hydrological response to future glacier mass loss. Nature Clim Change 8, 135–140 (2018). https://doi.org/10.1038/s41558-017-0049-x

[4] Mackay, J. D., Barrand, N. E., Hannah, D. M., Potter, E., Montoya, N., and Buytaert, W.: Physically based modelling of glacier evolution under climate change in the tropical Andes, The Cryosphere, 19, 685–712, doi:10.5194/tc-19-685-2025, 2025.

[5] Noël, B., Aðalgeirsdóttir, G., Pálsson, F., Wouters, B., Lhermitte, S., Haacker, J. M., & van den Broeke, M. R. (2022). North Atlantic cooling is slowing down mass loss of Icelandic glaciers. Geophysical Research Letters, 49, e2021GL095697. https://doi.org/10.1029/2021GL095697

[6] Keywan Riahi, Detlef P. van Vuuren, Elmar Kriegler, Jae Edmonds, Brian C. O'Neill, Shinichiro Fujimori, Nico Bauer, Katherine Calvin, Rob Dellink, Oliver Fricko, Wolfgang Lutz, Alexander Popp, Jesus Crespo Cuaresma, Samir KC, Marian Leimbach, Leiwen Jiang, Tom Kram, Shilpa Rao, Johannes Emmerling, Kristie Ebi, Tomoko Hasegawa, Petr Havlik, Florian Humpenöder, Lara Aleluia Da Silva, Steve Smith, Elke Stehfest, Valentina Bosetti, Jiyong Eom, David Gernaat, Toshihiko Masui, Joeri Rogelj, Jessica Strefler, Laurent Drouet, Volker Krey, Gunnar Luderer, Mathijs Harmsen, Kiyoshi Takahashi, Lavinia Baumstark, Jonathan C. Doelman, Mikiko Kainuma, Zbigniew Klimont, Giacomo Marangoni, Hermann Lotze-Campen, Michael Obersteiner, Andrzej Tabeau, Massimo Tavoni (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. Global Environmental Change, Volume 42, p153-168, https://doi.org/10.1016/j.gloenvcha.2016.05.009

[7] Ultee, L., Coats, S., and Mackay, J.: Glacial runoff buffers droughts through the 21st century. Earth System Dynamics, 13, 935–959, https://doi.org/10.5194/esd-13-935-2022, 2022.

[8] Vicente-Serrano, S. M., S. Beguería, and J. I. López-Moreno, 2010: A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. J. Climate, 23, 1696–1718, doi:10.1175/2009JCLI2909.1.

[9] Vicente-Serrano, Sergio M. & National Center for Atmospheric Research Staff (Eds). Last modified 2025-04-29 "The Climate Data Guide: Standardized Precipitation Evapotranspiration Index (SPEI)." Retrieved from https://climatedataguide.ucar.edu/climate-data/standardized-precipitation-evapotranspiration-index-spei on 2025-12-09.

[10] Wiersma, P., Aerts, J., Zekollari, H., Hrachowitz, M., Drost, N., Huss, M., Sutanudjaja, E. H., and Hut, R.: Coupling a global glacier model to a global hydrological model prevents underestimation of glacier runoff, Hydrol. Earth Syst. Sci., 26, 5971–5986, doi:10.5194/hess-26-5971-2022, 2022.

[11] Zekollari, H. et al., Glacier preservation doubled by limiting warming to 1.5°C versus 2.7°C. Science 388, 979-983 (2025). DOI:10.1126/science.adu4675