Anonymous Referee #2

In this manuscript the authors quantify the effects of historical and future glacial and climatic conditions on simulated glacier melt in the Patagonian Andes using the Open Global Glacier Model (OGGM) and Random Forest Regression approach. They focus on six categories, or sources of uncertainty, namely glacier outline inventories, ice thickness, historical climate, GCMs, bias correction methods, and emission scenarios. They examine the importance of each source on ten runoff metrics (e.g., peak water year and magnitude, interannual variability, seasonal contribution and variability, etc.) and conclude that the choice of reference climate is the predominant source of uncertainty.

The authors have undertaken a fairly comprehensive assessment, exploring 1920 cases across the six categories, to identify how the choice of input and forcing data affects the outcome in a glacialhydrological modeling workflow. This is overall a well-written paper, and the figures support the narrative well. The authors have demonstrated the methods clearly and the results are presented in a logical way.

The areas which require considerable improvement are the Discussion and Results sections which lack coherency and do not tie the various pieces together. There are differences across the hydrological zones/catchments and the ten glacier runoff metrics which require more nuanced discussion. Further, the authors need to explicitly highlight the implication/s of this work and explain if/how the findings here can be used for other glaciated regions of the world (i.e., does this study domain encompass a comprehensive set of climatic, glacial, and hydrological conditions to make general deductions).

I recommend moderate revisions before the manuscript is ready for acceptance. These revisions mostly require either clarification or further elaboration in the text. Comments below are provided section-wise and are not in order of importance.

R: Thank you for your thorough review of our manuscript. We appreciate your positive evaluation and constructive feedback. We take note of your suggestions for improvement, in particular to improve the coherence of the Discussion and Results sections. We are committed to making the necessary revisions to address these concerns and provide a clearer and more comprehensive understanding of our findings. In addition, we will explicitly highlight the implications of our work and discuss how the results can be applied to other glaciated regions worldwide. Please see our detailed response on how we plan to restructure the Discussion section in the related comment below.

Introduction

Ln 61-62: Can you clarify the statements related to "*significantly increased streamflow*" and "*significant trends*". I presume the latter refers to statistical significance and the former to a large magnitude increase? Also linking with the previous statement, is it possible to extract information from these past work on how much (%) the streamflow has increased because of the accelerated glacial mass loss?

R: Thanks for the opportunity to clarify these points. As you mentioned, the "significant" refers to the existence of a statistical significance. This will be clarified in the text (only Pasquini et al. 2021 used a

statistical test): "recent studies have reported increased flows in rivers with important glacier cover (Masiokas et al., 2019; Vries et al., 2023), some of which have only begun to show significant trends ($p < 0.001$) in the last decade (e.g., Santa Cruz; Pasquini et al., 2021)"

To our knowledge, there have been no comprehensive studies to date that would allow us to determine the percentage of streamflow that has increased as a result of accelerated glacier mass loss. We are addressing these important questions in an ongoing study that we expect to be available as a preprint this autumn (March - April).

Ln 67-68: Is this referring to overestimation of precipitation? Also, what does "*diverged towards*" mean?

R: Yes, this will be clarified in the text: "the different approaches and data sources have overestimated the precipitation according to numerical simulations of regional moisture fluxes."

Ln 70 onward: I am somewhat unsure about the purpose of this table (and generally this paragraph). The table summarizes past mass balance assessments; however, the current study is not considering various MB schemes or calibration processes as a source of uncertainty. This table seems superfluous and if the authors want to keep it, please consider moving to the supplementary material. Instead, please expand on the two Patagonian glacier hydrological contribution studies (*Mernild 2017* and *Caro 2023*). That will link better with the preceding paragraph and the overall theme of the paper.

R: Thanks for the suggestion. Table 1 will be moved to the Supplementary Material, and we will add a few sentences about the two previous hydrological studies as suggested.

Ln 89: "… *adding additional data to the calibration*…" What type of additional data is this referring to?

R: In parenthesis, we will add the most frequent complementary data used by glacio-hydrological models (snow cover area and glacier mass change; Van Tiel et al., 2020)

Ln 91: Please explicitly mention the six sources of uncertainty here, bringing information in Ln 94 – 96 earlier, and then mention the tools used (OGGM, random forest).

R: Thanks for your comment. The specific paragraph will be reorganised according to your suggestions.

Ln 91 onward: The preceding paragraph talks about different sources of uncertainty in the modeling chain in current literature (beyond what the authors have explored in this work). Various statements in this study imply that it focuses on comprehensive sources of uncertainties across the modeling chain. In reality, it only considers two sources i.e., model input data (glacier outlines, thickness) and atmospheric forcing (historical data, GCMs, climate scenarios, and bias-correction methods). I would not consider this encompassing "*uncertainty sources in the modeling chain*" (e.g., Section 5.1 heading). The full chain will include considerations such as model workflow, parameter space, MB or ice flow schemes, etc. (this is duly mentioned as a study limitation in Ln 517).

I recommend removing the term "glacier modeling chain" here and in other instances, including the abstract, and explicitly mention that various configurations of model input and forcing data are used to identify the dominant sources of uncertainty. This is an extensive assessment with these two broad categories as it is, so there is no need to overstate the study's scope.

R: We appreciate the insightful feedback and acknowledge the concerns raised about the term 'glacier modelling chain'. After careful consideration, we agree that the term may not accurately reflect the scope of our study. We will remove the term 'glacier modelling chain' and explicitly clarify in various sections, including the abstract, that our focus is on the impact of different sources of 'data' uncertainty. The revised version will better emphasise our comprehensive assessment within the specified categories of model inputs and forcing data, without overstating the scope of the study.

Study area

Ln 103: What is meant by pristine environment here? Is this referring to glaciers or the water resources?

R: The specific sentence will be removed.

Ln 111: These nine hydrological zones are referred to extensively throughout the paper. Please fully name all of them here.

R: The nine hydrological zones will be defined in the indicated paragraph. Particular attention will be given to the northern zones of PPY and PCA which were only defined in Fig. 1.

Ln 111: I understand how the nine hydrological zones were identified and demarcated, but can you please elaborate how the 847 catchments were selected. Was having a 0.1% glacier area of the total catchment area the only criteria? Why was this threshold selected?

R: Thanks for the question. We selected 847 glacierised catchments, each with at least one glacier and a glacier area greater than 0.1%. The 0.1% was selected as a conservative threshold for drought buffering (please see Fig. 3 in Ultee et al. 2022; cited in main text). This clarification will be added to the main text.

Ln 113: Can you please explain "*high explanatory power of recent glacier change*" or reword this statement.

R: The statement will be clarified: "that shown a strong capacity to reproduce recent glacier changes"

Ln 123: "*are characterized by many small catchments*". Replace characterize by 'which hosts many small catchments' or something similar.

R: Thanks for the suggestion. This will be changed as suggested ("which hosts many small catchments").

Methods

Ln 142: Please mention the upper and lower thresholds here in brackets.

R: This will be added to the sentence in brackets [0, 2 ºC]

Ln 146: OGGM is a flowline model so what is meant by "local grid" of 10 – 200m? Is it referring to the spacing between cross-sections across the centerline?

R: Yes, in the context of OGGM, the "local grid" refers to the spacing between grid points on the glacier surface. This grid is used to obtain the model glacier geometry by overlaying glacier inventory outlines and NASADEM elevation data. The resolution of the grid varies with the glacier size, ranging from 10 to 200 m. Glaciers are then segmented into elevation bands following the algorithm proposed by Werder et al. (2020), each of which covers an elevation difference of 30 m. This response will be used to clarify this procedure in Section 3.1.

Ln 151: If the precipitation factor is set a single value of 1, then it is no longer accounting for biases due to topography, missing processes etc. (Ln 137). Also, what does it mean to "… *assess influence of different reference climates* …" Is this referring to the temperature sensitivity parameter as the precipitation factor is set to 1 for all cases.

R: Thanks for the comment. It is true that setting the precipitation factor (Pf) to 1 no longer accounts for the missing local processes, and therefore these processes are included in the modelling framework with the temperature sensitivity parameter. Regarding the second point, the statement also includes the historical precipitation of the reference climate dataset, as it influences, for example, the interannual variability of the glacier mass balance. The decision to use Pf = 1 is based on the fact that some of the regional climate datasets used in this study (e.g. PMET and CR2MET) already include a bias correction process to correct for potential underestimation of precipitation. Unlike many previous studies which strongly "correct" the input datasets before using them to drive the impact model, we chose to incorporate the uncertainty of the driving dataset in our evaluation. This will be added to the discussion on model calibration.

Ln 163: Replace "*new value*" with initial value or first guess.

R: The "new value" will be replaced by "initial value"

Ln 172: Is the simulated glacier volume set to match the input volume accumulated over each of the nine hydrological zones and not the individual 847 sub-catchments?

R: Yes, please see the reply on Volume-Area scaling (VAS) below

Ln 187: Remove "it" after compared.

R: Thanks! The "it" will be removed.

Ln 191: This might be a typographical error, what is N˚1?

R: Nº1 refers to the first step. This will be clarified.

Ln 201: What does "*latitudinal patterns in terms of area*" mean?

R: The specific sentence will be replaced by "both datasets show similar areas across different latitudes"

Ln 209: It helps to keep the language simple. Can this be rephrased as M22 having 13% larger volume than F19… or something along those lines.

R: Thanks for the suggestion. The sentence will be modified as suggested ("In the southern Andes, Hock et al. (2023) reported that M22 had 13% more total ice volume than F19")

Ln 210: "Both alternatives" - consider replacing with the 'two volume data sources' or something similar.

R: Thanks, we will follow this suggestion.

Ln 213: By dataset do you mean the 4 gridded products?

R: This paragraph refers to the two ice volume datasets. To clarify this, we will replace "datasets" with "volume data sources", as suggested in the previous comment.

Ln 213: VAS was done for the 9 hydrological zones; can you explain here why not for the 847 individual catchments?

R: Thanks for the question. Considering that most catchments have only a few glaciers, and the theoretical basis of VAS is determined for samples of glaciers spanning a wide range of sizes (and is validated by observations covering a wide range of sizes; Bahr et al. 2015), we decided to calculate the VAS parameters using the hydrological zones.

Bahr, David B., W. Tad Pfeffer, and Georg Kaser. "A review of volume-area scaling of glaciers." Reviews of Geophysics 53.1 (2015): 95-140.

Ln 228: Can you please elaborate on why/how these specific GCMs were selected? Was this a subset from a larger initial pool and the 10 GCMs were selected based on their TCR and ECS?

R: Thanks for the question. We will add a sentence to clarify this point: "Using the full CMIP6 ensemble, the selection of the 10 GCMs was based on the recommendations of Hausfather et al. (2022), who suggest focusing on a subset of GCMs that are most consistent with the assessed warming projections of the Sixth Assessment Report (AR6)." For more details on the source of the full ensemble, please see the "Data availability" section.

Ln 234 – 236: This is rather confusing: all GCMs have ECS falling in "very likely" range but only 80% in "likely"? Is the very likely range broader than the likely range? The "likely' range in *Hausfather22* was narrower (1.4 - 2.2 C).

R: Thanks for the comment, it gives us the opportunity to clarify this point. Hausfather et al. 22 recommended screening out models with a TCR that lies outside the 'likely' range of 1.4–2.2 °C, or alternatively using a "likely" ECS of 2.5-4°C, which also reproduces the AR6 results well. In view of these alternatives, the sentence relating to ECS will be removed from the main text to improve the clarity of the GCM selection.

Ln 257: It is somewhat unclear where, when, and why catchments and hydrological zones are considered separately, e.g., volume-area scaling was done at the hydrological zone scale, glacio-hydrological signatures were assessed at catchment scale (Ln 288), etc.

R: Thanks for the comment. Our initial intention was to conduct the complete study using the catchment scale only. However, in order to present the glacio-hydrological projections in Section 4.2 (Fig 8, 9 + Fig S1 and S2) we needed a clear level of aggregation (i.e., hydrological zones). Additionally, the theoretical basis of VAS is determined for samples of glaciers spanning a wide range of sizes, and therefore, the proposed zones were more appropriate than the catchments (please see the previous response about VAS procedure).

Ln 259: Not sure I understand what is meant by area/volume aggregation based on terminus location.

R: Thanks for the comment. This will be clarified in the specific paragraph: "For area and volume, we calculated the relative and absolute differences for each catchment and hydrological zone defined in Fig. 1. To calculate these differences, we aggregated glacier area and volume for a given catchment by selecting all glaciers with their terminus location within that catchment. It is assumed that, if the inventory outlines are correct, all the water flowing out of the glacier will flow via its terminus".

Ln 261: This lapse rate was used because it is default in OGGM and also used within this domain in the past studies (Ln 141). Table 1 however also shows that literature used different lapse rates depending on the region, e.g., 6.5 for NPI and 5.8 for GCN. Is there a north-to-south gradient in the mean annual lapse rates? Does that affect the downscaling of ERA5 data? Also, is this downscaling step the same as what is mentioned in Ln 141? What about total precipitation, how was that handled when going from a quarter degree to 0.05˚ in ERA5?

R: Thank you for the question, it gives us the opportunity to elaborate on our decision to continue using a constant environmental lapse rate of 6.5 ºC km^{-1} to downscale the temperature. Although recent studies in Patagonia have demonstrated the variability of this value (Bravo et al. 2019 in SPI), recent efforts during the development of the PMET (Aguayo et al. 2024; citation will be updated in the text) have shown that seasonal and regional variations of this value in Patagonia do not significantly improve performance (results based on > 100 meteorological stations). Using a lapse rate of 6.5 ºC km^{-1}, all reference climates used in this study are able to achieve mean biases of less than 1 ºC (Fig. 5 in Aguayo et al. 2024). No downscaling method was used for precipitation. This information will be added to the main text.

Bravo, C., Quincey, D. J., Ross, A. N., Rivera, A., Brock, B., Miles, E., & Silva, A. (2019). Air temperature characteristics, distribution, and impact on modeled ablation for the South Patagonia Icefield. Journal of Geophysical Research: Atmospheres, 124(2), https://doi.org/10.1029/2018JD028857, 907-925.

Aguayo, R., León-Muñoz, J., Aguayo, M., Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Fernández, A., and Jacques-Coper, M.: PatagoniaMet: A multi-source hydrometeorological dataset for Western Patagonia, Sci Data, 11, 6, https://doi.org/10.1038/s41597-023-02828-2, 2024.

263: Please mention these default parameters either here or in supplementary material. I suppose Ln 153-154 mentions a couple of these, but please enlist all the parameters for reproducibility. Model "defaults" can change over time as new information becomes available.

R: The parameters will be added here (T_{melt} = -1 ºC, T_{solid} = 0 ºC and T_{liquid} = 2 ºC).

Ln 264, 273, 361, etc.: "*Glacierized grid cells*" – to confirm, these grid cell-based computations are only done for input & forcing data because OGGM is a flowline model and the output information (area, volume, melt) will be for a specific glacier?

R: Yes, OGGM uses the climate data from the nearest grid cell to generate outputs for each glacier. In the case of air temperature, the raw climate data is downscaled to the glacier surface (elevation bands) using a constant lapse rate.

Ln 275: To clarify, when uncertainty metrics are computed all glaciers are considered (Ln 256), but for hydrological assessments only this subset is considered? Section 4.1 is for all the glaciers in RGI and Section 4.2 is for these ~2000 glaciers?

R: Yes, you are correct. Due to computational limitations, we used the OGGM model to estimate the evolution of only the glaciers with an area > 1 km^2 (~2,000 glaciers) (Section 3.4). The comparative analysis (Section 3.2) was performed considering all glaciers in RGI6 and RGI7.

Ln 278 – 279: Up till this point, I do not think historical (outline, volume, ref. climate) vs future (scenarios, GCMs, BCM) sources are explicitly mentioned and separated in the text. So, the "*2 . 2 . 4*" reference is not clear.

R: Thanks for the comment. We agree that it is important to explicitly mention this here: "For each glacier, we evaluated 16 scenarios generated by the historical sources of uncertainty: glacier outlines (n = 2), volume datasets (n = 2) and reference climates (n = 4). These scenarios were used to project the future evolution given by different GCMs (n = 10), future scenarios (n = 4), and bias correction methods (n = 3), resulting in 120 future scenarios for each historical simulation (a total of 1920 potential scenarios; Fig. 2)"

Ln 283-284: I understand on- and off-glacier liquid precipitation, and "on-glacier" melt. But what is off glacier melt? Is it referring to snow on non-glaciated areas? Also, the term on-glacier melt is unusual, I believe this is referring to direct melt from glacier?

R: The melt off-glacier melt corresponds to snow melt on areas that are now glacier-free (i.e. 0 at the start of the simulation; in our case 1980). The term "melt on glacier" will be renamed to glacier melt and clarified in this text: "we also extracted the melt on glacier (hereafter glacier melt), which is the sum of ice and seasonal snow melt on the glacier (Fig. 2c)".

Ln 284: It seems this line and onward is now referring to the full 1920 scenarios and not the 16 historical ones? How is climate uncertainty influence being overestimated? What is meant by 'climate' here, and what is precipitation reduction?

R: Thanks for the questions. We will divide the specific paragraph into two: one for the scenarios and the following for the variables extracted from OGGM. The second paragraph will start with: "For all 1920 scenarios, we extracted…"

Regarding the second point, this will be clarified in the text: "To disaggregate the impact of projected precipitation changes, we also extracted the melt on glacier (hereafter glacier melt)…"

I suppose "glacier runoff" here is the sum of glacial melt and liquid precipitation on glacier? Please consider rephrasing the 'melt on glacier' with an equivalent term commonly used in literature.

R: Yes, at the beginning of the simulation (i.e., year 1980) glacier runoff is the sum of glacier melt and liquid precipitation. As the glacier retreats, off-glacier melt (seasonal snow) and liquid precipitation are also considered (glacier runoff corresponds to all water originating from the initially glacierized area).

The term "melt on glacier" will be renamed to glacier melt and clarified in this text: "we also extracted the melt on glacier (hereafter glacier melt), which is the sum of ice and seasonal snow melt on the glacier (Fig. 2c)".

Ln 286: Is this referring to the melt and precipitation time series? Also, I do not understand the meaning of 'according to glacier terminus' here.

R: In the previous comments, we have defined and clarified the glacier runoff and melt time series. The glacier terminus reference point will be clarified in Section 3.3, and a reference to that clarification will be included here: "As in the comparative analysis (Section 3.3), the time series were initially aggregated at the catchment scale according to the location of the glacier terminus".

Ln 288: Now this assessment is at catchment scale and not hydrological zone scale, again it gets fuzzy where hydrological zones vs catchments are considered. Also, aren't SSP-based scenarios 4, and n=1920 the total number of scenarios?

R: As mentioned above, our original intention was to conduct the entire study at the catchment scale only. However, in order to present the glacio-hydrological projections in Section 4.2 (Figs. 8, 9 + Figs. S1 and S2), we needed a clear level of aggregation (i.e. hydrological zones). In addition, the theoretical basis of VAS is determined for samples of glaciers covering a wide range of sizes, and therefore the proposed zones were more appropriate than catchments. On the other hand, in order to identify clear patterns using the permutation feature importance of RF regression models, we need to have a representative number of samples greater than just 9 hydrological zones. We hope that this clarification will help to understand our decisions that prevented us from using only one scale of analysis.

Results

Ln 321-322: Can you please make the statements explicit here, e.g., RGI7 has 4% and 15% greater area than RGI6. The "*showed positive/negative difference*" is perhaps not the best way to state this.

R: The specific sentence will be rephrased: "showed increases ranging from 4% to 15% relative to RGI6". We agree that "showed positive/negative difference" is not the clearest way to show the differences and therefore we will rephrase similar sentences.

Ln 325: There is only a 1–year difference in the acquisition date for most glaciers from RGI6 to RGI7 (Fig. 3c) – did that make such a large difference in terms of area reduction?

R: Although a detailed comparison between the different sources of data uncertainty is beyond the scope of this study, the differences reported here are similar to the detailed comparison made by Zalazar et al. (2020) (see Fig. 11 there). For example, Zalazar et al. (2020) showed that the largest absolute differences are in the Patagonian Icefields, and some (one-degree) latitudinal bands can show relative differences of more than 50% in the Patagonian Andes. It should also be noted that improved outlines and corrections from local inventories in RGI7 contribute to the observed differences in glacier area.

Ln 335: Just for clarification, both M22 and F19 have the same glacier outlines taken from RGIv6?

R: Yes, in order to make a proper comparison we use RGI6 in both ice volume datasets. This was mentioned in caption of Fig. 4 and will be added to the indicated sentence.

Ln 336: Fig. 4b is the M22 minus F19, correct? Most of the area is in yellow shades visually so M22 shows less volume. Perhaps it is the colorbar that needs to be changed.

R: Yes, that's correct. Panels B and C are differences in percentage relative to F19 ((M22 - F19) / F19). We will adjust the range of the colormap from [-100, 100] to [-50, 50] to have more visible differences.

Ln 351: Replace "in" with "over 51% of the glaciated area".

R: This will be changed as suggested.

Ln 352: Throughout the paper, the reference to glacier/glaciated area and catchment area together is very confusing, for example this line mentions 51% of glacier area, 22% of the catchment area for the precipitation reference and 95% of the glacier area, 99% of the catchment area for temperature reference. Please consider presenting this information in some other way (perhaps focus on one of these only).

R: Thanks for the comment. This was one of the general comments of Reviewer 1. The percentages related to catchment area will be removed from the main text in order to the information clearer to the future reader.

Ln 356: Similar patterns in terms of spatial patterns?

R: Thanks for the question, the wording was not clear. The sentence will be clarified: "In relation to PMET, CR2MET and MSWEP showed a difference of nearly -50% in solid precipitation, while ERA5 showed smaller differences close to -20% (Fig. 5c)."

Ln 356: The spatial resolution of the native data is quite different (0.25 for ERA5, 0.05 for PMET/CR2MET), does that come into play at all?

R: Please see our previous response about the spatial resolution of the climate data (use of lapse rates)

Ln 367: Also, are there other characteristics that create these latitudinal differences in (a) precipitation change sign and (b) model agreement/disagreement?

R: The projected changes in precipitation follow a latitudinal pattern which generates a low agreement in the intermediate zone (SPI and GCN) due to the fact that it is a transitional zone (ΔPP = 0%). These projections are consistent with historical trends that has been attributed to the Southern Annular Mode (SAM), an index that describes the movement of the low-pressure belt that generates westerly winds (Fogt et al. 2020). This index has recorded a transition to a positive phase, causing a decrease in the intensity of westerly winds at mid-latitudes, which generates a significant part of the decrease (increase) in precipitation in northern (southern) Patagonia.

Fogt, Ryan L., and Gareth J. Marshall. "The Southern Annular Mode: variability, trends, and climate impacts across the Southern Hemisphere." Wiley Interdisciplinary Reviews: Climate Change 11.4 (2020): e652.

Ln 370: Why/how is ice volume relevant to climate projections? This is a rather obtuse statement, talking about GCM model disagreement and then ice volume estimates.

R: The specific sentence ("areas characterized by a high ice volume") will be removed from the paragraph.

Ln 374: What do you mean by the "main catchments".

R: The main catchments correspond to the catchments that have an area greater than 5,000 km^2 (Fig. 1). All maps have the following sentence to indicate the main catchments: "The names in grey correspond to the names of the main catchments"

Ln 405: What do you mean by the prolongation of the mass loss? Is this referring to Fig. 8b?

R: Yes, but more generally to the whole of Fig. 8, as all panels refer to volume loss, which will be clarified in the specific sentence (volume instead of mass): "Projections from OGGM indicate a prolongation of the glacier volume loss of recent decades (Fig. 8)."

Ln 406: Again, please reconsider the discussion related to catchment and glacier area. It is surprising that 18% of glacier area is equivalent to 43% of the catchment area. Also, how is catchment losing volume (it is only the glaciers that will be losing the volume).

R: Thanks for the comment. This was one of the general comments of Reviewer 1. The percentages related to catchment area will be removed from the main text in order to the information clearer to the future reader. In this case, the difference between 18% and 43% is explained by the fact that the glaciers are located in the northern area which is characterised by the largest catchments (Petrohue, Puelo, Yelcho, Palena, etc.)

Ln 417: Is this the maximum range across all the basins? It seems some basins have less or more spread across the scenarios (e.g., SPI-N). Again, can you comment on the difference between the hydrological zones here or in discussion.

R: Thanks for the question. We will address this by calculating the differences across hydrological zones (in this case, specific mass balance) and discussing the spatial differences as part the proposed new structure of the discussion section. This will provide deeper insights into the observed differences between sub-regions and scenarios.

Ln 427: 61% of the total catchment area contains 30% of the total glacier area i.e., the remaining 70% of the glaciated area in the ~40% of the catchment area? Here it is best to talk about glacier area (because it is talking about glacial melt), catchment area reference is somewhat misleading.

R: We agree that the catchment area could be misleading, so the reference to catchment area will be removed as before.

Ln 428: Is this melt now talking about all the glaciated area in the full study domain?

R: Yes, this will be clarified in the specific sentence.

Ln 430: Please consider rephrasing "the evolution of the melt on glacier…". Do you mean changes in melt rates?

R: Thanks for the question. This will be clarified in the main text: "The projected trajectories of glacier melt varied slightly among emissions scenarios"

Ln 435: This is referring to Fig. 9e for SPI, correct?

R: Yes, we will correct the reference to the panel figure.

Ln 436: Which panel of Fig. 5 is this referring to? Please rephrase "*melt on glacier evolution*".

R: The reference to Fig. 5 was incorrectly located there and has been removed: "Relative to glacier runoff (Fig. S2), the uncertainty (i.e., standard deviation) in the glacier melt was lower due to a lower influence of the reference climate"

Ln 455: Remove comma after 'contribution'.

R: Thanks for the comments. The comma after 'contribution' will be moved before 'metrics': "This was especially clear for (…) metrics, where the reference climate accumulated…"

Ln 468: What is meant by lower importance of climate here? Reference climate?

R: Yes, this will be clarified in the revised version ("The lower importance of the reference climate was limited to…").

Ln 470: Can you explain why there are differences between seasonal contribution, variability and shift in Fig. 10? What are the mechanisms causing these differences in sources of uncertainty.

R: Before addressing the differences observed in Fig. 10, it is important to acknowledge an error in the calculation of seasonal variability. The seasonal variability was incorrectly based on the full period rather

than the reference period, which increases the importance of future sources of uncertainty such as GCM, SSP, and BCM (bias correction method). The results and discussion section will be updated using the corrected values.

Considering the updated Fig. 10, the future sources of uncertainty do not contribute to the uncertainty in seasonal contribution and variability. In both cases, the reference climate accumulated more than 60% of the feature importance, followed by the glacier inventory and the volume. There are not clear differences in these metrics between glacier runoff and melt. The seasonal shift, which is the absolute change in summer contribution (DJF) between the reference and future periods, shows a different pattern. For this metric, the reference climate shows a similar importance to the SSPs and GCMs. It is important to note that the greatest importance of the bias correction method is for this metric. Mechanisms that underscore the importance of the reference climate as a major source of uncertainty are likely to include its role in defining the baseline conditions against which future changes are assessed. In addition, the influence of the reference climate on temperature and precipitation patterns directly affects seasonal glacier response (melt / accumulation), contributing to its importance in determining seasonal metrics.

Discussion

Overall, the discussion section needs some attention from the authors. This is not a study that can be easily replicated for other domains in terms of time commitment and computing resources needed. That makes it important to highlight the big picture findings, i.e., statements that provide insights on how to interpret the results for other regions (Are the findings applicable globally? If yes, then how was the conclusion reached? If not, then what are the differences or other cases that might need to be considered?).

R: Both reviewers have raised concerns about the discussion section. To address this, we have proposed a new structure for this section. This revised structure aims to minimise repetition of results and to consolidate previous sections on uncertainty. In addition, we plan to incorporate the suggestions of both reviewers by focusing more on the implications of the results, rather than simply comparing them with other studies. This approach will improve the clarity and coherence of the discussion and allow readers to better understand the significance of our findings.

The proposed new structure is as follows: Section 5.1 ("Changing glacier hydrology") will present regional projections of changing glacier hydrology and compare them with results from previous studies. Section 5.2 ("Hydrological importance of data uncertainty") will discuss the hydrological importance of data uncertainty, summarising comparisons and highlighting the importance of sources of uncertainty, emphasising spatial differences and the importance of domain characteristics, as suggested by reviewer 2. In addition, the importance of model calibration is addressed, as suggested by Reviewer 1. Finally, Section 5.3 ("Limitations and global implications") discusses limitations such as unconsidered sources and potential global implications suggested by both reviewers.

Also, for this specific domain, the authors should discuss the differences across the various hydrological zones e.g., how does domain characteristics in terms of climate, topography, glacier size, etc. affect the six sources of uncertainty and the ten runoff metrics?

R: We appreciate the suggestion for the discussion. As part of the newly proposed Section 5.2, "Hydrological significance of data uncertainty", we will explore the spatial differences between different hydrological zones, with particular emphasis on climate. Given the hydrological importance of climate (Fig. 10) and its diversity in our study area, we will analyse the differences between historical reference climates and their impact on hydrological metrics. In addition, we will discuss the potential implications of our findings for other high mountain regions, which often face challenges in constraining gridded precipitation products due to a low density of gauging stations (Section 5.3). This will provide valuable insights into the spatial variability of hydrological dynamics and have implications for glacier modelling and water resource management in similar regions.

Ln 484-485: This statement implies that the difference in acquisition year played a significant role in glacier area, while in Fig. 3d there is only a 1-year difference between the two RGI versions. Are you saying there was a large area loss between 2000 and 2001?

R: In addition to the previous response regarding differences in glacier area between the inventories (results consistent with Zalazar et al. 2020), we will highlight that improved outlines and corrections from local inventories in RGI7 also contribute to the observed differences between inventories.

Ln 486 – 488: The sentence from "*While the 69% of the total catchment … estimates*" is unclear, please clarify and rephrase it. Also in Ln 488, what observational data are you referring to (glacier outlines or climate)?

R: The specific sentence will be moved (and rephrased) to the results to avoid repetition in the discussion section: "In terms of glacier area, 69% showed a relative difference of less than ± 10% between the two glacier inventories, while only 27% showed differences below this threshold when considering both ice volume estimates"

In Ln 488, we are referring to ice thickness estimates. This will be rephrased for clarity: "The use of direct or indirect observations of ice thickness, such as those based on ground-penetrating radar or airborne surveys, can help to select a better dataset for the study area, potentially reducing the data uncertainty. However, these observations are spatially and temporally scarce in the Patagonian Andes."

Ln 491 – 493: Not sure what this sentence is getting at.

R: The specific sentence will be clarified and rephrased: "Furthermore, the generation of large-scale ice thickness datasets requires the compilation of numerous datasets derived from different acquisition dates, often spanning years or even decades, which hinders regional validation".

Ln 537: What is meant by "*continue changing*"?

R: The use of "continue changing" suggests an ongoing and evolving process, driven by the sustained loss of glacier mass (historical and projected) in the region. The sentence will be clarified.

Ln 539: Just to clarify, 43% of the catchment will lose 80% of their glacial volume? Also, there is latitudinal dependence (North-South divide). Can you talk more about that?

R: Yes, considering the mean derived from the full set of SSP scenarios (n = 1920), 18% of the total glacier area will lose more than 80% of its current volume this century (the use of % of catchment area will be removed from the manuscript to avoid adding unnecessary complexity to the study). The regional projections and their spatial differences will be discussed as part of the proposed new structure of the discussion.

Ln 545: How was this conclusion reached? Is this talking about *Rounce et al. (2023)* or the current study?

R: The conclusion was reached by comparing the uncertainty (i.e. error bars in Fig. S4) associated with the volume loss in 2100 between our study and Rounce et al. (2023). Although Rounce et al. (2023) only considered GCM and SSP as sources of data uncertainty, the error bars, which are calculated using one standard deviation, have surprisingly similar values in both cases (Fig S4). We recognize that this deduction requires further clarification, and we will provide additional context and explanation in the revised version of the conclusion.

Ln 565-566: Statement regarding peak water already reached – can this be supported with observational data for the region (at least that can help understand the historical metrics between 1980 – 2023, if the data is available)?

R: Thank you for the question. Ground-based validation is very important, but in high mountain areas the availability of ground-based data is generally limited, and the Patagonian Andes are no exception. Despite recent efforts during the development of PMET-obs (the ground-based alternative to PMET), only 9 (out of 109) catchments with stream gauges have more than 10% glacier area, making it difficult to properly isolate the glacier influence. Nevertheless, it is planned to use these data for hydrological modelling in an ongoing effort, which we expect to be available as a preprint this autumn (March - April).

Conclusions

Much of this reads like results or discussion section. For example, the second bullet can be moved to results.

R: We will re-evaluate the distribution of information to ensure a more coherent structure. The main findings of the study will be presented and highlighted in the results section, and the conclusions section will focus on summarising the key takeaways and implications of these findings.

Ln 629: This study did not provide insights into "*local calibration choices*". As the authors rightly point out in the following sentence, future studies need to look into MB schemes, parameterizations, etc. This study is fairly comprehensive as it is, so the authors should try to highlight the significance of the work done and how it can be interpreted for other domains.

R: Thanks for your comment. We agree with this point, and therefore we will replace "local calibration choices" by "impacts of data uncertainty".

Figures

Fig. 1 and throughout: Can you please make the delineation between the 9 hydrological zones more prominent (e.g., in thick black line). The zone labels in 1b and c are not visible.

R: We will change the transparency of the delineation lines in order to make them more prominent. We will also change the colours of the zone labels in Fig. 1b,c

Also, it is better not to use sequential colormap for discrete categories. For example, it is hard to see the difference between <500 and <1500 in 1c. Readers should be able to extract this information quickly.

R: Thanks for the suggestions. We will replace the discrete categories in b) and c) with sequential colormaps.

Fig. 2: This flowchart is hard to follow, can you provide step numbers along the way (preferably in same sequence as the text from Ln 162).

R: Thanks for the suggestion, we will add roman numerals in Fig. 2 to refer to the calibration steps in Section 3.1

Fig. 3: Please consider using a more distinct diverging colormap for 3a, with white in the middle. It seems most of the regions are around yellow, so there is no (visible) difference between RGI 6 and 7. Also the bar colors in panel b are barely visible. Perhaps remove the grey background and make all the colors in darker shades.

R: Thanks for the suggestions. We tried to use to white in the colormap of Fig 3b, but the contrast with the background was not adequate. We will adjust the range of the colormap from [-100, 100] to [-50, 50] to have more visible differences in the northern and southern zones. We will also add black outlines and a zero line to the bar plot in Fig 3b to prevent the bars being confused with the background.

Fig. 4: For panel b, please see same comment as before. For panel a, consider a discrete colorbar with set ranges (0-20, 20-40, etc.). It's very hard to see the differences between the domains.

R: Thank you for your suggestions. For Panel B, we will also adjust the range of the colour map (as in Fig. 3). We agree that the different order of magnitude makes it difficult to analyse the differences in volume between catchments. As an alternative to the proposed discrete colourmap, we will use a log scale for Panel A to address this issue.

Fig. 5, 6, 7: Same as 4a. I do not see the dotted line for the mean value mentioned in the caption.

R: The dotted lines were places within the boxplots to indicate the mean values. Considering that the violin plots already show the distribution of the values, we have decided to remove the dotted lines in the box plots.

Fig. 6: The dots in panel a are not visible.

R: Similar to Fig. 7, where the low agreement is indicated with black outlines, we will replace the dots in Panel A with black outlines ("The catchments with black outlines indicate low model agreement, where less than 80% of the models agree on the sign of the change.")