

Editor

Dear Rodrigo Aguayo, and co-authors,

First of all I want to thank you and your co-authors for the dedicated and thorough point-by-point answers to all review comments and the respectively planned actions. In the following, I only want to highlight some of the major changes that you indicate.

- According to one major request by reviewer #1 (supported by reviewer #2), you plan to avoid the catchment terminology in your revised manuscript. I agree that keeping the glacier focus in your analysis will make you manuscript more accessible.

R: We acknowledge the concerns raised about the introduction of "catchments" and the resulting confusion for the reader. We recognise the importance of clarity in our presentation, and therefore we agree that the use of "% of catchment area" in the comparison between the different sources of uncertainty may add unnecessary complexity to the study. Consequently, we have decided to remove these values from the manuscript, and we have changed the title of the manuscript to "Unravelling the sources of uncertainty in glacier runoff projections in the Patagonian Andes".

In addition, we have simplified the presentation in the revised version by i) focusing our analysis only on 'glacier area (at inventory date)', ii) presenting the relative differences between the different sources of uncertainty in terms of glacier area (%), and iii) adding glacier area-weighted averages to summarise the differences. These changes have been incorporated into Figures 5-7 and 10 and will contribute to a clearer and more concise presentation of our results. Finally, we have retained the level of aggregation of hydrological zones in order to examine spatial differences over a wide area encompassing different climates and geographical features (Figs. 8, 9, S1 and S2).

- Both reviewers have requested a re-organisation of the discussion section to make it more coherent and concise. Both also request a focus on implications. You suggest a new structure for the discussion with 3 sub-sections. You briefly outline their content but it still remain rather vague.

R: Thanks for the comment. In the revised version, the new structure aims to minimise repetition of results and to consolidate previous sections on uncertainty. In addition, we have incorporated the suggestions of both reviewers by focusing more on the implications of the results based on current limitations. This approach will improve the clarity and coherence of the discussion, and we hope that it will allow readers to better understand the significance of our results and the current challenges in assessing uncertainty in large-scale glacier models. The new structure of the discussion in the revised version is as follows:

- 5.1 Hydrological response of Patagonian glaciers to climate change: This section summarises the regional glacier projections and their associated uncertainties. In this section we have maintained the comparison of glacier runoff with results from previous studies.

- **5.2 Hydrological importance of data uncertainty:** This section links the analysis of sources of uncertainty with their influence on different glacio-hydrological signatures, emphasising spatial differences due to domain characteristics, as suggested by reviewer 2.

- **5.3 Influence of model calibration:** As the calibration approach varies between glacier models and has a direct influence on glacier discharge (through the precipitation factor), we have added a section to discuss this issue, as suggested by Reviewer 1.

- **5.4 Limitations and global implications:** This section discusses limitations such as unconsidered sources and potential global implications suggested by both reviewers.

- I am glad that you intend to follow the 2nd reviewer's suggestion on partitioning your uncertainty assessment into two over-arching categories: model inputs & forcing data. I am convinced that these categories will make it more easy to distill your main conclusions.

R: In the revised version, we have incorporated the suggestion to partition our uncertainty assessment into two overarching categories. However, the proposed categories aggregate the sources of uncertainty by time period (historical and future data), instead of model inputs and forcing data. We believe that these categories are better suited to understanding how the historical sources of uncertainty are projected into the future, given that the primary objective of glacio-hydrological modelling studies has been to assess the future impacts of climate change.

- Concerning your selection of 10 GCMs from the CMIP6 ensemble, your explanation refers to the AR6 decision criteria using global temperature metrics. Yet I wonder about the performance of the selected GCMs over South America. Please provide any information. Are they well suited for this region?

R: Thanks for the question, it gives us the opportunity to examine the regional performance of the selected GCMs based on recent studies. In Chile, we have two studies that have investigated the adequacy of large ensembles of GCMs (Salazar et al., 2024; Gateño et al., 2024).

Given that raw GCM time series are bias-corrected using three different techniques as part of the sources of uncertainty, we will focus on Gateño's study, which proposes a framework that goes beyond bias-related metrics to include metrics related to the ability of GCMs to reproduce teleconnection responses to the El Niño Southern Oscillation (ENSO) and the Southern Annular Mode (SAM), which can affect regional climate variability and trends. For example, the SAM has shown a transition to a positive phase, causing a decrease in the intensity of mid-latitude westerlies, which generates a significant part of the decrease (increase) in precipitation in northern (southern) Patagonia.

Specifically, the GCM information in Table S2 was complemented with regional Past Performance Indexes for precipitation (PPI_{PP}) and temperature (PPI_{T2M}). In addition, we have included the selected GCMs from the screening approach based on the analysis of Gateño et al. All the selected GCMs showed an adequate performance in reproducing temperature, while only half of them showed a good performance for precipitation. Out of the 10 GCMs selected, 4 are included in the screening recommendation of Gateño et al.

(initial pool of 27 GCMs), highlighting the appropriateness of our selection. This information has been included in the discussion section as a recommendation to reduce the future overall uncertainty.

Salazar, Á., Thatcher, M., Goubanova, K. et al. CMIP6 precipitation and temperature projections for Chile. *Clim Dyn* 62, 2475–2498 (2024). DOI: 10.1007/s00382-023-07034-9

Gateño, F., Mendoza, P.A., Vásquez, N. et al. Screening CMIP6 models for Chile based on past performance and code genealogy. *Climatic Change* 177, 87 (2024). DOI: 10.1007/s10584-024-03742-1

- Please also consider my initial comments from the access review in your revised manuscript. I am very pleased that you intend to include a paragraph on the influence of model calibration on your uncertainty analysis. In how far is the high importance of the historical sources of uncertainty a result of the model calibration. This comment is in line with request raised by reviewers #2.

R: Thanks for the comment, we have added a new section to the discussion entitled "Influence of model calibration" (Section 5.3). In this section we address the complexities of calibrating large-scale glacier models. We highlight GloGEM's sequential calibration approach and its role in mitigating sensitivity to climate forcing and draw insights from regional studies to underline the importance of reference climate choices. With regard to the scaling effect of the precipitation factor on glacier runoff, our decision not to correct historical climate datasets aims to capture the full range of potential precipitation values while utilising existing bias correction procedures. In addition, we discuss the potential for future studies to refine uncertainty assessment through ensemble meteorological datasets and explore recent advances in model calibration techniques, including Bayesian inference methods.

Altogether I gladly retain your manuscript for consideration in TC. Based on the brief summary of the main actions above, I suggest that your manuscript undergoes a second review round after you revised your manuscript according to your plans.

All the best,

The editor, Johannes Fürst

Anonymous Referee #1

In the manuscript “Assessing the glacier projection uncertainties in the Patagonian Andes (4-56 S) from a catchment perspective”, a large set of glacier model simulations obtained with the OGGM model is used to assess the uncertainty in glacier runoff and glacier melt projections. The study looks at the effect of different glacier outlines, different glacier volume estimates, various historical climate datasets, different GCM, different emission scenarios and different bias correction methods. Each of these different datasets are discussed and it is shown how they vary, mainly focusing on their spatial patterns. Using a random forest regression method, the relative importance of each of these “model choice” uncertainties is examined. The study concludes that the reference climate is the most important source of uncertainty for a range of glacio-hydrological signatures, even for signatures that represent a signal beyond the reference period.

Overall, I think that the study presents an impressive amount of model run comparisons at a large regional scale and shows clear insights into spatial differences of glacier volume and runoff changes in the Patagonian Andes. Moreover, the effect of each source of uncertainty on the glacio-hydrological signatures nicely illustrates their variable importance for different aspects of the change in glacier runoff. However, I feel that some parts of the manuscript could be improved, such as the use of “catchments”, and the discussion section, which should describe more the implications and possible hypotheses of the findings, rather than only a summary of the findings and comparison with other studies. Please find below a list of more detailed comments, also explaining these two examples.

R: Thank you for your thoughtful and constructive feedback on our manuscript. We appreciate your positive assessment of the comprehensive model run comparisons and the insights they provide into spatial variations in glacier volume and runoff changes within the Patagonian Andes. We will carefully consider your suggestions for improvement, particularly regarding the use of "% of catchment area" and the discussion section. We recognise the importance of delving deeper into the implications of our results and proposing potential hypotheses, rather than simply summarising and comparing them with other studies.

To address these concerns, we have revised the discussion section to provide a more thorough exploration of the implications of our findings. In addition, we have chosen to prioritise the use of 'glacier area (at inventory date)' over '% of catchment area' throughout the manuscript to improve clarity for future readers. Detailed responses to these general comments are provided below.

Abstract

The introduction of “catchments” starts here in the abstract. Without stating what “catchment” represents in this study, it is quite hard to follow. The reader doesn’t know if the study is about a few catchments, or many, and how they are delineated/defined. Accordingly, all the statements with “xxx% of the catchment area” are difficult to interpret. On a more general note, it did not become clear why catchments are used in the study? Most of the results are rather described on the basis of the hydrological zones. In the study area description, it is noted “847 catchments were selected” without any further information. On the basis of what were these catchments defined? Since the study is not using any downstream information or non-glacierized catchment info (“only using glacierized grid cells”), I think the “catchment” part should not be part of this study. Results could then be just presented as xx% of glaciers show this and that, while also keeping the aggregation level of the hydrological zones.

R: We acknowledge the concerns raised about the introduction of "catchments" and the resulting confusion for the reader. We recognise the importance of clarity in our presentation, and therefore we agree that the use of "% of catchment area" in the comparison between the different sources of uncertainty may add unnecessary complexity to the study. Consequently, we have decided to remove these values from the manuscript, and we have changed the title of the manuscript to "Unravelling the sources of uncertainty in glacier runoff projections in the Patagonian Andes".

In addition, we have simplified the presentation in the revised version by i) focusing our analysis only on 'glacier area (at inventory date)', ii) presenting the relative differences between the different sources of uncertainty in terms of glacier area (%), and iii) adding glacier area-weighted averages to summarise the differences. These changes have been incorporated into Figures 5-7 and 10 and will contribute to a clearer and more concise presentation of our results. Finally, we have retained the level of aggregation of hydrological zones in order to examine spatial differences over a wide area encompassing different climates and geographical features (Figs. 8, 9, S1 and S2).

Introduction

L44 undergoing "a" shift – shouldn't shift not be an increase, as otherwise there is no increase in risk? Or does it relate to larger volumes that are stored before bursting out?

R: The specific sentence on GLOFs has been removed to improve the continuity with the following paragraph.

L65 "have had to be used" – have been used? Also check the "but" which is more an additional problem than a contrast?

R: The specific sentence has been removed

I would suggest to move table 1 to the SI. The abbreviations in the first column are not yet clear and in general the aim/message of the table did not become clear in the introduction

R: Thanks for the comment. Table 1 has been moved to the Supplementary Material. The abbreviations in the first column have been defined in the table caption.

L93 "to project the evolution of the glacier area" – the model is not only used for projecting the area?

R: The specific sentence says, "to project the evolution of each glacier (area > 1 km²)". The glacier model outputs used are not only the glacier area, but also the volume and hydrological outputs (glacier runoff and melt from formerly glacierised areas, see Section 3.4).

Study area

L104-105 "crucial", "essential" in the same sentence, maybe one is enough?

R: The sentence has been summarised: “the seasonal melting of glaciers is essential for the long-term sustainability of the local ecosystems and coastal human populations”.

Methods

Step 1 of the calibration procedure – I wondered if it should be discussed what the effect is of only calibrating the melt factor/temperature sensitivity using geodetic mass balances. If the precipitation is off (e.g. too little) then the temperature sensitivity might be smaller, to compensate?

R: It is true that the calibrated temperature sensitivity parameter depends strongly on the assumed precipitation, and thus on the chosen precipitation factor (Pf; see e.g. Schuster et al. (2023)). However, for most glaciers we have only one robust observation of mass change, i.e., the 20-year geodetic mass-balance average. Therefore, we decided to set the Pf to 1, as some regional climate datasets used in this study (e.g. PMET and CR2MET) already include a bias correction process to correct for potential precipitation underestimation, and therefore we expect the range of "true" precipitation to be covered by the different products. Based on this comment, the importance of the model calibration in large-scale glacier models has been added to the discussion (please see Section 5.3: Influence of model calibration).

Step 2 – the description is a little bit unclear, is the gradient adjusted? Or what is meant with “a mass balance residual to add”?

R: Sorry for the confusion, the gradient of the mass balance is not adjusted. In fact, the whole mass balance profile is shifted so that the resulting apparent mass balance is zero, to satisfy the equilibrium assumption. We have clarified this in the manuscript (step 2).

Step 3 – maybe the equation could be given here? I think it is similarly known as equation 1, but helps the reader understanding what is needed for the inversion and at what scales it is applied

R: Thanks for the suggestion. We think that the equations for the inversion are not that relevant for the paper and are already very well documented in Maussion et al. 2019. However, we have added the equation for the ice velocity in Section 3.1 to illustrate the control of the ice creep parameter A (which is calibrated in this step).

Step 4 – what is a “constant mass balance run”?

R: A constant mass balance run means that we define a mass balance over a period of time (e.g. the average mass balance of all the years in the period), this is our constant mass balance. Then we use this constant mass balance and let the glacier evolve for a few years. We have clarified this concept in step 4.

Figure 2 – I think it would be helpful to add the number of the steps in the figure. And what is the arrow from reference climate to bed inversion? It may refer to one step before that?

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

L259 “according to the location of the glacier terminus” – I am not sure to understand this addition? Maybe some more explanation of glaciers that are crossing catchment/hydrological zone borders could be added here

R: Thanks for the comment. This has been clarified according to your suggestion: “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.”.

L291 here the text could benefit from some explanation/ careful discussion how streamflow metrics can be used to apply on aggregated glacier runoff data (i.e. glacier runoff is not the same as downstream streamflow and so their effect on the aquatic ecosystem is not 1:1 comparable)

R: Thanks for the comment. We have added a sentence to highlight the limitations of our analysis: “However, our analysis of glacier runoff should not be considered as downstream streamflow because our simulations only considered the initially glacierised area and did not include the interaction with other hydrological fluxes (e.g., evaporation and infiltration)”

Table 2 misses information about which period was used to calculate the signatures, apart from the ones that explicitly state “ref and future period”

R: We have added a “Period” column for each signature.

Header of 3.5 – maybe choose the titles of section 3.3 and 3.5 in such a way that it is more clear what their different content is

R: The original header of 3.5 (“Uncertainty analysis”) has been renamed to “Hydrological importance of sources of uncertainty”

L299 – where comes 329 from?

R: Out of 847 catchments, only 329 catchments have one or more glaciers in both inventories (RGI6 and RGI7). This has been clarified in the text.

In the calculation of RMSE, what is the baseline? i.e. how is RMSE calculated?

R: Thank you for your question. The permutation feature importance measures the change in model performance (in this case, the Root Mean Square Error; RMSE) after the values of a single model feature have been permuted (also known as shuffled), with more important features resulting in greater decreases in performance when permuted. The baseline corresponds to the model performance before the permutation of the model features (in this case, the six sources of uncertainty). This has been clarified in Section 3.5.

Results

Figure 2: possibly remove the blue/gray background so the results are better visible

R: We think that this comment refers to Fig 3, as there are no results in Fig 2. We have added black outlines and a zero line to the bar plot in Fig 3b to prevent the bars being confused with the background.

L335 – why was the glacier thickness divided by the catchment area, while all other variables are focused on the glaciers/glacier grid cells?

R: In the case of ice volume, which is derived from ice thickness and glacier outline, we decided to normalise the value by catchment area to facilitate the comparison between catchments - hydrological zones.

L374 – maybe I missed it, but what are “the main catchments”?

R: The main catchments correspond to the catchments that have an area greater than 5,000 km² (Fig. 1), which account for the 68% of the total catchment area. All maps have the following sentence to indicate the main catchments: “The names in grey correspond to the names of the main catchments”

L408 “historical sources” – climate?

R: Yes. This has been clarified in the text (“Considering the prolongation of historical climate conditions, 26% ± 9% of the total glacier ice is committed to melt in the long term”)

L435-L440 Does isolating the melt from glacier runoff result in more or less uncertainty? There is a hint that melt only has an effect of temperature, but it also states that precipitation can compensate for the change in melt? The 60% suggests that 40% of the glacier runoff is generated from off-glacier melt, or liquid precipitation on+off glacier?

R: Compared to total runoff (Fig. S2), the uncertainty (standard deviation of the annual glacier melt) was reduced because the reference climate, which is the main source of climate uncertainty, had less influence, as the strongly past climate-dependent liquid precipitation is only included in the glacier runoff components (on- and off-glacier melt and on- and off-glacier liquid precipitation). This is mentioned in the subsection 5.1 (“Hydrological response of Patagonian glaciers to climate change”). Therefore, the 60% suggests that 40% of the glacier runoff is generated by off-glacier melt and on- and off-glacier liquid precipitation.

Discussion

For the discussion in general, I found it sometimes hard to follow what was described in the different parts. The first part discusses the uncertainties, but the other two parts as well. Maybe by less repetition of the results, and more focusing on the implications of the results would help in restructuring the discussion. I think the comparison with other studies is mostly well described, but would also benefit from an additional thought on what are the implications of that it agrees well or not.

R: Both reviewers 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 have incorporated the suggestions of both reviewers by focusing more on the implications of the results based on current limitations. This approach will improve

the clarity and coherence of the discussion, and we hope that it will allow readers to better understand the significance of our results and the current challenges in assessing uncertainty in large-scale glacier models.

The new structure of the discussion in the revised version is as follows:

- **5.1 Hydrological response of Patagonian glaciers to climate change:** This section summarises the regional glacier projections and their associated uncertainties. In this section we have maintained the comparison of glacier runoff with results from previous studies.
- **5.2 Hydrological importance of data uncertainty:** This section links the analysis of sources of uncertainty with their influence on different glacio-hydrological signatures, emphasising spatial differences due to domain characteristics, as suggested by reviewer 2.
- **5.3 Influence of model calibration:** As the calibration approach varies between glacier models and has a direct influence on glacier discharge (through the precipitation factor), we have added a section to discuss this issue, as suggested by Reviewer 1.
- **5.4 Limitations and global implications:** This section discusses limitations such as unconsidered sources and potential global implications suggested by both reviewers.

I was wondering if there should be maybe a discussion on the way such glacier models are calibrated? If all model runs are calibrated equally well (RGI areas and glacier volumes are used for calibrating), then how come the results are so different, especially regarding reference climates? These parameters propagate in all other simulations, right? This must then come from processes that are not captured when looking at annual and long-term metrics only when calibrating model parameters? Or maybe there are other processes that are relevant to discuss for improving glacier modelling?

R: Thanks for the comment, we have added a new section to the discussion entitled "Influence of model calibration" (Section 5.3). In this section we address the complexities of calibrating large-scale glacier models. We highlight GloGEM's sequential calibration approach and its role in mitigating sensitivity to climate forcing and draw insights from regional studies to underline the importance of reference climate choices. With regard to the scaling effect of the precipitation factor on glacier runoff, our decision not to correct historical climate datasets aims to capture the full range of potential precipitation values while utilising existing bias correction procedures. In addition, we discuss the potential for future studies to refine uncertainty assessment through ensemble meteorological datasets and explore recent advances in model calibration techniques, including Bayesian inference methods.

L484 "acquisition dates" – this sentence could benefit from more explanation about what the difference is in both RGI outlines with respect to acquisition dates.

R. Thanks for the comment. The specific sentence was removed from the discussion to avoid repetition with the results (subsection 4.1.1): "These regional differences may be due to several factors, including improved outlines and corrections from local inventories and differences in acquisition dates (Fig. 3d) ..."

L487 "this threshold" – do you mean a 10% threshold for volume?

R: Yes, the sentence compared the relative differences between different sources of glacier inventory and ice thickness estimates. The specific sentence has been removed from the discussion in the revised version.

L494 – Isn't this paragraph suggestion a contradiction? It would help starting the paragraph like that.

R: The main topic of the specific paragraph was the uncertainty associated with the reference climate (in specific precipitation). In the revised version, the content of this paragraph has been summarised and included in the "Limitations and potential implications" subsection. Specifically, the content of this paragraph is included in the context of sources of uncertainty that can be reduced using ground-based data.

L510 – a space needs to be inserted before "The selected..."

R: Thanks for the catch. The specific sentence ("The selected GCMs showed a high agreement in most hydrological zones...") was removed from the discussion to avoid repetition with the results.

L512 "the ten selected...of the assessed warming – not sure how this sentence fits in?"

R: The specific sentence has been removed from the discussion to avoid repetition of the methodology/results.

L526 Why are the "older" estimates named after the newer estimate and referred to as "recently"?

R: Thanks for the question. The "recently" was misplaced there and has been removed from the sentence.

L539 "of its current volume" – although it is clear that the study only discusses glaciers, the confusing use of "catchments" means that here there "glacier" needs to be added

R: Thanks for the comment. The reference to "catchments" has been removed from the sentence to clarify the source of the volume loss.

L594 – possibly add that Mackay et al. deals with glacio-hydrological modelling in Iceland

R: Thanks for the comment. We have added this to clarify the scope of the antecedent.

Conclusions

L610 – "differences", "different", "varied" in one sentence – consider rephrasing.

R: The specific sentence has been rephrased: "The six sources of data uncertainty showed differences of varying magnitude".

L613 isn't it a separate point, the one on reference climate being most important for uncertainty?

R: We decided to aggregate the main points of the conclusions based on subsections of the Results: i) Analysis of sources of uncertainty, ii) Hydrological importance of data uncertainty, and iii) Limitations and potential implications.

L622 “tended to converge towards an overall decrease” – decrease of what?

R: Decrease of glacier runoff and melt. This has been clarified in the conclusions.

L625-628 double?

R: Yes, the specific sentence indicated that regardless of the spatial scale of interest (catchment or glacier area), the main source of uncertainty was the reference climate. This sentence has been clarified as the “catchment area” has been removed from the Results and Discussion sections.

L629 – what is meant with “local”?

R: The specific word has been removed from the sentence.

L634 “relative contribution of non-glacial water sources” – shouldn’t future studies not focus on the dynamics of these sources, rather than the relative contribution? And what should future studies do with these non-glacial water sources? The follow up sentence does not directly fit here, i.e. there is some gap between knowing a relative contribution and understanding other catchment stores. Maybe a general sentence on extending the scope from glaciers to downstream hydrology fits better?

R: Thanks for the suggestion. We agree that a general sentence about broadening the scope is more appropriate than focusing only on non-glacial water sources. In the revised version, we have modified this as suggested: “future studies should address sources of uncertainty not considered in this study (...) and extend the scope from glaciers to downstream hydrology. Downstream hydrology can play a critical role in the seasonal and interannual water release during dry seasons (Drenkhan et al., 2022), attenuating the consequences of glacier shrinkage (e.g., Somers et al., 2019)”

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 detailed review of our manuscript. We appreciate your positive assessment and valuable feedback. We have addressed the issues raised, in particular regarding the coherence of the Discussion and Results sections. Furthermore, we have explicitly emphasised the implications of our findings and their relevance to glaciated regions worldwide. These revisions aim to improve the clarity and comprehensiveness of our work.

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 has been 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.

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

R: Yes, this has been clarified in the introduction: “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 2024*). That will link better with the preceding paragraph and the overall theme of the paper.

R: Thanks for the suggestion. Table 1 has been moved to the Supplementary Material, and we have also added 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 have added 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 has been 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. In the revised version, we have emphasised our assessment within the specified categories of model inputs (by time period, historical and future), 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 has been 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 have been defined in the indicated paragraph. Particular attention has been 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 has been added to the study area section.

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

R: The statement has been 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 has been changed as suggested (“which hosts many small catchments”).

Methods

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

R: The thresholds have been added to the sentence in brackets [0, 2 °C]

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 procedure has been clarified 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 has been added to the new discussion entitled (“Influence of model calibration”).

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

R: The “new value” has been 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” has been removed.

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

R: N°1 refers to the first step. This has been clarified (“If not, a new μ^* is defined and the process starts again from the beginning”).

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

R: The specific sentence has been 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 has been modified as suggested.

Ln 210: “Both alternatives” - consider replacing with the ‘two volume data sources’ or something similar.

R: Thanks for the suggestion. We have replaced “both alternatives” with “two volume data sources” in this paragraph.

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 have replaced “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. The initial pool consisted of all GCMs that have at least one realisation (r1i1p1f1) in all four emissions scenarios (SSP). We have added a sentence to clarify this point: “Considering only GCMs with at least one output in all emission scenarios, 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).”

In the revised version of the discussion, we have also discussed the suitability of the GCM to reproduce seasonal cycles, monthly probabilistic distributions, spatial patterns of climatological means, and the ability of GCMs to reproduce teleconnection responses based on the analysis of Gateño et al. (2024). For more details, see the Editor's comment.

Gateño, F., Mendoza, P.A., Vásquez, N. et al. Screening CMIP6 models for Chile based on past performance and code genealogy. *Climatic Change* 177, 87 (2024). DOI: 10.1007/s10584-024-03742-1

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 has been 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 has been 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 OGM 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 °C 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 has been 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 °C km⁻¹, all reference climates used in this study were able to achieve mean biases of less than 1 °C (Fig. 5 in Aguayo et al. 2024). No downscaling method was used for precipitation (this has been 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 used in this study have been added ($T_{\text{melt}} = -1$ °C, $T_{\text{solid}} = 0$ °C and $T_{\text{liquid}} = 2$ °C).

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 km² (~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” has been 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 have divided 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 has been 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” has been 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 responses, 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 has been 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 has been 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 similar sentences have been rephrased.

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; cited in the main text). 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 be noted that improved outlines and corrections from local inventories in RGI7 also 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 has been 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 have adjusted 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: 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 have been 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 specific sentence has been removed”

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 ($\Delta PP = 0\%$). 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”) has been 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 km² (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 (all panels refer to volume loss), which has been clarified in the specific sentence (volume instead of mass): “Projections from OGGM indicate that the glacier volume loss of recent decades will continue (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 have been 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. The referred differences in NPI-E between contrasting scenarios are not the maximum across hydrological zones (see Table R1). The greatest differences are found in the SPI (e.g., NPI-N and SPI-C) as you noticed. To simplify the discussion in the revised version, we have omitted the specific example of NPI-E and focused instead on broader patterns across hydrological zones (please see section 5.2).

Table 1. Mean specific mass balance for the period 2070-2099 for different hydrological zones.

Hydrological zone	Specific mass balance (kg m ⁻²)	
	SSP1-2.6	SSP5-8.5
PCA	-362	-1524
PPY	-312	-1491
NPI-E	-476	-1712
NPI-W	-1484	-3083
SPI-N	-1828	-4196
SPI-C	-1203	-2914
SPI-S	-794	-1996
GCN	-683	-1643
CDI	-313	-955

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 has been removed as previously mentioned.

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

R: Yes, this has been 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 has been 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, the reference to the panel figure has been corrected.

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.

Ln 455: Remove comma after ‘contribution’.

R: Thanks for the comment. The comma after ‘contribution’ has been 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 has been 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 have been updated in the revised version.

Considering the updated Fig. 10 in the revised version, 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 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. 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. This analysis has been included in Section 5.2 (“Hydrological importance of data uncertainty”)

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 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 have incorporated the suggestions of both reviewers by focusing more on the implications of the results based on current limitations. This approach will improve the clarity and coherence of the discussion, and we hope that it will allow readers to better understand the significance of our results and the current challenges in assessing uncertainty in large-scale glacier models.

The new structure of the discussion in the revised version is as follows:

- **5.1 Hydrological response of Patagonian glaciers to climate change:** This section summarises the regional glacier projections and their associated uncertainties. In this section we have maintained the comparison of glacier runoff with results from previous studies.

- **5.2 Hydrological importance of data uncertainty:** This section links the analysis of sources of uncertainty with their influence on different glacio-hydrological signatures, emphasising spatial differences due to domain characteristics, as suggested by reviewer 2.

- **5.3 Influence of model calibration:** As the calibration approach varies between glacier models and has a direct influence on glacier discharge (through the precipitation factor), we have added a section to discuss this issue, as suggested by Reviewer 1.

- **5.4 Limitations and global implications:** This section 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 have considered your suggestion and incorporated it into our discussion. Specifically, in Section 5.2 (Hydrological importance of data uncertainty), we have analysed the variations across different hydrological zones within the Patagonian Andes (Figs. S3 and S4). Our results show that, despite significant variability in climate, geography and glacier characteristics across the region, the influence of reference climate conditions is limited to specific zones. In the revised version, we have analysed these specific zones and the factors that could explain the greater importance of future sources of uncertainty, such as GCMs and SSPs.

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 have highlighted in Section 4.1.1 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 has been removed and indirectly included in the discussion of the hydrological importance of data uncertainty (Section 5.2): “Despite the larger relative differences in glacier volume than in glacier inventories (11.1% vs. 4.0% of overall difference; Fig. 3 and 4), the selection of glacier inventory was more important for most metrics (Fig. 10)”.

In Ln 488, we are referring to ice thickness estimates. This has been rephrased for clarity (and moved to subsection 5.3): “The use of 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. 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 has been clarified and rephrased: “Furthermore, the generation of large-scale ice thickness datasets requires the compilation of numerous datasets derived from different acquisition dates which hinders regional validation (subsection 5.3)”.

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 has been removed in the revised version.

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 has been removed from the manuscript to avoid adding unnecessary complexity to the study).

Regarding the latitudinal dependence, it's interesting to note that while there are clear latitudinal patterns in precipitation projections (Fig. 6), no clear latitudinal pattern was observed in volume loss (Fig. R1). This suggests that volume loss is influenced by a variety of factors beyond just latitude-dependent variables. Initial volume, ice dynamics and climatological features all play an important role in shaping patterns of ice loss. This underscores the complexity of the processes driving changes in ice volume and highlights the need for a comprehensive understanding of the different sources of uncertainty.

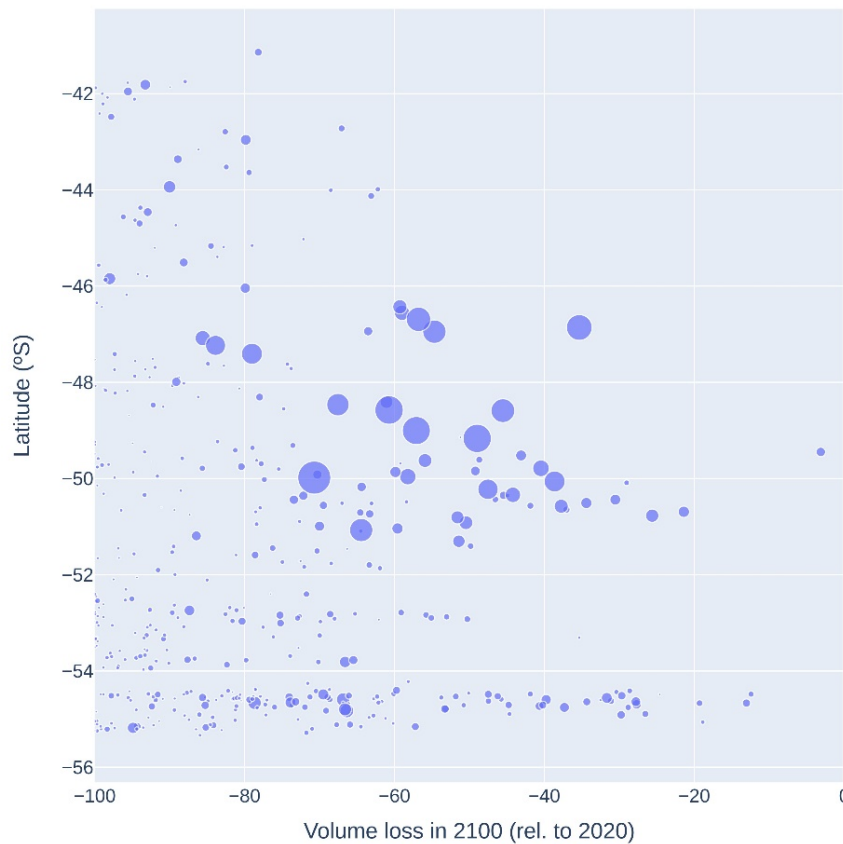


Figure R1. Volume loss (in percentage) vs. Latitude. The size of the circles is proportional to the glacier volume.

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. S5) 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 S5). Nevertheless, we recognise that this statement requires more detailed analysis, given that Rounce et al. account for uncertainty in model parameters, and we have therefore removed this inference from the discussion.

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 summer (June - July).

Conclusions

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

R: We have re-evaluated the main points of the conclusions to ensure a more coherent structure. In the revised version, the main findings of the study are presented in the results section, and the conclusions section focuses on summarising the key findings, with one point per subsection of the Results.

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 have replaced “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 have changed the transparency of the delineation lines in order to make them more prominent. We have also changed 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 suggestion. We have replaced the discrete categories in (b,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 have added 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 have adjusted the range of the colormap from [-100, 100] to [-50, 50] to

have more visible differences in the northern and southern zones. We have also added 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 have also adjusted 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 have used 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 have replaced 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.”)