Review of "**Assessing the glacier projection uncertainties in the Patagonian Andes (40–56°S) from a catchment perspective**" by *Aguayo et al.*

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 glacial-hydrological 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.

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?

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

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.

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

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).

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.

Study area

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

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

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?

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

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

Methods

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

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?

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.

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

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?

Ln 187: Remove "it" after compared.

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

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

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.

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

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

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

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?

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).

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.

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

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?

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.

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?

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?

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.

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?

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?

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.

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.

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?

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.

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?

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

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.

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

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).

Ln 356: Similar patterns in terms of spatial patterns?

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?

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

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.

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

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

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).

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.

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.

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

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

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

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

Ln 455: Remove comma after 'contribution'.

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

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.

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?).

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?

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?

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)?

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

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

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?

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

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)?

Conclusions

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

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.

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.

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).

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.

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.

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

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