Review of "Hydrological response of Andean catchments to recent glacier mass loss" by Caro et al.

This study investigates the changes in glacier mass, area, and runoff for different glaciated catchments in the Southern Andes from 2000-2019. The study uses the Open Global Glacier Model, calibrated with geodetic mass balance data from 2000-2019, and forced by a bias-corrected climate dataset. The focus of the study is on all land-terminating glaciers (i.e., lake- and marine-terminating glaciers are excluded) and specific attention is given to how the changes in climate over 2000-2009 vs. 2010-2019 affect the glacier runoff across the different catchments, which notably span different climatologies. The main conclusions are that most glaciers are losing mass leading to increases in runoff in the Tropical Andes and Dry Andes. Furthermore, results are consistent with previous studies and the glacier contribution to runoff is highest for some catchments in the summer and others in the transition season prior to summer.

While I was excited to read this study, and believe it's an important topic(!), I found the writing to be challenging to follow. The readability of the study would be greatly improved by providing more details and context to many aspects of the methods and results. For example, values were often reported as a change, but it was unclear what the change was related to. Explicit definitions of the glacier runoff and how contributions are calculated would also help. This might also clarify potential issues with comparisons to other studies where different variables or different time periods appeared to be compared. The study also provides a lot of model results for various catchments and various parameters such that paragraphs of the results almost read as a catalogue of the changes with even more details provided in the supplementary material. Making the results a bit more concise and highlighting the key aspects may improve the readability and highlight the novel aspects of the study as well.

I believe the challenges associated with my ability to understand the methods and results greatly affected my review and made it hard for me to understand the novelty of the study and the major gap that this fills in the literature or our advanced understanding of the hydrological response. Contextualizing the results in terms of the impacts (e.g., is a change in runoff a good or bad thing for downstream water availability?) or showing the added value of the bias-corrected climate data may also help highlight some of the novelty. For example, my understanding is the model focused on 2000-2019 because that's when the climate data was available as well as in-situ observations. However, a 20-year time period is fairly short for evaluating the impacts of climate change; therefore, I'm left wondering what is the added value of focusing on this time period? Does it lay the foundation for an improved model to model historic or project future changes? Does it improve the predictive capabilities of seasonal runoff?

Given the lack of novelty and major issues with respect to readability, I believe the manuscript requires major revisions and another round of review. I believe the topic of understanding the hydrological response of glaciers to climate change, especially in an important area like the Southern Andes, is important and warrants publication; however, significant improvements are required and it's unclear to me if the scope of these revisions would be suitable for a typical response to reviewers or if a reject and resubmit would be better. See additional comments below.

General Comments

I found that a good portion of the text lacks context or details that are needed to understand what the authors are stating. For example, the abstract provides good overall numbers, but lacks context such that if the reader begins by reading the abstract it is difficult to understand what the numbers of referring to. Similar challenges in reading the text due to a lack of context were found throughout the text. See specific comments.

I also found the description of the methods to be quite confusing. Figure 1, for example, doesn't have an arrow going from the corrected data to forcing the simulations. The caption also makes it sound like these are two separate workflows (one input data to run the model and another to perform a correction); however, it's unclear to me how these are done separately if the mass balance model is then calibrated after these are performed? I highly recommend modifying the methods section to make things clear. See specific comments below.

More details on how glacier runoff is defined (e.g., fixed vs. moving gauge) and what glacier contribution to runoff means (is this the ratio of the glacier melt to the glacier melt plus precipitation just at the outlet of the glaciers?) are needed. This would help provide context to interpret the results.

The novelty of the study is unclear. For example, the conclusion that most glacierized catchments are losing glacier mass is already known from the observations used to calibrate the model. The third conclusion is that the results are consistent with previous studies. It would be useful for the novel aspects of this study to be included.

Specific Comments

L20 – a bit unclear what "reduced in 93% of the catchments between the periods 2000-2009 and 2010-2019" means. Is this meaning that 7% of the glaciers grew? Is this by area? by number?

L21 – -9% and +0.4 degC compared to what? 2010-2019 vs. 2000-2009? Or compared to a prior year?

L23 – consider adding "calibrated" prior to "melt factors" to make this clear that it's a model parameter.

L25 – specify "contribution" to what? Total runoff in the catchments? Glacier melt vs. glacier runoff?

L27 – "a high mean correlation" between what? Annual mass balance measurements? Discharge?

L29 – "increases" compared to what?

L33 – I assume "largest ice concentration" is referring to glacier area or glacier volume; however, there are many glacierized areas with more area and volume (e.g., Alaska, Arctic

Canada, High Mountain Asia) (Millan et al. 2022). Hence, is this the glacier area divided by catchment area? Please clarify what is meant here as this currently does not appear to be correct.

L34 – Consider changing "provide the water supply" to "supply water"

L44 – Suggest defining what "glacier contribution" refers to. I assume fraction of glacier runoff relative to the total catchment runoff?

L47-48 & L54-55 – There appears to be an implicit changing of definitions of glacier runoff, since the area loss being compensated by the increase mass loss suggests a "moving-gauge" runoff framework (i.e., calculating runoff at the outlet of the glacier as it moves over time), while Huss and Hock (2018) use a "fixed-gauge" runoff framework where they include snow melt and rainfall of off-glacier areas in their comparisons.

L99-103 – "the elevation difference between a glacier inside a TerraClimate grid and the mean TerraClimate grid elevation" is unclear. What part of the glacier was used? The min, mean, max elevation? Or I assume this was used to adjust the temperature for every elevation bin? The cTC_t appears to refer to both the mean glacier elevation and the glacier catchment, which doesn't make sense. Please be specific on how variables are defined and calculated.

L113 – Suggest stating how these were selected. Might be just continuing the sentence with the following criteria; however, unclear as to why 10 glaciers were selected and this analysis wasn't performed over all 18 glaciers? Did something go wrong or is there a lack of confidence in the modeling or data from some glaciers?

L118 – "using a similar methodology as that for …" : this is just a pearson correlation, right? I would recommend stating that to improve readability. I also don't think it saves space by saying similar methodology when the methodology is only a few words.

Equation 2 – is there a reason this comes after the paragraph instead of in line like equation 1?

L119 – Check journal standards, but normally section comes first, i.e., "Section 2.2"

L133-134 – what was quantified? This statement is very vague. I suggest improving readability by being explicitly. For example, "The OGGM toolbox was used to derive the glacier-wide mass balance from the elevation change estimates from Hugonnet et al. (2021)." is very clear about what the Hugonnet product is, what "quantified" refers to, and what processing was done by the toolbox. Otherwise, Hugonnet et al. (2021) quantified mass changes, so very unclear what was actually done.

L150 – consider lake- and marine-terminating as I don't think "fresh terminating" makes sense and often see tide as tidewater glaciers, not tide-terminating.

L152 – state OGGM version here. Also consider changing to "was not included in the version of OGGM used in this study" or something similar since OGGM does have a calving parameterization that could be used.

L154 – what are contradictory variations?

L156 – selected for what? “to represent glaciological regions with different climatic and morphometric characteristics”? Sentence doesn’t make sense as currently written.

L158-159 – if previous hydro-glaciological studies have already quantified the impact, then what is the novelty of this study? Suggest adding something regarding limitations of previous studies here or how the proposed work will be an advance.

L170 – “short description of the” seems unnecessary. Consider just “OGGM” or “OGGM details”.

L172 – “contains enough default input data” is incredibly vague. Do you mean provides pre-processed datasets such as DEMs, glacier hypsometry, glacier flowlines, etc. that can be used to explicitly simulate …?

L174 – Previous text states that the study is done by glacier, not glacier catchment. Results are aggregated to glacier catchments, but as the text currently reads it sounds like the model is running at the glacier catchment.

L176 – Suggest deleting “it is possible” and replacing with what was done.

L178 – How is the geodetic mass balance rate calibrated? Isn’t this input data used for calibration?

L180-182 – Strikes me as odd that the ice thickness inversion for OGGM cites the ice thickness model intercomparison as opposed to Maussion et al. (2019; GMD)

L188 – “where” should be lowercase?

L188 – I found this description of the precipitation/snow difficult to understand. Was the solid precipitation really scaled by the precipitation factor? If so, isn’t this problematic since the TerraClimate data is at a different elevation (previously described) and thus the snow/rain differentiation will be off? Thus, the precipitation factor would be also accounting for the fact that the amount of precipitation that falls as snow increases at higher elevations due to colder climates as well as any biases in the precipitation datasets itself? Please clarify.

L190-192 – Again, I’m confused by how this temperature threshold is performed since it states that 100% of precipitation is classified as snow between 0-2.1 degC and then 0% between 2-4.1 degC. First, the bounds don’t match (i.e., one goes to 2.1 and the other starts at 2). Second, make this clear that this is a model parameter (realized this from Table 1) with supposedly a 2.1 degree range and the linear interpolation varies over the span of two degrees? Perhaps explicitly state the model parameters in a separate sentence too. The way it’s currently written is hard to understand even for someone quite familiar with OGGM.

Table 1 – bottom appears to be cutoff

L199 – Great! Include this information earlier as its clear to understand.

L203-205 – Perhaps I'm misunderstanding the term "model output", but normally I think of model output as a result and thus the result is not calibrated. Rather, the model parameters are calibrated. It's also normally easier to understand "[insert model parameter] is calibrated such that [insert model and observations used]". Please clarify.

L211 – I'm not suggesting redoing the study for something this small, but why wasn't the mass balance output monthly or the model timestep switched from calendar to hydrological years such that the comparison is done properly?

L214 – Suggest referencing this workflow figure earlier in the description where it would be useful to understand the workflow as opposed to it being referenced for the first time once the description is complete.

Figure 1 – I may be misunderstanding the figure, but "observed and simulated MB" do not appear to come from the Corrected climate data. This is rather confusing as my understanding was that the corrected climate data was used to run the MB simulations, no? Glacier melt and rainfall on glacier also appear to be floating in the figure: I believe these are meant to define runoff, but suggest putting them in the runoff box; otherwise, they don't make much sense.

L226 – typically climate change is referring to long-term changes. Here, climate change is being used to refer to changes between two decades. Couldn't differences at this short of time scales be due to interannual climate variability (AMOC, El Nino, etc.)?

L237 – what is meant by amplitude? Isn't it just the mean monthly temperature?

L248-252 – This language is very confusing to follow. Why was only 36% of the glacierized area simulated and not all of the glaciers here, which I thought is what the methods suggested? Furthermore, why does this differ so much from the 85% and 79% stated next? Or is this just meant to state the fraction of the total glacier area in the whole region? If the latter, then this should go in the Methods.

L258 – delete "negative" as the value is shown as a negative number.

L261 – specific wording comment, but Figure 2b technically shows the specific mass balance, not the volume.

Figure 3 – how is the mode shown if these are floating values? Or was the calibration performed in a step-wise fashion? It strikes me as odd that the modes are always a minimum or maximum value indicating the bounds are often what the melt factor reaches for calibration.

L290 – if mentioning sublimation, may be worth noting that sublimation is "implicitly" included in the model, since the observed mass loss is including sublimation while the modeled mass loss

does not. That means that the model parameters are implicitly being calibrated to account for this process.

L293-294 – why not include this where the simulations are initially mentioned in L278, "To test our results …"

Figure 4 – if using an acronym (e.g., G Melt) need to state what it stands for in the caption (even if obvious).

L421-426 – see previous not about comparing fixed-gauge runoff (Huss and Hock 2018) to moving-gauge runoff (this study – this is my assumption as its not particularly clear what happens as the glacier retreats). This information should be provided. If comparing two different ways of estimating runoff, then the comparison isn't really meaningful.

L440 – at first use, it may be useful to mention what this id refers to since it's clearly not RGI which the inventory stated was used.

L466 – typo? "considering a largest glacierized area"? "larger" perhaps?

L464-469 – This suggests that mass balances over two different time periods are being compared? If that's true, this should definitely not be done as the climate forcing may be completely different. Why not model the same time periods (which OGGM can do) and thus be able to do a proper comparison?

L478 – "is" close to?

L479-L481 – sentence does not make sense. Please clarify/rephrase what is being stated.

L492 – temperature index "model" …

L497 – how were all of these differences taken into account? The present study appears to only perform a single set of simulations with one set of model parameters for each glacier. Am I missing something?

L502 – higher "melt" factor values?

L502 & L512-514 – The problem being described seems identical to the overparameterization problem associated with glacier evolution models (e.g., Rounce et al. 2020; JoG). Hence, it might be worth mentioning that this might be what is occurring, i.e., changes in temperature (essentially a temperature bias correction performed using in-situ measurements) are being compensated by changes in the degree-day factor, and thus caution should be taken to avoid overinterpreting calibrated model parameters (especially given that they'll compensate for other factors not accounted for in the model such as changing surface conditions, avalanching, etc. as mentioned in the limitations section). Changing the melt temperature threshold and the rain/snow temperature thresholds by 2 degrees for the Tropical Andes compared to the Dry/Wet Andes

(Table 1) is essentially the exact same thing as using a temperature bias value of 2 degrees (i.e., an additive value that adjusts the temperature by 2 degrees).

L530 – how did the geodetic mass balance define the maximum melting per glacier? Again, this may stem from my misunderstanding of what's actually being calibrated and how the calibration was performed.

L532 – what does a "true seasonal melting distribution" mean? Is this the observations?

L551-554 – Are these percentage increase in the Tropical Andes and Dry Andes referring to the same Inner Tropic and Dry Andes 1 zones or is this statement meant to refer to different zones? Please clarify as this is unclear.