

Author Response to RC1

Review of Shutkin et al. 'Modeling the impacts of climate trends and lake formation on the retreat of a tropical Andean glacier (1962-2020)'.

Summary

Shutkin et al. have examined the past behaviour of Queshque Glacier, a glacier found in the Cordillera Blanca of Peru. They use observations of the recession of the glacier between 1962 and 2008 and several other datasets relating to glacier volume and dynamics, as well as local climate, to calibrate the Open Global Glacier Model (OGGM), which they then use to simulate contemporary glacier behaviour. They analyse the variability of OGGM simulations in relation to input data to establish model sensitivity to input parameters. Despite their use of a simplified temperature-index model, which the authors acknowledge has well established limitations, their findings show how they are able to replicate characteristic behaviour of Queshque Glacier in its climatologically complex setting. The authors also illustrate the impact of the transition of the glacier from land- to lake-terminating during the study period, suggesting that this process now largely dictates the glaciers mass loss trajectory over climate.

Overall, the paper is well structured and written and the authors have constructed their study in a comprehensive manner. The main findings are of relevance across the field of tropical glaciers and it is good to see that additional work is now being done to incorporate ice loss processes experienced by lake-terminating glaciers, which are prevalent across the Cordillera Blanca and, increasingly, worldwide.

Major comments:

We believe the study currently has one main limitation which we suggest requires revisiting – that is the approach employed to generate one of the two DEMs used to estimate glacier mass balance, and the subsequent treatment of elevation change data. The 1962 DEM produced by the authors has been generated using a manual photogrammetric plotter, which is methodologically dated and has produced a result which is very contrasting in data quality to their 2008 lidar derived DEM. Given the availability of software specifically designed to process optical stereo imagery to produce high-quality DEMs (example recommended later in the review) we would suggest reprocessing of the 1962 imagery should be attempted, if the authors have access to the images, to bring that dataset in line with the 2008 lidar DEM. This reprocessing would ensure that any biases associated with DEM difference data derived from methodologically contrasting DEMs (e.g. <https://xdem.readthedocs.io/en/stable/>) are minimised. Similarly, the subsequent treatment of elevation change data derived from this current DEM pair is lacking rigor, namely consideration of outlier identification and removal and subsequent gap filling, which are both required prior to geodetic mass balance estimation (Piermattei et al., 2024). Finally, the authors make no attempt to calculate the uncertainty associated with their glacier mass balance data, on which the rest of their analyses is based. This needs revisiting.

Considering the above, we recommend that major revisions are needed before the paper can be considered for publication.

Author Response

We thank the review team for their constructive feedback. Please find our detailed responses below, which are also reflected in a revised manuscript and supplement. The reviewers' major comments concerned the 1962 DEM quality, DEM outlier identification, and treatment of uncertainty in the difference of DEMs. We respond to these issues in logical order below before addressing the minor comments.

1. 1962 DEM Quality

We appreciate the concern that methodological differences in DEM construction may impact their intercomparison and potentially introduce biases into the DEM difference map. Although we no longer have access to all metadata required for using the most state-of-the-art photogrammetric software, we note that the same DEM was used previously for a similar analysis (Mark & Seltzer, 2005). Nonetheless, we have conducted multiple comparisons indicating that our DEM is of acceptable quality.

First, we have accessed a 10 m resolution DEM constructed from the same 1962 stereo imagery using ERDAS Leica Photogrammetry Suite version 11. This DEM was used for a similar analysis of Queshque Glacier and others in the region (Huh et al., 2017). A comparison between Huh et al.'s (2017) DEM and the one used in the present study shows considerable quality differences favoring our choice of DEM. We attribute quality concerns in the latter DEM to extremely low contrast in much of the accumulation zone that hampered the effectiveness of the DEM generation software. This resulted in obviously unnatural terrain artifacts that are absent from our chosen DEM product. On this basis, and on the basis that our data product has already been accepted for publication in reputable journals, we believe that while more state-of-the-art photogrammetry software could improve our DEM, this is not guaranteed.

Second, we compare differences between our 2008 and 1962 DEMs over stable terrain to quantify the resulting uncertainty in elevation change over the glacier (see below).

2. Geodetic Mass Balance Uncertainty

If we safely assume that the 2008 LiDAR is of much higher quality than the 1962 DEM, then differences in the elevation of stable terrain between the two data products can be attributed to artifacts or inaccuracies in the earlier product.

As described in the supplemental material, we aligned the 1962 DEM to that of 2008 using a 3-dimensional coregistration process. We have subsequently compared our methodology to the common Nuth & Kääb algorithm (Nuth & Kääb, 2011). Comparison of residual error over stable ground indicates that our results are more robust than those accomplished using the methods from Nuth & Kääb. We have updated Fig. S3 to highlight this result and reproduce it here as Fig. 1.

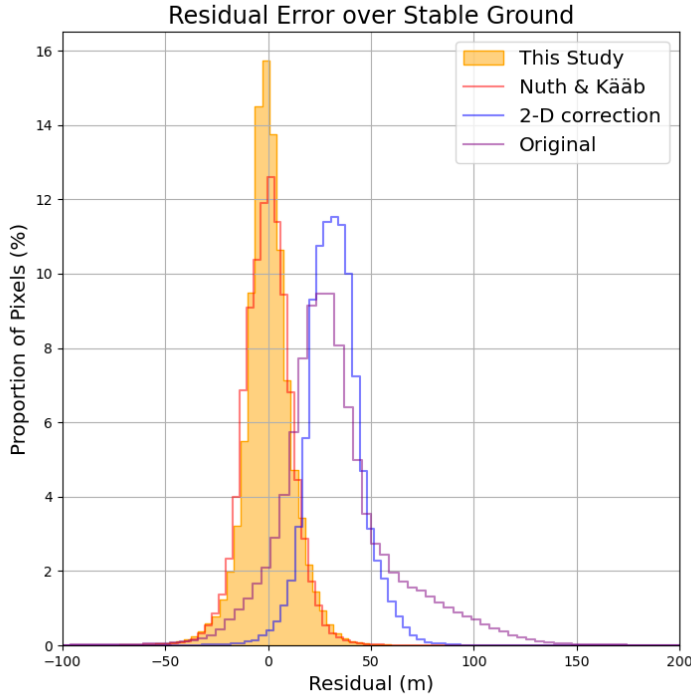


Fig. 1: Difference of DEMs residual error over stable ground at different stages and using different methods of coregistration.

Following the reviewers' recommendation, we use the xDEM python package (xDEM contributors, 2024) to evaluate the uncertainty of our geodetic mass balance. Having minimized systematic error through DEM coregistration, we follow Hugonnet et al. (2022) to evaluate random error over stable terrain then infer uncertainty in elevation change over the glacier. Random error is quantified by considering both the heteroscedasticity and spatial correlation of error. Heteroscedasticity is evaluated across gradients of DEM slope and curvature, calculated using methods from Horn (1981) and Zevenbergen & Thorne (1987), respectively. The spatial correlation of error is estimated by an empirical variogram using Dowd's estimator (Dowd, 1984). The uncertainty in elevation change within the 2008 glacier boundary is then calculated as the average pairwise product of pixel uncertainties times the spatial correlation of error between each two pixels (Hugonnet et al., 2022 eqs. 17-19).

Following these methods, we arrive at a mean elevation change of -22.61 ± 0.81 m across the glacier (Fig. 2). Maintaining the density assumption of 900 kg m^{-3} as used throughout the manuscript, this translates to a geodetic mass balance of $-442 \pm 16 \text{ mm w.e. a}^{-1}$. The change in elevation across the entire DEM including stable and unstable terrain is shown in Fig. 2. We consider the level of uncertainty arrived at using these methods to be acceptable for the purpose of our study. Since most error over stable ground (and therefore elevation change uncertainty) is attributable to artifacts and errors in the 1962 DEM, we also consider the acceptable uncertainty range as testament to the adequacy of our 1962 dataset.

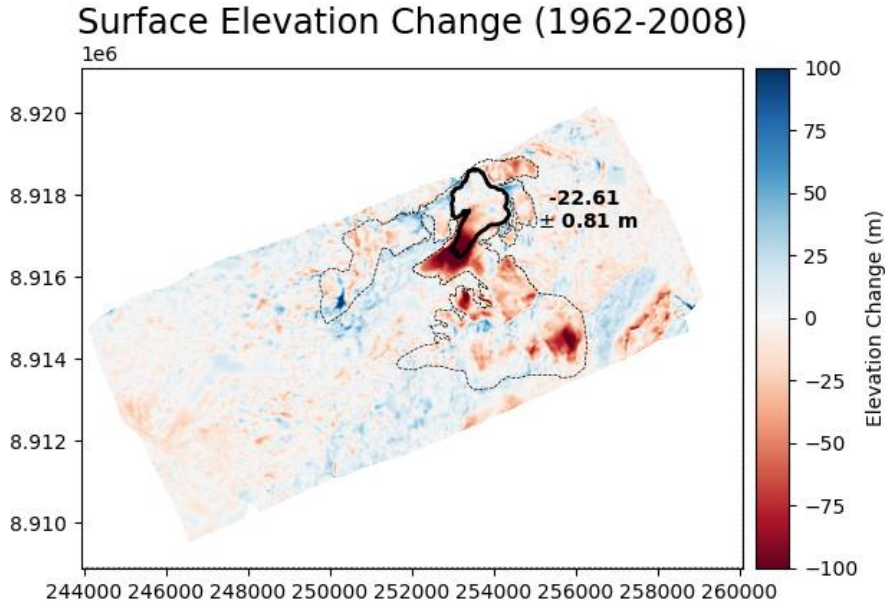


Fig 2: Difference of DEMs (1962-2008) shows significant surface height change across the glacier ablation zone. The 2008 glacier boundary is outlined in bold and unstable terrain (including the 1962 glacier boundary) is delineated by a dashed line. Note that additional terrain above 5000 m in 2008 was also considered to be unstable.

3. DEM Outlier Identification

We have assessed the presence of outliers (95th percentile) in the difference of DEMs between 1962 and 2008. These results are compared to the DEM difference used in the original submission.

The outlier detection procedure is as follows:

1. The data were binned according to their positioning in 50 altitude bins (~16 m) according to the 2008 DEM.
2. Pixel values for surface height change were compared to the mean value of each altitude bin. Pixel values with z-score absolute values greater than 1.96 (two-tailed 95th percentile) were considered to be outliers resulting from DEM or coregistration errors and were replaced by the mean value from the appropriate altitude bin.
3. The resulting map of change in surface height was used to recalculate the specific (area averaged) mass balance across the entire glacier surface. These results were compared to the original value used.

After removing outliers, the new SMB was calculated to be -435 mm w.e. per year, a 1.5% positive change from the originally published figure. Maps depicting the original DEM difference, altitude-binned averages, and detected outliers are included below (Fig. 3).

We note that the difference in specific mass balance is within the uncertainty window estimated in the previous step and that it is difficult to distinguish between outliers caused by map artifacts

versus extreme natural phenomena. Furthermore, previous glaciological studies using outlier detection and gap filling have operated with very different data constraints. For example, Piermattei et al. (2024) use ASTER and TanDEM-X data, both of which are known to contain artifacts and data voids in mountainous regions. Geodetic mass balance estimation using these global datasets therefore may require outlier correction. In our case using local datasets without issues such as cloud cover, this requirement is less apparent. Given this ambiguity and the negligible impact of outlier correction on the resulting geodetic mass balance estimation when compared to the overall uncertainty, we believe that our original figure is of sufficient quality. We thank the reviewers for their methodological recommendations which have allowed us to more rigorously defend this claim.

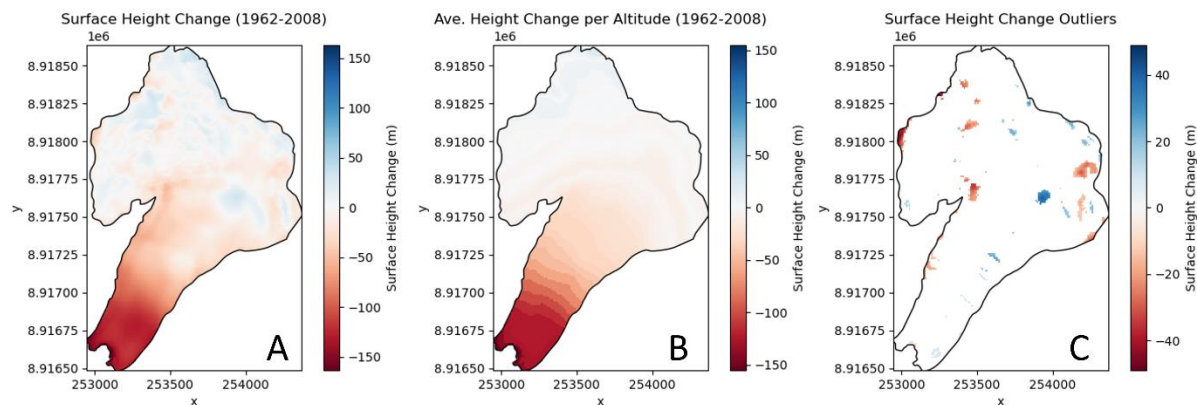


Fig 3: Outlier detection began from the map of elevation difference (a), which was then binned into 50 elevation bands (b). Pixel values exceeding 2σ deviation from the elevation band mean were considered to be outliers (c).

More minor suggested amendments:

L57: perhaps 'dictate' rather than 'direct' retreat patterns?

Accepted

L95: Refer to Fig 1 here.

Accepted

L95: It would be easy and useful to visualise this increased recession after the transition of the terminus type from land- to lake-terminating. Repeat mapping of terminus position from optical imagery at timesteps a few years apart should yield a nice set of ice front position estimates and the distance between them should show the increase in recession rate related to terminus type transition? Landsat 7 and 8 images would capture this well.

We have conducted the suggested analysis using Landsat 8 and Sentinel 2 images and will include an additional supplemental figure demonstrating accelerated retreat during the period of lake calving (Fig. 4). It must be noted, however, that the absence of annual imagery earlier in the timeseries precludes an assessment of normal variability in the actual retreat rate. Furthermore, our decision to use the elevation band flowline approach in OGGM limits the model's accuracy

when it comes to glacier length. The linear retreat rate is therefore an inferior validation metric, but does still yield the useful insight that retreat greatly accelerated as frontal ablation developed.

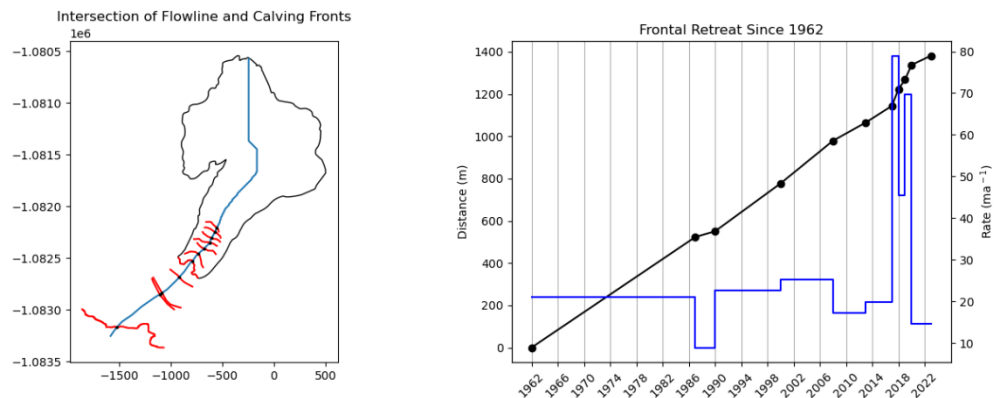


Fig 4: (left) Positions of the glacier terminus mapped from aerial, Landsat 8, and Sentinel 2 images from 1962-2023. (right) The cumulative retreat of Qeshque Glacier (black) with mapped years plotted as points. The retreat rate, or slope of the black line, is plotted in blue.

3.1 We have a few suggestions here:

-We'd really encourage the authors to provide an illustration of the elevation change data they have derived from their respective DEMs across the full study area. This is needed to provide the reader with an indication of the overall quality of the DEMs and the presence or absence of any biases within the derived elevation change data. It'd also provide a powerful illustration of the changes the glacier has experienced.

We provide a map showing the residual differences on stable and unstable terrain between the 1962 and 2008 DEMs with the 1962 and 2008 glacier boundaries outlined in black:

Stable ground was considered to exist off glacier and at altitudes below 5000 m. This second exclusion accounted for resolution differences that produced some large errors over steep ridgelines. As seen in Fig. 1, residuals over stable ground are normally distributed with a mean of approximately 0 m and a standard deviation of about 12 m. Systematic (e.g., aspect-related) bias is minimized, though may be evident locally near the southeast corner of the model domain. Note that this localized issue was persistent across coregistration methods, including when using the Nuth and Kääb (2011) algorithm in the xdem Python package. The large positive residual situated off glacier is due to an artifact in the 1962 DEM wherein a peak is represented as a plateau. This peak is located above 5000 m, however, so did not bias the DEM coregistration process.

-I'd ask the authors to consider the impact of outliers in the elevation change data, which they don't currently mention. As the two DEMs have been generated using very different approaches, there will be outliers which do not represent real elevation change. These should be removed considering the statistical characteristics of elevation change data within a similar altitudinal band (e.g. Gardelle et al., 2013). Over glacier surfaces, values more typical of the elevation change experienced by the glacier within the same altitudinal band should then be used to fill the resulting gaps (e.g. using <https://xdem.readthedocs.io/en/stable/>), prior to mass balance calculation.

See response in “Major Comments” section above.

-If the authors have access to the 1962 images in their original form, I'd encourage them to explore the possibility of generating a DEM using photogrammetric software now readily available online (e.g. CATALYST <https://catalyst.earth/>, user-friendly tutorials are available online and a fully functional 7-day trial can be acquired online). The techniques used to generate the two DEMs used to calculate elevation change couldn't currently be more contrasting and various local and broad scale biases could be present as a result, which the reader can't currently see without the data being shown.

See response in “Major Comments” section above.

-The approach to DEM coregistration seems logical and robust and the figures in the supplement suggest good agreement between the DEMs, but a map of elevation change over the glacier and surrounding areas really is needed to confirm this.

See response in “Major Comments” section above.

-The uncertainty associated with the mass balance estimate on which the rest of the modelling is based does not seem to have been considered at all. There are multiple sources of error in the technique the authors have employed (Hugonnet et al., 2022) which can bias the mass balance towards higher/lower overall ice loss. This certainly needs to be estimated to reassure the reader that the mass balance signal is realistic and beyond the level of uncertainty.

See response in “Major Comments” section above.

L134: How many of these point measurements were used to evaluate error? Where were they located? How does the difference between derived and point based measurements vary spatially?

Only 17 georeferenced point measurements were provided by the 2009 GPR survey report. As elaborated upon in the main text, the points show general consistency with the subsequent 2014 survey (Fig. 5a). The 2009 GPR points span from the bottom of the glacier in the southwest towards the center of the glacier in the northeast of Fig. 5b. The points are located approximately along the centerline of the glacier and are each in proximity to multiple measurements from the subsequent GPR survey. There is no apparent relationship between the mean distance from the 2009 data points to their respective nearest neighbors and the resulting difference between measured and derived thickness. There does, however, appear to be a slight spatial bias, with derived thicknesses being more likely to underestimate the 2009 measurements at lower elevations. The significant outlier where the derived thickness is approximately 28 m thinner than observed occurs at a discontinuity in the 2009 survey, suggesting that the technician may have needed to navigate an obstacle which may have produced abnormalities in the ice thickness profile or potentially caused an error in measurement or radargram interpretation.

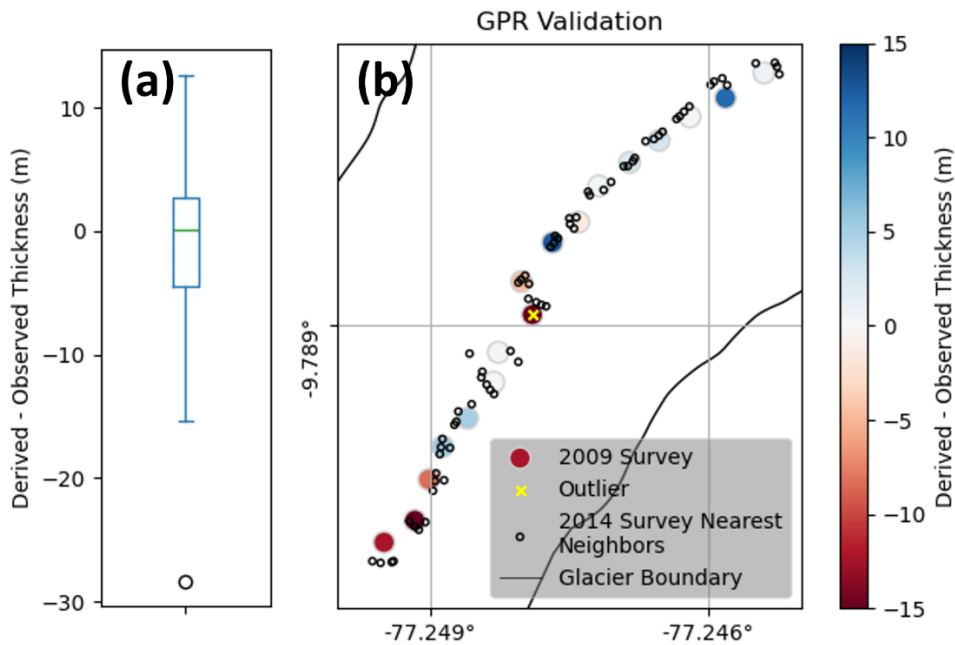


Fig. 5: (a) Box and whiskers plot showing the distribution of error between the derived and observed thickness values, including the position of the single negative outlier. (b) A map of the 2009 GPR survey showing the error calculated at each point.

Line 156: It might be useful to provide a summary figure of the ‘average’ of the climate data used as input to the temperature index model, as much of the discussion is focused on seasonality (or lack thereof) later in the paper. It would help the reader relate the simulated accumulation and ablation (Fig. 3) to the climate the glacier experiences.

We agree that this would be useful for the reader. We provide Fig. 6 below, which will be included in a revised manuscript.

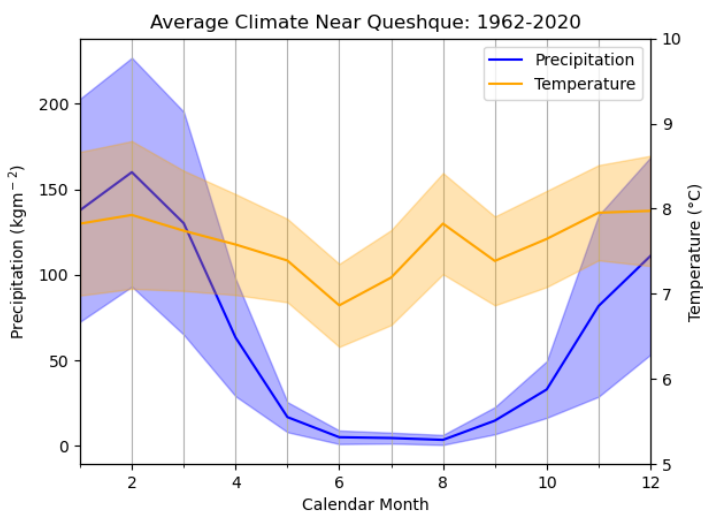


Fig. 6: Climatology near Queshque Glacier during the period 1962 through 2020. Mean temperature (2 m) and precipitation are depicted with 1σ bounds as shaded regions. Climatology data combine CRU (Harris et al., 2014; New et al., 2002) and PISCO (Aybar et al., 2020; Huerta et al., 2018) products. Note that the values shown here are subsequently adjusted by the precipitation factor and temperature bias within the mass balance model.

Line 163: ‘lapse rate of -6.5°C ’, while this is the global average, was there any testing on using a different value? The tropical Andes can have lapse rates can be lower than -6.5°C , maybe as low as -3.5°C (Navarro-Serrano et al., 2020). This could change the temperature index substantially I suspect.

Lapse rates in the tropical Andes are a critical, yet highly uncertain parameter in tropical glaciological studies. Our group’s sensor network in the Cordillera Blanca, as well as atmospheric modeling using the WRF, show that regional lapse rates are seasonally variable, increasing in magnitude during the dry season. Measured lapse rates vary from $\sim -9.1^{\circ}\text{C km}^{-1}$ to $\sim -6.0^{\circ}\text{C km}^{-1}$ between seasons, while modeled lapse rates varied from $\sim -7.5^{\circ}\text{C km}^{-1}$ to $\sim -5.8^{\circ}\text{C km}^{-1}$ (Hellström et al., 2017). One limitation of our model is that it cannot incorporate seasonal lapse rate variability during the mass balance model calibration, despite this playing a potentially crucial role in the tropical Andes. A compromise between the measured and modeled seasonal extremes was therefore selected and we opted for the conventional $-6.5^{\circ}\text{C km}^{-1}$ for the sake of consistency and comparability with other studies.

We will clarify this point in the text and address it further in a new limitations section.

Table 1: For the water level, the source could not be Sentinel 2 (optical imagery). The source would be the DEM you used, which needs specifying.

We identified the position of the lake boundary using Sentinel 2 imagery then adopted a water level from the 2008 DEM using this position. This can be clarified in the text and in Table 1.

L236: This description of how the climate data were analysed belongs more in section 3.3 I feel. If it is moved up, it should also be bolstered by citation to appropriate literature to confirm that this is a standard approach to processing these datasets.

We accept moving this description into section 3.3. The detrending method used is a common practice for removing the multidecadal trend from a climatological dataset and converting the timeseries into anomalies from this trend (Wu et al., 2007). This is a necessary step for comparing the climatology to detrended indices like the SOI or ONI. We further standardize the data using the common standard anomaly approach by dividing absolute anomalies by the timeseries standard deviation (e.g., Dabernig et al., 2017). This facilitates comparison between datasets that oscillate on varying orders of magnitude.

Table 2: The authors state ‘low sensitivity’ and ‘high sensitivity’ here, but it is only briefly mentioned in the text. The authors could be more specific on why you have assigned certain model runs as low/high sensitivity. This sensitivity could also be added in the table caption. The mention of the sensitivity from the model runs is only sparsely mentioned in the discussion.

Furthermore, if the authors are not varying the Precipitation Factor, is it necessary to place it in Table 2? Would it have been better placed in table 1 as a ‘constant’?

Here “sensitivity” refers to the magnitude of the temperature sensitivity parameter, which dictates the ablation response to a unit change in temperature. We will clarify this point throughout the text and discuss this parameter’s relation to the temperature bias more thoroughly in the discussion.

We agree that the precipitation factor should be moved to Table 1.

Table 3: The authors have used different names for their ‘sensitivity’ models. First, they were ‘low – high sensitivity,’ now they are ‘least – most climate sensitive’. If these contrastingly named model runs are actually the same, the naming needs to be consistent, or if they are different, a section explaining how the experiments were conducted would be useful.

The “low” and “high” sensitivities indicated in Fig. 3 (main text) represent models number 2 and 5 from Table 3, whereas models 1 and 6 are indeed the least and most temperature sensitive. We recognize the confusion this may have caused and will be sure to clarify our nomenclature in the revised manuscript.

Section 5.1 The authors may be able to bolster this section by comparing their modelled glacier mass balance evolution against available measurements of glacier mass balance over the period 2000-2019 (Hugonnet et al., 2021). According to the dataset of Hugonnet et al. (2021), the mass balance of RGI 16.02060 (Queshque Glacier) was -0.59 m w.e.a-1 from 2000-05, -0.73 m w.e.a-1 from 2005-10, -0.83 m w.e.a-1 from 2010-15 and -0.88 m w.e.a-1 from 2015-2019. These estimates provide a point of comparison for the authors modelled results and could also be discussed alongside the climate variables the authors have analysed.

We have compared our ensemble mean specific mass balance (SMB) model outputs for each of the 42 epochs provided in the Hugonnet et al. (2021) study. We find that averaged across the years of a given epoch, our ensemble consistently overestimates mass loss as compared to the Hugonnet et al. study. In other words, our model results suggest that Queshque Glacier is retreating faster than the best estimate from the global geodetic mass balance study.

With this said, our results are within the uncertainty bounds provided by Hugonnet et al. during all but 3 of the 42 epochs. All three of these epochs consider change as of 2020 including 2000-2020, 2015-2020, and 2016-2020. This suggests that a systematic bias exists either in the 2020 data used in taking the geodetic mass balance or in our simulation occurring around that time. Other epochs up to 10 years in duration (the second longest duration behind the single 20-year measurement) show agreement between Hugonnet et al.’s and our own data. In summary, this comprehensive comparison indicates general agreement between the two datasets, bolstering confidence in our mass balance simulation.

Figure 5: It might be good to add a second panel to this figure to illustrate where the centreline (assuming these are centreline velocities) of the glacier runs in these modelled velocity profiles.

The elevation band flowline approach used in this study represents glaciological variables as mean altitude-binned values. The modeled and observed velocities from Fig. 5 (main text) each represent such mean values.

We will clarify this point in the figure caption.

Section beginning L334: A good section acknowledging the limitations of the applied approaches.

We appreciate this feedback and intend to expand our limitations section considering the other points made by the reviewers.

L405: It may be worth considering the findings of Malles et al. (2023) in the discussion of the impact of lake development on the studied glacier. Malles et al. (2023) also establish changes in the sensitivity of glaciers to climate following the introduction of glacier-lake interactions to the same model (OGGM), so their findings may well support the inferences made by the authors here.

Malles et al. (2023) presents an interesting study on the effect that considering frontal ablation during OGGM's mass balance calibration has on projected glacier meltwater contributions to sea level rise. The authors find that considering this additional process reduces the overall projected contributions of tidewater glaciers to 21st century sea level rise. While on first glance this appears contrary to our conclusions and those of various studies of lake terminating glaciers (e.g., King et al., 2018), this is not the case. As stated in Malles et al. (2023), the modeled reduction in mass loss above sea level when calving is included "is due to the lowering of the sensitivity to atmospheric temperatures..." which is the logical outcome of calibrating a temperature-index model using mass balance data that are already to some extent decoupled from climate. If one assumes that all observed mass loss relates to changes in the climatic mass balance, then any mass loss due to complex calving processes will be falsely attributed to a change in temperature within the OGGM framework. Consequently, projections under warming conditions will overestimate mass loss, as Malles et al. (2023) observe.

In the case of our study, we calibrate OGGM using observations spanning a long period throughout most of which the glacier was not calving. We can therefore be more confident that our calibration data relate directly to the climatology. We therefore feel it would be inappropriate to draw comparisons with the Malles et al. study.

References cited:

Gardelle, J., Berthier, E., Arnaud, Y., and Kääb, A.: Region-wide glacier mass balances over the Pamir-Karakoram-Himalaya during 1999–2011, *The Cryosphere*, 7, 1263–1286, <https://doi.org/10.5194/tc-7-1263-2013>, 2013.

Hugonnet, R., McNabb, R., Berthier, E. et al. Accelerated global glacier mass loss in the early twenty-first century. *Nature* **592**, 726–731 (2021). <https://doi.org/10.1038/s41586-021-03436-z>

R. Hugonnet et al., "Uncertainty Analysis of Digital Elevation Models by Spatial Inference From Stable Terrain," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 6456–6472, 2022, doi: 10.1109/JSTARS.2022.3188922.

Malles J-H, Maussion F, Ultee L, Kochtitzky W, Copland L, Marzeion B. Exploring the impact of a frontal ablation parameterization on projected 21st-century mass change for Northern Hemisphere glaciers. *Journal of Glaciology*. 2023;69(277):1317-1332. doi:10.1017/jog.2023.19

Navarro-Serrano, F., López-Moreno, J. I., Domínguez-Castro, F., Alonso-González, E., Azorin-Molina, C., El-Kenawy, A., et al. (2020). Maximum and minimum air temperature lapse rates in the Andean region of Ecuador and Peru. *Int. J. Climatol.* 40, 6150–6168. doi: 10.1002/joc.6574

Piermattei, L., Zemp, M., Sommer, C., Brun, F., Braun, M. H., Andreassen, L. M., Belart, J. M. C., Berthier, E., Bhattacharya, A., Boehm Vock, L., Bolch, T., Dehecq, A., Dussailant, I., Falaschi, D., Florentine, C., Floricioiu, D., Ginzler, C., Guillet, G., Hugonnet, R., Huss, M., Kääb, A., King, O., Klug, C., Knuth, F., Krieger, L., La Frenierre, J., McNabb, R., McNeil, C., Prinz, R., Sass, L., Seehaus, T., Shean, D., Treichler, D., Wendt, A., and Yang, R.: Observing glacier elevation changes from spaceborne optical and radar sensors – an inter-comparison experiment using ASTER and TanDEM-X data, *The Cryosphere*, 18, 3195–3230, <https://doi.org/10.5194/tc-18-3195-2024>, 2024.

Additional References Cited in Author Response

Aybar, C., Fernández, C., Huerta, A., Lavado, W., Vega, F., & Felipe-Obando, O. (2020). Construction of a high-resolution gridded rainfall dataset for Peru from 1981 to the present day. *Hydrological Sciences Journal*, 65(5), 770–785. <https://doi.org/10.1080/02626667.2019.1649411>

Dabernig, M., Mayr, G. J., Messner, J. W., & Zeileis, A. (2017). Spatial ensemble post-processing with standardized anomalies. *Quarterly Journal of the Royal Meteorological Society*, 143(703), 909–916. <https://doi.org/10.1002/qj.2975>

Dowd, P. A. (1984). The Variogram and Kriging: Robust and Resistant Estimators. In G. Verly, M. David, A. G. Journel, & A. Marechal (Eds.), *Geostatistics for Natural Resources Characterization: Part 1* (pp. 91–106). Springer Netherlands. https://doi.org/10.1007/978-94-009-3699-7_6

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *International Journal of Climatology*, 34(3), 623–642. <https://doi.org/10.1002/joc.3711>

Hellström, R. Å., Fernández, A., Mark, B. G., Michael Covert, J., Rapre, A. C., & Gomez, R. J. (2017). Incorporating Autonomous Sensors and Climate Modeling to Gain Insight into Seasonal Hydrometeorological Processes within a Tropical Glacierized Valley. *Annals of the American Association of Geographers*, 107(2), 260–273. <https://doi.org/10.1080/24694452.2016.1232615>

Horn, B. K. P. (1981). Hill shading and the reflectance map. *Proceedings of the IEEE*, 69(1), 14–47. <https://doi.org/10.1109/PROC.1981.11918>

Huerta, A., Aybar, C., & Lavado-Casimiro, W. (2018). *SENAMHI PISCO temperatura versión 1.1 (PISCOt v1.1)*. SENAMHI - DHI. <https://iridl.ldeo.columbia.edu/SOURCES/.SENAMHI/.HSR/.PISCO/.Temp/>

Hugonnet, R., Brun, F., Berthier, E., Dehecq, A., Mannerfelt, E. S., Eckert, N., & Farinotti, D. (2022). Uncertainty Analysis of Digital Elevation Models by Spatial Inference From Stable Terrain. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 6456–6472. <https://doi.org/10.1109/JSTARS.2022.3188922>

Huh, K. I., Mark, B. G., Ahn, Y., & Hopkinson, C. (2017). Volume change of tropical Peruvian glaciers from multi-temporal digital elevation models and volume–surface area scaling. *Geografiska Annaler: Series A, Physical Geography*, 99(3), 222–239. <https://doi.org/10.1080/04353676.2017.1313095>

- King, O., Dehecq, A., Quincey, D., & Carrivick, J. (2018). Contrasting geometric and dynamic evolution of lake and land-terminating glaciers in the central Himalaya. *Global and Planetary Change*, 167, 46–60. <https://doi.org/10.1016/j.gloplacha.2018.05.006>
- Mark, B. G., & Seltzer, G. O. (2005). Evaluation of recent glacier recession in the Cordillera Blanca, Peru (AD 1962–1999): Spatial distribution of mass loss and climatic forcing. *Quaternary Science Reviews*, 24(20), 2265–2280. <https://doi.org/10.1016/j.quascirev.2005.01.003>
- New, M., Lister, D., Hulme, M., & Makin, I. (2002). A high-resolution data set of surface climate over global land areas. *Climate Research*, 21(1), 1–25. <https://doi.org/10.3354/cr021001>
- Nuth, C., & Kääb, A. (2011). Co-registration and bias corrections of satellite elevation data sets for quantifying glacier thickness change. *The Cryosphere*, 5(1), 271–290. <https://doi.org/10.5194/tc-5-271-2011>
- Piermattei, L., Zemp, M., Sommer, C., Brun, F., Braun, M. H., Andreassen, L. M., Belart, J. M. C., Berthier, E., Bhattacharya, A., Boehm Vock, L., Bolch, T., Dehecq, A., Dussailant, I., Falaschi, D., Florentine, C., Floricioiu, D., Ginzler, C., Guillet, G., Hugonnet, R., ... Yang, R. (2024). Observing glacier elevation changes from spaceborne optical and radar sensors – an inter-comparison experiment using ASTER and TanDEM-X data. *The Cryosphere*, 18(7), 3195–3230. <https://doi.org/10.5194/tc-18-3195-2024>
- Wu, Z., Huang, N. E., Long, S. R., & Peng, C.-K. (2007). On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proceedings of the National Academy of Sciences of the United States of America*, 104(38), 14889–14894. <https://doi.org/10.1073/pnas.0701020104>
- xDEM contributors. (2024). *xDEM* (Version v0.1.0) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.11492983>
- Zevenbergen, L. W., & Thorne, C. R. (1987). Quantitative analysis of land surface topography. *Earth Surface Processes and Landforms*, 12(1), 47–56. <https://doi.org/10.1002/esp.3290120107>

Author Response to RC2

We thank the reviewer for her constructive feedback. Please find our detailed responses below.

Review of Shutkin et al. Modeling the impacts of climate trends and lake formation on the retreat of a tropical Andean glacier (1962-2020) submitted to The Cryosphere

Shutkin et al. present OGGM simulations of Queshque Glacier in the Peruvian Andes. The model is calibrated through comparison with measured mass balance from 1962-2008, with the dynamic glacier simulations run from 2008-2020. They run a range of simulations based on different temperature biases, which influence the other calibrated parameters, with model results compared against observed glacier velocities. They also show results obtained through maintaining a constant temperature or precipitation while allowing the other parameter to vary (to determine the drivers of historical changes), as well as a test comparing the influence of including frontal ablation on modelled glacier volumes.

Overall it is good to see a study such as this which is able to distinguish the long term drivers of glacier change, through these modelling scenarios. The results related to the temperature/precipitation changes are clear and interesting, as is the result on the impact of including frontal ablation. The correlations with ENSO are also quite instructive and add to the growing literature on the impact on ENSO conditions on glacier change in this region. In general the paper is clearly written and has an appropriate structure.

I do though have a couple of more major points to address:

1. Model validation/confirmation

In the paper I only see direct comparison of the models with the observed glacier velocity (Figure 5). However, this comparison is not clearly quantified aside from the comparison in the figure. As well as better quantification in terms of the glacier velocities, it would also be useful to add any other validation data that may be available. For instance, the calving front positions (which are shown in Figure 1a) could be added into Figure 6. It might also be good to compare with any mass balance measurements which are available, ideally from the field, but if not the dataset of Hugonnet et al. (2021) might also be beneficial. I think this would allow the authors to better determine which of their model runs is most likely to represent conditions correctly and give confidence that the parameters in that run are reasonable.

The reviewer's major comments categorized under "Model validation/confirmation" concern quantification of the surface velocity validation, the inclusion of additional mass balance and terminus position validation datasets, and an assessment of relative model strength across the parameterizations used throughout the paper. We address these concerns below.

a) Surface Velocity

We have quantified the error of modelled altitude band average surface velocity during 2018 against measurements (Millan et al., 2022) derived from feature tracking on satellite imagery from the same period. The results suggest that in terms of reproducing velocity, models with temperature biases of

-7.0 through -8.0 °C (models 2-4 in Table 2 of main text) perform with the highest accuracy. This is in visual agreement with Fig. 5 (main text). These results are presented in Table 1 below:

Model Number	Model Temp Bias (°C)	RMSE (ma ⁻¹)	MAE (ma ⁻¹)
1	-6.5	9.8	4.9
2	-7.0	3.5	2.5
3	-7.5	2.7	2.2
4	-8.0	3.8	3.2
5	-8.5	5.6	5.2
6	-9.0	7.6	6.9

Table 1: RMSE and MAE values of elevation band flowline surface velocities versus observations from Millan et al. (2022).

b) Glacier Terminus Positions

Direct comparison of glacier terminus position (glacier length) to satellite imagery is complicated by the elevation band flowline method selected in this study. This method was chosen to facilitate comparison between observed and modeled surface velocities. However, the decision to use the elevation band flowline approach in OGGM limits the model's accuracy when it comes to glacier length (see OGGM documentation: <https://docs.oggm.org/en/latest/flowlines.html#elevation-bands-flowlines>).

Nonetheless, we have conducted the suggested analysis using Landsat 8, Sentinel 2, and historical aerial images (Fig. 1). To circumnavigate uncertainties surrounding model glacier length, we have instead opted to use the elevation of the 2008 DEM where it intersects with each terminus position shown in Fig. 1. This is a more reliable approach for aligning modeled and observed glacier positions, as the elevation band flowline is built around the 2008 DEM. We use the Zonal Statistics as Table tool in ArcGIS Pro to extract mean and standard deviation elevations from the 2008 DEM where it intersects with mapped termini. The resulting values are depicted in Fig. 2. Note, however, that these calculated elevations reflect the surface of the glacier in 2008, not the actual terminus altitude in a given year. For this reason, standard deviations for years prior to 2008 (ice-free in the 2008 DEM) are much greater than for years where ice is present in the DEM. This reflects the fact that glacier ice has lower surface roughness than its surroundings.

We then leverage the terminus surface altitudes to identify the point along the elevation band flowline corresponding to the terminus position at each year. Specifically, we query the elevation band flowline such that the glacier surface elevation in 2008 is within ± 2 m of the calculated terminus elevation of a given year since 2008. This small range accounts for the fact that not every possible elevation is included as an index in the elevation band model. The results show close agreement between models and observations by 2020 (Fig. 3), though performance quality differs from year to year. In particular, it appears that our models exaggerate frontal retreat rates by both 2018 and 2019 before slowing to match observations by 2020.

We hypothesize that the overestimation of retreat at certain years relates to the calving model's first order dependence on lake depth (see methods, main text), which we believe to be over-estimated in our model (see discussion, main text). This source of uncertainty may also lead to an overestimation of frontal retreat rates during certain years. It is notable that 2019-2020 marks the

transition to shallower modeled lake depths, accounting for the deceleration of retreat. We elaborate on this model limitation in our discussion section.

Finally, although a calving rate parameter calibration could theoretically force each of our six models to match observed frontal retreat perfectly, doing so would eliminate the insight we gain through our calving experiments (as per RC2 comment on Line 412 and throughout). Rather, deploying a constant calving parameterization across models ensures that model performance is not the result of calibration and does indeed reflect the impact of the onset of frontal ablation. This is addressed more specifically in the line edits below.

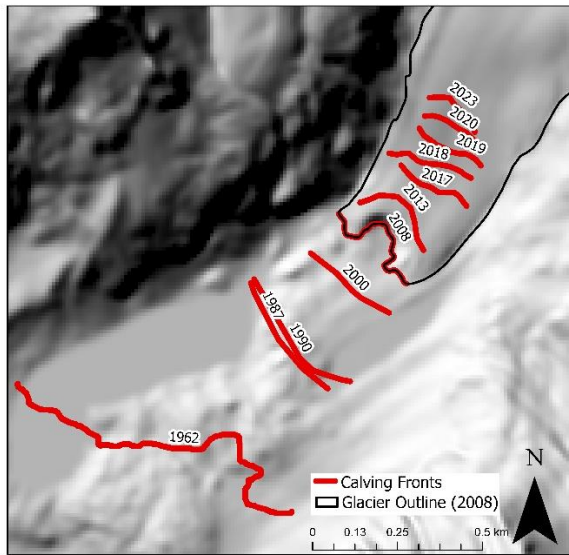


Fig. 1: The mapped glacier terminus positions overlaid on the 2008 DEM.

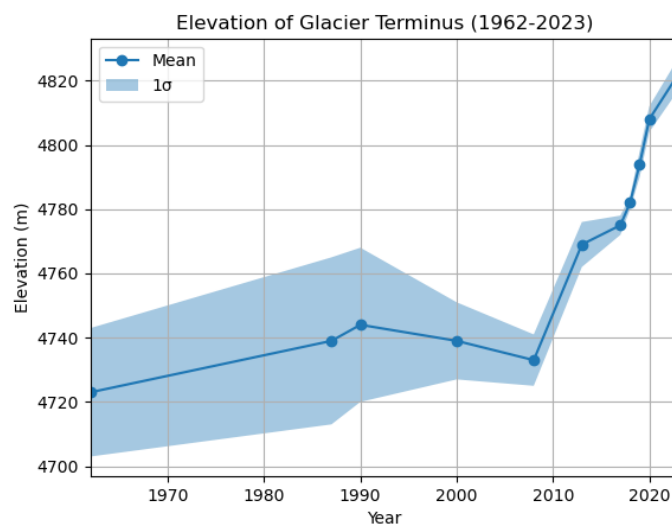


Fig 2: The glacier terminus elevations for each observation year derived from the map in Fig. 1 by averaging the DEM elevations where they overlap mapped glacier termini.

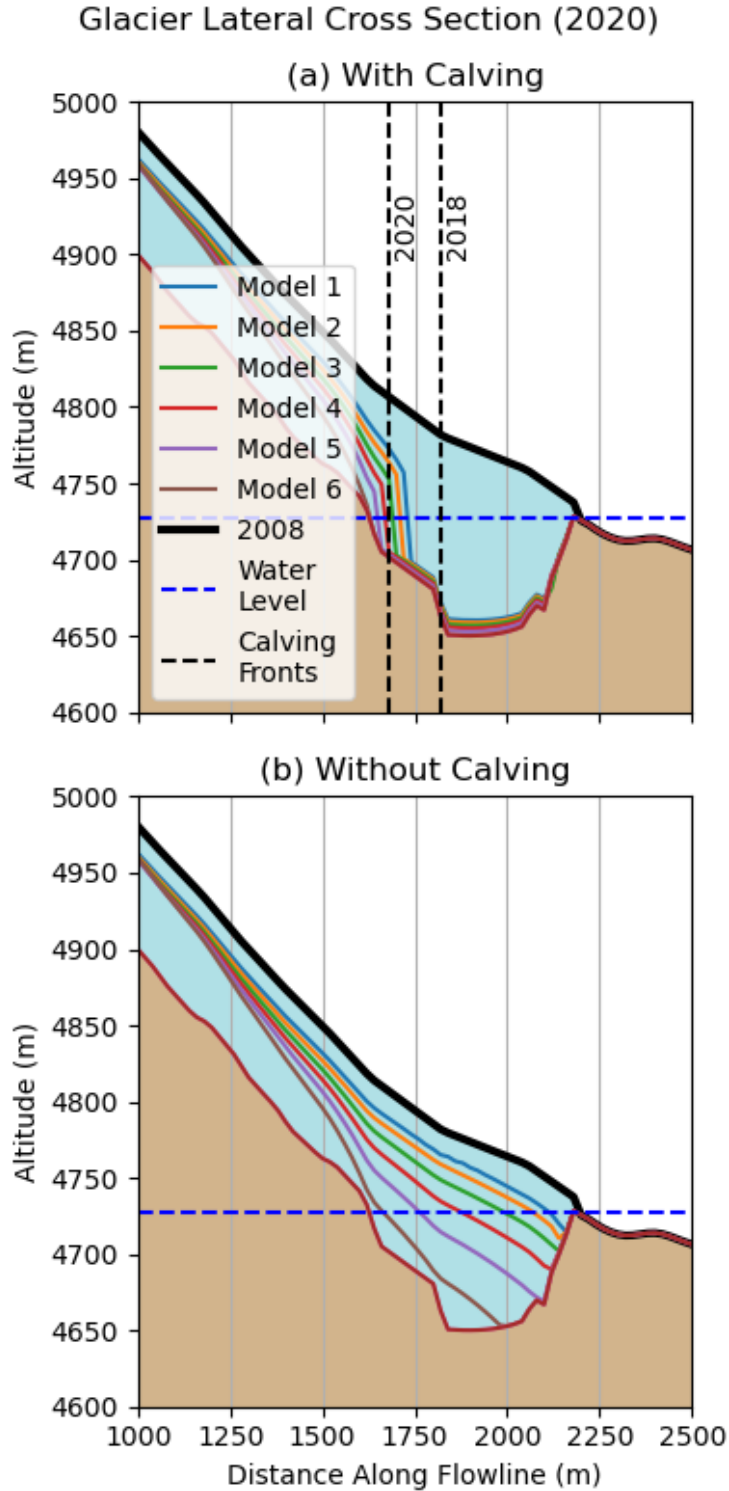


Fig 3: An update of Fig. 6 (main text) including 1) different symbology for observed (2008) vs. modeled glacier surfaces, 2) vertical lines representing mapped glacier terminus positions from 2018 and 2020, corresponding with the ice fronts mapped in Fig. 1 (main text).

c) Geodetic Mass Balance Observations

We have compared our ensemble mean specific mass balance (SMB) model outputs for each of the 42 epochs provided in the Hugonnet et al. (2021) study. We find that averaged across the years of a given epoch, our ensemble consistently overestimates mass loss as compared to the Hugonnet et al. study. In other words, our model results suggest that Queshque Glacier is retreating faster than the best estimate from the global geodetic mass balance study.

Nevertheless, our results are within the uncertainty bounds provided by Hugonnet et al. during all but 3 of the 42 epochs. All three of these epochs consider change as of 2020 including 2000-2020, 2015-2020, and 2016-2020. This suggests that a systematic bias exists either in the 2020 data used in taking the geodetic mass balance or in our simulation occurring around that time. Other epochs up to 10 years in duration (the second longest duration after the single 20-year measurement) show agreement between Hugonnet et al.'s and our own data. In summary, this comprehensive comparison indicates general agreement between the two datasets, bolstering confidence in our mass balance simulation.

d) In-Situ Ablation Measurements

We use ablation stake data measured by the National Water Authority of Peru between the years 2015 through 2019 to further evaluate our mass balance models and to compare the relative performance of individual parameterizations. This dataset was not available during our original analysis. The data comprise individual ablation stake measurements spanning about 4700-5150 m in altitude that have been converted to water equivalence. Some measurements report altitudes occurring below the glacier terminus elevation in 2008. While lower altitudes in the stake data may be in part linked to glacier thinning, altitudes below our calving water level of 4727 m cannot be explained in this way. It appears, rather, that some level of inaccuracy or negative bias exists in the altitude data. We therefore apply a uniform bias correction of 26 m across all elevations reported in the stake data such that the lowest stake measurement reaches 4727 m. We recognize that this correction is a source of considerable uncertainty, however, we determined it to be necessary since we lack additional GPS metadata. Due to inconsistencies in the duration of the stake measurements, ablation measurements were converted to m w.e. d^{-1} then multiplied by 365 days to arrive at standard units of m w.e. a^{-1} . Having made these corrections and standardizations, we can then compare the observed ablation to our modeled mass balance profiles.

Two analyses of the ablation measurements inform our evaluation. First, we consider the magnitude of ablation in the lower altitudes (defined as 4800 m and below) of the ablation zone as a constraint on the realistic melt rates near the glacier terminus. We then consider the observed ablation gradient in comparison to our models.

i) Magnitude of Ablation

Observed annual melt rates below 4800 m range from about -11.4 to $-3.5 \text{ m w.e. a}^{-1}$, averaging at $-7.5 \text{ m w.e. a}^{-1}$. Melt rates are greatest during the El Niño year of 2016 which is consistent with our simulation of overall specific mass balance. The range described above provides a limit on the magnitude of ablation we should expect to produce near the glacier terminus in our models. Average ablation rates at the lowest altitudes (4727-4800 m) during the years 2015 through 2019 are

presented in Table 2 below. We find that models 2, 3, 4, and 5 fall within the bounds of observations, with model 4 producing ablation rates closest to the observed mean. These results are generally consistent with our velocity validation, in which models 2-4 outperform the others and model 3 performs best (Table 1). We note that this is unsurprising, as the magnitude of ablation is related to overall ice flux, which controls surface velocity.

Model Number	Degree-Day Factor (mm w.e. $d^{-1} \text{ }^{\circ}C^{-1}$)	Mean Terminus Ablation (m w.e. a^{-1})
1	1.74	-3.1
2	2.58	-4.2
3	3.87	-5.5
4	5.92	-7.3
5	9.31	-9.7
6	15.31	-13.0

Table 2: Mean annual ablation produced by each model at 4727-4800 m during the years 2015-2019 which coincide with the timespan of in-situ mass balance measurements. The degree-day factors are included for reference.

ii) Ablation Gradient

While our models can reproduce the observed ablation rate at the glacier terminus, we find that their ability to reproduce the observed ablation gradient is limited. Fitting a linear trend to all negative stake observations, we calculate that on average ablation decreases (becomes less negative) by 8.5 mm w.e. d^{-1} for every 100 m in elevation. By dividing this value by the lapse rate of $-0.65^{\circ}C$ per 100 m, we arrive at an estimated temperature sensitivity (positive degree-day factor) of approximately 13 mm w.e. $d^{-1} \text{ }^{\circ}C^{-1}$. Based on the maximum lapse rate seasonality identified by Hellström et al. (2017), we note that this value could in reality range between 9.3 and 14.6. However, our model mass balance calibration was performed under the assumption of seasonally consistent lapse rate and we therefore adopted the conventional value.

In order to ensure that our model matches the observed ablation gradient, we recalibrated the model by adjusting the temperature bias to fit the observed geodetic mass balance using a fixed temperature sensitivity parameter (DDF) of 13 mm w.e. $d^{-1} \text{ }^{\circ}C^{-1}$. This DDF falls between models 5 and 6 of the submitted manuscript (Table 2). We find that this calibration overestimates the ablation rate at the glacier terminus. We next introduce an additional temperature bias, cooling the model until it approximates the observed average mass balance profile in both magnitude of accumulation/ablation and gradient. We find, however, that this model vastly overestimates the specific mass balance and would indeed induce glacier growth since 1962. To further investigate the threshold between glacier growth and retreat, we conduct a sensitivity experiment wherein the fixed-gradient model is cooled until balanced conditions are achieved between 1962 and 2008. The results indicate that all stake observations except from the extreme El Niño year of 2016 show a more positive mass balance than would be required for balanced conditions (Fig. 4).

This experiment highlights a fundamental limitation of our model, which is that we cannot simultaneously fit the observed magnitude of ablation and the total observed mass change across the glacier. However, various model assumptions may be able to explain this discrepancy. For example, the assumption of perfect continuity (that all mass in the accumulation zone contributes

to the ablation zone) which is inherent in OGGM may result in an overestimation of the true contributing accumulation area. This would in turn require more intensive ablation to compensate for the inflated accumulation, resulting in a model that reproduces the gradient but not magnitude of ablation (i.e., blue curve in Fig. 4a).

Alternatively, one could correct for overestimated contributing area by raising the equilibrium line altitude thereby reducing the accumulation area. This forces lower ablation gradients, as relatively high ablation rates persist at higher altitudes. This strategy is represented in our models that fit the magnitude of observed ablation without matching the observed ablation gradient (i.e., models 2-5, Table 2).

In summary, we have compared our mass balance models against total mass change from 1962-2008, the magnitude of ablation during 2015-2019, and the ablation gradient during the same period. The models used in the submitted manuscript fit the first two metrics while missing the third. We find that it is impossible within our model to fit both the first and third, and therefore conclude that we have chosen adequate validation metrics. This is further supported by our surface velocity and terminus position validations. Both the in-situ ablation validation and the surface velocity validation suggest that our model numbers 2-4 produce the most reliable output. The frontal position mapping shows closest agreement with models 3 and 4.

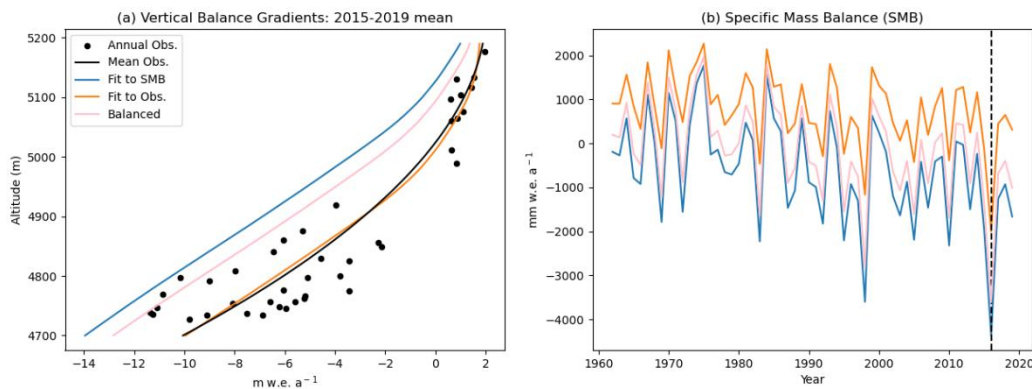


Fig. 4: (a) The observed and modeled vertical balance gradients (mass balance profiles) fit to the observed SMB (blue), the observed mass balance profile (orange), and to a glacier in long-term (1962-2008) climatic equilibrium (pink). (b) Annual SMB from the same same models, 1962-2020. The dashed line represents the strong El-Niño year of 2016, which is the only year that observed ablation rates (points left of pink curve in Fig. 4a) would produce negative mass balance across the glacier.

2. Calibrated parameters and their effect

The calibration process is quite clearly described, but it is a little odd that the temperature bias range then effects the DDF and potentially also the creep parameter (this second point should be made clearer in the paper). This effect makes sense from a calibration perspective, since the other parameters are compensating for the temperature bias to ensure the modelled mass loss matches that observed, but it has quite strong impacts on the other parameters. Note also that the precipitation lapse rate is the same in all models – if it was allowed to vary it might influence the

change in the DDF, since the accumulation could change as well as ablation. There might also be impacts on the modelled frontal ablation rates, but I am not completely sure. I think the section on mass balance calibration should make clear the resulting influence of the temperature bias on all the parameters.

We agree that we should make the connections between parameters more clear in the text and that impacts of the temperature bias on other parameters are the result of calibration. In the case of flow model parameters, we calibrate against radar (thickness) data. The A parameter therefore must be adjusted to ensure that thickness matches the observations despite different ablation rates across models. The ablation rates are themselves a product of the DDF, which is strongly influenced by the temperature bias.

The precipitation factor is indeed a free parameter in our mass balance model. However, it is calibrated subsequently to the DDF only if the model fails to converge in the previous step. This was never the case, so our default value of 2.5 was used across all models. We accept that this parameter is highly uncertain as very few reliable records of accumulation exist from the tropical Andes. Our direct, though limited in duration, mass balance measurements discussed above record accumulation as high as 1.95 m w.e. a^{-1} at 5150 m on Queshque Glacier. Using the precipitation factor of 2.5, our mass balance models produce average annual accumulation rates of 1.9 ± 0.4 m w.e. a^{-1} . This value is consistent with the limited direct accumulation measurements we have available. In contrast, the nearby Huascarán Col experiences lower annual accumulation of about 1.4 m w.e. a^{-1} (Thompson et al., 1995; Weber et al., 2023). Although the default precipitation factor of 2.5 does seem to produce realistic accumulation values for Queshque, we note that changes to this parameter do not introduce bias into the mass balance model so long as the DDF is recalibrated, as we have done in our study (Maussion et al., 2019).

Given this, it would be useful also if the authors could compare the calibrated DDF and creep parameters with those from the literature (ideally from this region, or similar glaciers). This would help to understand to what extent the calibrated parameters are reasonable. If DDFs themselves are not available then comparison with modelled melt rates would also be useful (which are available, as ablation is shown in Figure S5).

We will include a section elaborating on our DDFs as compared to previously published literature.

At Zongo glacier (16° S), Fuchs et al. (2013) use DDF values of 6.5 and 30 mm w.e. $d^{-1} \text{ } ^\circ\text{C}^{-1}$ to model observed glacial discharge for the dry and wet seasons, respectively. Averaged over an entire hydrological year, these values are somewhat consistent with the DDF calculated above using stake observations and support the higher sensitivity models used in our study.

Fyffe et al. (2021) find that on glaciers in the Cordillera Blanca, 5 °C warming induces a melt increases from 0.75-1.25 mm w.e. h^{-1} (estimated from Fyffe et al. 2021 Fig. 8). This translates to 3.6-6.0 mm w.e. $d^{-1} \text{ } ^\circ\text{C}^{-1}$, matching the sensitivity of our models quite well.

Relating to this the authors should be clear to not mistake the effects of calibration (so the differences between the models with different temperature biases) for processes related to more or less temperature sensitive glaciers (e.g. in the discussion section 5.4 and conclusion).

We understand the comment about conflation of effects of calibration for processes related to temperature sensitivity to be in reference to lines 408-414. In this paragraph, our aim is to highlight the effect of calving on variability across models. We observe that when the same calving parametrization is included across models, variability in mass loss is reduced. This is not the impact of calibration and we feel it is therefore valid to conclude that the lake calving process does indeed reduce the effects that differing temperature sensitivities have on relative mass loss rates.

We will alter the wording in the conclusion (line 444) from, "...with otherwise different climatic sensitivities," to, "...with otherwise different responses to climate warming."

One idea would be to add text in the discussion about the calibration processes, including how reasonable the calibrated parameters are, as well as how the use of these parameters (including the constant ones) may influence the results (for instance the temperature versus precipitation sensitivity of the mass balances and the influence of ENSO).

We agree that the discussion would benefit from a dedicated section comparing the calibrated parameters to data and previous literature. It would also be beneficial to discuss model sensitivities to parameterization decisions, including the constant parameters. In addition to what was elaborated upon above, we will note that although the precipitation factor does not introduce model bias, it may alter model sensitivity to temperature versus precipitation changes, with higher precipitation factors leading to more precipitation sensitive glaciers (Maussion et al., 2019).

I also include my minor comments below, which I hope are useful.

Yes, thank you; we appreciate the helpful edits.

With kind regards,

Catriona Fyfe

Minor comments

L21: 'precipitation amounts'

Accepted

L40 'other challenges'

Accepted

L64-73 – consider including the ETI methods you mention later also in this paragraph

Accepted

Figure 1a. Is the 2008 outline only for the Queshque Glacier, and not the other glaciers that were previously attached to it? If so maybe make this clearer in the caption.

Yes. We do so to indicate the extent of our model domain. We will clarify this in the figure caption.

L126 It might be nice to show the GPR bed topography in a little more detail. It is shown in Figure 1 but it's not so easy to see, and perhaps a specific figure in the SI would be useful.

We will include the following figure (Fig. 5) in the supplement.

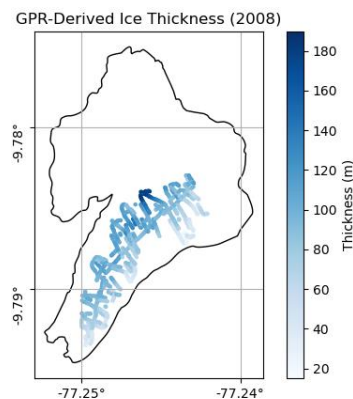


Fig. 5: Map of the GPR data used to calibrate the OGGM flow model. Though the GPR was taken in 2014, ice thicknesses in 2008 were derived by subtracting the GPR bed topography from the 2008 lidar DEM (see methods).

L131 It would be useful to add here the number of point observations you used for validation.

Only 17 georeferenced point measurements were provided by the 2009 GPR survey report. As elaborated upon in the main text, the points show general consistency with the subsequent 2014 survey (Fig. 6a). The 2009 GPR points span from the bottom of the glacier in the southwest towards the center of the glacier in the northeast of Fig. 6b. The points are located approximately along the centerline of the glacier and are each in proximity to multiple measurements from the subsequent GPR survey. There is no apparent relationship between the mean distance from the 2009 data points to their respective nearest neighbors and the resulting difference between measured and derived thickness. There does, however, appear to be a slight spatial bias, with derived thicknesses being more likely to underestimate the 2009 measurements at lower elevations. The significant outlier where the derived thickness is approximately 28 m thinner than observed occurs at a discontinuity in the 2009 survey, suggesting that the technician may have needed to navigate an obstacle which may have produced abnormalities in the ice thickness profile or potentially caused an error in measurement or radargram interpretation.

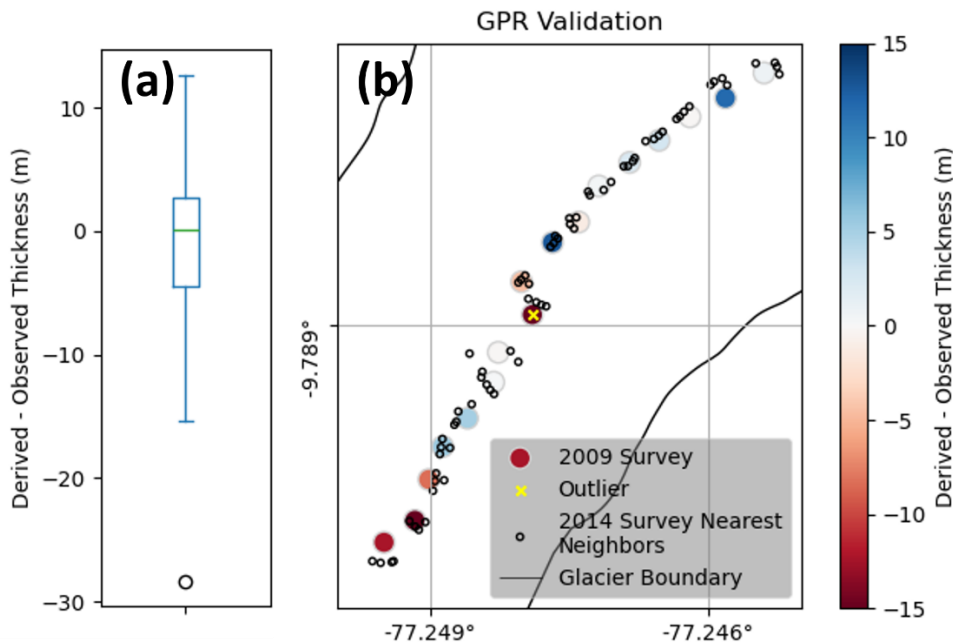


Fig. 6: To be included as an additional supplementary figure. (a) Box and whiskers plot showing the distribution of error between the derived and observed thickness values, including the position of the single negative outlier. (b) A map of the 2009 GPR survey showing the error calculated at each point.

L168 Where did you get this range of temperature biases from? It seemed that the cold bias was identified from the comparison of CRU and PISCO – was this the source?

The temperature bias range was determined by the simple requirement that temperatures needed to fall at or below freezing for a significant duration within the altitude range of the glacier. We will clarify this point in the text.

Section 3.4.2 It might be a good idea to have a sentence at the beginning of this section which summarises the approach here. This is just because the need to calculate initial ice thicknesses is unclear at the beginning since you have the measured GPR data, but then you mention later (line 190) that the GPR data is used for calibration.

We will add the following sentence at the beginning of section 3.4.2: “The GPR dataset is leveraged to calibrate the ice flow model by minimizing error between modeled and observed ice thickness.”

L185 Consider adding a reference for this statement

Here we follow Pelto et al. (2020).

L189 Why did you choose a minimum slope parameter of 7.5 (and what is the unit of this parameter)? Do you have a reference for this? The sentence above about exaggerated overdeepenings would also benefit from a reference.

The correct unit for the minimum slope parameter is degrees. The text should therefore read “7.5°”. This parameter clips the glacier surface slope to $\geq 7.5^\circ$ during the ice thickness inversion process, resulting in the flattest sections of ice retaining higher flow rates. This is necessary because otherwise the model will overestimate ice thickness to satisfy ice flux continuity with the steeper slopes above. The value of 7.5° was accepted after a series of sensitivity experiments examining the impact of this parameter on terminal ice thickness (eventual lake depth). However, as elaborated upon in the discussion section of the main text, we still suspect that terminus ice thickness (lake depth) is over estimated. This parameter is only discussed briefly in the OGGM documentation (<https://github.com/OGGM/oggm/blob/master/oggm/core/inversion.py>).

Table 1. The melt threshold of -1°C may be a little cold for tropical glaciers, as there tends to be strong diurnal temperature gradient so it takes some energy in the morning to warm the ice/snow surface to melting. Pellicciotti et al. (2008) mentions this general idea, although they used a 1°C temperature threshold and only calibrated the TF and SRF factors (but TF became negative, they say due to this effect, see Figure 12). My general point is that ‘Alpine’ ETI/TI parameters may not work as well in the Andes. I imagine in reality the calibrated TI factor will compensate for this, but it might be worth mentioning somewhere in the discussion.

This is an excellent point that we have elaborated upon as a limitation in the discussion section of the manuscript. We agree, ultimately, that the DDF calibration should compensate for other parameter uncertainties.

Section 3.5 Just a suggestion – but it might be useful to name each of your experiments so you can easily refer back to them later.

We appreciate this suggestion but feel it is easier for the reader to describe the experiments (e.g., “constant climatological mean temperature”) rather than returning to an index with experiment names.

Lines 218-224 Although it is fairly well explained, I think you could be even more explicit in explaining how the changes in the temperature bias go through to influence the model sensitivity. For instance instead of saying ‘low magnitude ϵT ’ you could write ‘less negative’ (and vice versa), and also say that under these conditions the calibrated DDF is lower, meaning melt is less sensitive to air temperature. It might have also helped to directly compare the PISCO data with local weather station information and use this for a bias correction. Was there a reason why this wasn’t considered? You should also mention that the creep parameters are also influenced by the temperature biases – so that you have higher creep values under the most negative temperature biases.

We accept these recommendations regarding being more explicit about relations between calibrated parameters in this paragraph. We will also add discussion here regarding the precipitation factor and creep parameter, both of which are elaborated upon above in this author response.

We attempted to locate meteorological station data from the Queshque valley, but were unsuccessful, as mentioned in lines 224-225.

Table 2: You have quite a large range of calibrated DDFs due to the range in the temperature biases. Which of these is more reasonable compared to calibrated degree day models from the literature? This might help you to work out which of the models is more sensible and would give a melt response to temperature that is reasonable for the region. This might be something worth adding to the discussion.

See response to major comments.

Figure 2 I am not convinced that showing these correlations as bar graphs is necessarily the best approach. Can you make clear in the caption which model was used for the correlations? I presume the ensemble mean?

Yes, these correlations refer to the ensemble mean values. We will clarify this point in the caption. We opted for the bar graphs over other visualization methods such as correlation heatmaps because we felt this approach was the most intuitive.

Lines 257-259 The percentages here for the wet and dry season accumulation and ablation don't add up to 100%, are you in reality speaking about DJF and JJA only?

This is exactly the case. We have excluded the shoulder seasons to highlight the largest seasonal difference. We have clarified this in the text.

Figure 3 (and related Figure S5) Why not show the ensemble mean as well? Especially as it is likely more reasonable, given that Figure 5 suggests that model 3 (in the middle) is likely the most reasonable.

We feel that Figures 3 and S5 (main text & supplement) illustrate differences between models in a way that is informative to the reader. In particular, they demonstrate that all model showcase the same seasonality albeit with different magnitudes of variability. We feel that presenting mean values instead would showcase the seasonality without illustrating the later point about our model ensemble.

L277 It would be useful to show a comparison of the ice thicknesses of the different models, especially since there might be quite some differences in the accumulation zone which are not shown anywhere.

As shown in Table 3 (main text), initial thicknesses across the ablation zone where GPR data are available are stable across models. The same table shows, however, that initial ice volumes range from $7.74 \times 10^7 \text{ m}^3$ to $8.10 \times 10^7 \text{ m}^3$. This is primarily due to differences in the accumulation zone ice thickness.

L282 Are the real-time monthly conditions from PISCO? If so please add this for clarity.

Yes. We will rephrase to clarify that we mean the contemporaneous PISCO climate conditions as opposed to the 1962-1992 climatological means (derived from CRU and PISCO).

L285 Please give the steady state ensemble mean volume

This refers to the ensemble mean initial (2008) volume of $7.9 \pm 0.1 \times 10^7 \text{ m}^3$ (1σ uncertainty) of ice (from Table 3, main text).

L287 Give here exactly the increase in the mass loss caused by implementing the frontal ablation scheme, especially since it is a key result of the paper.

By 2020, the ensemble mean of models with calving shows additional mass loss of approximately $0.4 \times 10^7 \text{ m}^3$ as compared to models excluding the calving process.

L292-293 Can you quantify the comparison of the models with the glacier velocity measurements, even only upstream of the modelled calving front? It also looks from Figure 5 like model 3 has the best correspondence with the observations, but ideally you should quantify this and say it exactly.

See response to major comments.

Figure 4 Explain in the caption what the solid lines represent, I imagine the ensemble mean?

Yes, the solid lines are the ensemble mean values. We will clarify this in the figure caption.

Figure 6 perhaps use a different line type to differentiate the observations from the model results. Also, are there any data for 2020 which help you to determine which model is most correct? I see in Figure 1 that there are calving front elevations determined in 2018 and 2020, it would make sense to add these here to understand if any of the models are able to replicate them.

Accepted. See response to major comments.

L321 I am not sure I would use the word ‘peaks’ here to describe the higher wet rather than dry season, as I think you are talking about the wet season in general. The highest melt rates (for Peru at least) tend to be at the end of the dry season/beginning of the wet season anyway.

Accepted. We will change the wording from “peaks” to “increases.”

L322 and L338 The exception for Shallap is likely only the case in the modelled time period in Fyffe et al. (2021) – potentially due to strong La Nina conditions. It might not generally be the case, and I would not in general say Shallap has an atypical ablation seasonality. For this reason please adjust the two sentences related to this. Gurgiser et al. (2013) found mass balances to vary between years precisely because of differences in the snow cover over the ablation zone.

Thank you for this comment. We will clarify that multiple studies of Shallap glacier have identified exceptional ablation seasonality during various years, which has been linked to the timing and phase of precipitation.

L329 ‘excluded by the nature’

Accepted.

Section 5.3 Just a suggestion but you could mention Lamantia et al. (2024)
<https://doi.org/10.5194/tc-18-4633-2024>

Agreed. We will add discussion of Lamantia et al. (2024) in this section, as this paper also supports our observation of high correlation between wet season ENSO indices (ONI, SOI, and Niño-3.4) and annual specific mass balance.

L382 ‘the El Niño/warm’

Accepted.

L386 Be specific about which ENSO indices correlate with the SMB anomaly

Here we refer to annual or dry-season indices including ONI, Niño-3.4, and SOI. These are then contrasted with wet-season specific metrics.

L402 ‘2008 through to 2020’ (although maybe it’s ok in US English)

We believe this is standard US English but will defer to the editor’s preferences.

L402 The terminology ‘least sensitive calving glaciers’ is a bit odd as you are really speaking about specific model runs. Consider something like ‘the mass loss from the most temperature sensitive model including calving’ and include the model number so it is clear which one you mean.

This statement refers to model number 1 (run with calving) surpassing the ensemble mean mass loss for all models that excluded calving. We will clarify this in the text.

L412-412 It is an odd way to talk about the model results, as it is phrased as if the glaciers themselves are accelerating frontal ablation to counteract their lower sensitivity. Instead, the calibration process is resulting in higher frontal ablation rates in order to compensate for low TF parameter values (or at least I think this is what is happening). It is very important here that you are clear what is an effect of calibration compared to what is a real glacier response.

We appreciate the reviewer bringing up this point of confusion regarding our calibration process. As frontal ablation was not a significant process throughout most of our calibration period (1962-2008), we initialize the model glacier as a land-terminating type, which is unaffected by calving. As described in section 2 (main text), the onset of frontal ablation occurred only after 2008 due to complex processes concerning the surrounding hydrology and sub-glacial topography.

Models 1-6 are therefore calibrated without respect to calving. The subsequent experiments comparing calving and non-calving glaciers (section 3.5, main text) therefore begin from model glaciers with differing temperature sensitivity parameters (DDFs). We show that in non-calving scenarios, these glaciers evolve quite distinctly, and with greater variance, than in calving scenarios. In other words, we have isolated the impact of calving on modeled glacier evolution. This is not the result of any sort of calibration procedure, as all models use the same calibration parameterization.

We will be sure to clarify this point in the text and to explicitly emphasize that the calving process is not a factor during the mass balance calibration procedure.

L420-429 – you already have a few other sources of data, but just to make you aware of the HydroLake database, I am not sure if it would be helpful or not (the depths are also estimated), but you could look at the specific depth to area ratio for lakes only in this region. As I mentioned above it might also be useful to compare with observations of the calving front, since you mention observations from 2018 and 2020, and I imagine you could find more from satellite information.
<https://www.hydrosheds.org/products/hydrolakes>

We appreciate learning of this additional dataset, but feel that the two datasets already discussed are sufficient for this section. We will consider the HydroLake database in future research.

See response to major comments regarding calving front positions.

L438 The phrase ‘out performs theoretical expectations’ is a bit odd and not precise. Perhaps add here specific comparison against data (e.g. the best models ability to replicate the glacier velocity or calving front location).

We will change this phrase to refer to the specific validation metrics discussed in the main text and above in our response to major comments concerning mass balance validation.

L444 The phrase ‘reduces the variability between glaciers with otherwise different climatic sensitivities’ is perhaps misleading, since this variability of the calving front rates related to the models with varying TF values could be due to the calibration process rather than a ‘real’ glacier response.

See comments concerning this point above.

References

- Hugonnet, R., McNabb, R., Berthier, E. et al. (2021) Accelerated global glacier mass loss in the early twenty-first century. *Nature* 592, 726–731, <https://doi.org/10.1038/s41586-021-03436-z>
- Lamantia, K. A., Larocca, L. J., Thompson, L. G., and Mark, B. G. (2024) El Niño enhances snow-line rise and ice loss on the Quelccaya Ice Cap, Peru, *The Cryosphere*, 18, 4633–4644, <https://doi.org/10.5194/tc-18-4633-2024>
- Pellicciotti, F., Helbing, J., Rivera, A., Favier, V., Corripio, J., Araos, J., Sicart, J-E., Carenzo, M. (2008) A study of the energy balance and melt regime on Juncal Norte Glacier, semi-arid Andes of central Chile, using melt models of different complexity, *Hydrological Processes*, 22 (19) 3980-3997

Additional References Cited in Author Response

- Fuchs, P., Asaoka, Y., & Kazama, S. (2013). Estimation of Glacier Melt in the Tropical Zongo with an Enhanced Temperature-Index Model. *土木学会論文集b1 (水工学)*, 69(4), I_187-I_192. https://doi.org/10.2208/jscejhe.69.I_187
- Fyffe, C. L., Potter, E., Fugger, S., Orr, A., Fatichi, S., Loarte, E., Medina, K., Hellström, R. Å., Bernat, M., Aubry-Wake, C., Gurgiser, W., Perry, L. B., Suarez, W., Quincey, D. J., & Pellicciotti, F. (2021). The Energy and Mass Balance of Peruvian Glaciers. *Journal of Geophysical Research: Atmospheres*, 126(23), e2021JD034911. <https://doi.org/10.1029/2021JD034911>
- Hellström, R. Å., Fernández, A., Mark, B. G., Michael Covert, J., Rapre, A. C., & Gomez, R. J. (2017). Incorporating Autonomous Sensors and Climate Modeling to Gain Insight into Seasonal Hydrometeorological Processes within a Tropical Glacierized Valley. *Annals of the American Association of Geographers*, 107(2), 260–273. <https://doi.org/10.1080/24694452.2016.1232615>
- Maussion, F., Butenko, A., Champollion, N., Dusch, M., Eis, J., Fourteau, K., Gregor, P., Jarosch, A. H., Landmann, J., Oesterle, F., Recinos, B., Rothenpieler, T., Vlug, A., Wild, C. T., & Marzeion, B. (2019). The Open Global Glacier Model (OGGM) v1.1. *Geoscientific Model Development*, 12(3), 909–931. <https://doi.org/10.5194/gmd-12-909-2019>
- Millan, R., Mouginot, J., Rabatel, A., & Morlighem, M. (2022). Ice velocity and thickness of the world’s glaciers. *Nature Geoscience*, 15(2), Article 2. <https://doi.org/10.1038/s41561-021-00885-z>

- Pelto, B. M., Maussion, F., Menounos, B., Radić, V., & Zeuner, M. (2020). Bias-corrected estimates of glacier thickness in the Columbia River Basin, Canada. *Journal of Glaciology*, 1–13. <https://doi.org/10.1017/jog.2020.75>
- Thompson, L. G., Mosleythompson, E., Davis, M. E., Lin, P. N., Henderson, K. A., Coledai, J., Bolzan, J. F., & Liu, K. B. (1995). Late-Glacial Stage and Holocene Tropical Ice Core Records From Huascaran, Peru. *Science*, 269(5220), 46–50.
- Weber, A. M., Thompson, L. G., Davis, M., Mosley-Thompson, E., Beaudon, E., Kenny, D., Lin, P.-N., & Sierra-Hernández, R. (2023). Drivers of $\delta^{18}\text{O}$ Variability Preserved in Ice Cores From Earth's Highest Tropical Mountain. *Journal of Geophysical Research: Atmospheres*, 128(19), e2023JD039006. <https://doi.org/10.1029/2023JD039006>