Answer to Reviewer 1

In the following, we provide a point-by-point Author Response (AR) to any of the Reviewer Comments (RC) obtained for the manuscript that was under discussion. When presenting suggestions for how the manuscript text could be revised, the text is presented in *italics*.

RC1: The uncertainties are generally poorly treated. A sub-section of the method explaining what are the sources of uncertainties is missing. Uncertainties and errorbars are sometimes shown or quoted in the text (e.g. fig 6, L7 and 391, L385), but it is never clearly stated what they refer to (one or two sigma? what is actually included in the uncertainties?). Even more worrying is a very strange statement in the discussion on L374-375 that states that the uncertainties originate solely "only account for differences in the geodetic ice volume changes used during calibration (see Sect. 3.2.3), as this is identified as the main impacting factor (Cremona et al., 2025)". First of all, there is no mention of uncertainties from geodetic volume estimates in Cremona et al. (2025), so the reference is wrongly cited. Second, I would be very surprised that the volume change estimates from very high resolution photogrammetric DEMs are the main source of uncertainties compared with the rest of the modelling chain, including the density conversion assumption, calibration and extrapolation of the melt and precipitation factors.

AR1: We thank the Reviewer for reviewing our manuscript and for the time invested in checking our submission. For what the specific comment is concerned, we acknowledge the limitation of our uncertainty estimation, and we consequently performed a new uncertainty assessment considering different sources as suggested. We thus suggest adding a new section and presenting it as follows:

Uncertainty assessment

To estimate the uncertainty in the modelled mass balance, we account for uncertainties stemming from (a) the geodetic volume change used during calibration, (b) the modelled evolution of the glacier area, (c) the extrapolation of the precipitation correction factor with machine learning, and (d) the presence and effect of supraglacial debris. Uncertainties introduced by different geodetic volume change estimates during calibration are derived by changing the input geodetic volume change used across the available options. Uncertainties arising from the modelled evolution of the glacier area are derived by conservatively varying modelled area change rates by ±50%. This value covers a broad range of glacier area responses and, hence, considers that our parametrization is rather simple. Uncertainties originating from the extrapolation of the precipitation correction factor with machine learning are derived by varying c_{prec} *by ±0.37, which corresponds to the RMSE found by cross-validation (see Sect. 3.2.4). Uncertainties associated with the presence of supraglacial debris cover are derived by running the model as if there were no debris. The deviation of the mass balance resulting from the different new model runs (a)-(d) from the reference run was statistically evaluated, providing the resulting uncertainties* $\sigma_{geodetic}$, σ_{area}, $\sigma_{extrapolation}$, *and* σ_{debris}. *These uncertainties are considered independent and are combined into an overall uncertainty in daily glacier-wide mass balance for every day as:*

$$\sigma_{\dot{b}} = \sqrt{\sigma^2_{geodetic} + \sigma^2_{area} + \sigma^2_{extrapolation} + \sigma^2_{debris}} \quad (1)$$

resulting in an average uncertainty in daily mass balance that ranges up to 0.037 m w.e. d^{-1} *with a mean of 0.0026 m w.e.* d^{-1}. *Uncertainties in inferred Swiss-wide winter and annual mass balance are 0.27 and 0.20 m w.e., respectively. The average uncertainty in the mass balance cumulated over the entire study period is 2.5 m w.e.*

RC2: The machine learning extrapolation based on XGBoost is difficult to grasp for non-specialist readers. It would be good to illustrate the model performance to predict annual mass balance expressed in m w.e., because the only evaluation metric shown in the text is the RMSE of c_{prec} which is not very intuitive. Afterwards melt parameters are calibrated a second time to match the geodetic estimates (L240-241). I do not fully understand why this correction is necessary and what is the actual impact of this two-step correction/calibration.

AR2: We apologize for the lack of clarity, but the model performance to predict seasonal and annual mass balance (in m w.e.) is presented in Section "3.3 Validation", i.e. with respective MAD and bias for the 13 glaciers without SCAF observations for which the precipitation correction factor is extrapolated. We suggest to rephrase as follows to promote better clarity:

To identify the best model, a randomised grid search with five-fold cross-validation over the hyperparameter space was performed. In each fold, approximately 70 glaciers are used for training and about 20 glaciers for testing. This results in a model that reproduces c_{prec} *with a mean cross-validation RMSE over the five folds of 0.37 (-) and a mean absolute percentage error (MAPE) of 13%. The model performance to predict seasonal and annual mass balance for glaciers relying on extrapolated* c_{prec} *is validated in Sect. 3.3.*

We apologize that we did not describe the essence of the procedure clearly enough. Melt parameters are not calibrated, since up to here they were not calibrated at all for these glaciers. In a first step, the precipitation correction factor is extrapolated with machine learning, and in the second step, melt parameters are calibrated to match geodetic ice volume changes. We refer to AR3 for the changes to "Sect. 3.2.3 Model calibration" in the revised version, which aims at a clearer description of the calibration procedure.

RC3: I do not fully understand the calibration procedure, which I find quite confusing. My understanding of Cremona et al. (2025) is that annual glacier-wide mass balances from GLAMOS are used to calibrate the multiple melt parameters. But in this manuscript, it is mentioned that the melt factors are calibrated on geodetic estimates only (L218-220), and I don't understand how a single observation is used to calibrate two or three parameters. Can you clarify this discrepancy? It is also important to highlight that your method does not rely on field measurements for calibration if this is actually true. I am also wondering if there isn't a kind of circular reasoning when comparing your annual mass balance estimates with GLAMOS because the later ones are also calibrated with the same DEMs and model used in this study.

AR3: Again, apologies for the confusion that is generated by our description of the calibration procedure. First of all, regarding Cremona et al. (2025), also there just one observation is used to calibrate melt parameters, which is the multi-year average mass balance that is indeed derived from GLAMOS observations. In this study, the multi-year average mass balance is derived from geodetic estimates.

The reviewer is correct that calibrating two model parameters with a single observation is challenging due to parameter equifinality. This means that multiple combinations of parameter values can produce the same result (depending on the initial condition of the optimization problem). Therefore, it is not possible to resolve the equifinality between melt parameters with our approach. However, we want to stress that the equifinality between the two parameters in the melt models (i.e. separating the effect of radiation and temperature) is less relevant compared to the equifinality that separates between accumulation and melt (that our approach actually solves) for providing sub-seasonal mass balance estimates.

And yes, the presented calibration does not rely on direct measurements.

We reworded Sect. 3.2.3 Model calibration as follows to avoid confusion by taking into account the above-mentioned comments:

The mass balance model is calibrated relying on the concept described in Cremona et al. (2025), i.e. by combining information from SCAF observations and geodetic ice volume changes with a three-step optimisation procedure (Fig. 3). In the first step, the parameters of the melt models (MF and $a_{\mathrm{snow/ice}}$

in Eq. 2, and TF and SRF in Eq. 3) are adjusted to reproduce the observed geodetic ice volume change. Here, a density of volume change of 900 kg m^{-3} *is assumed to convert mass change to ice volume change. This assumption is motivated by the almost complete disappearance of firn areas observed since the beginning of the 21st century in the Swiss Alps, resulting in a rapid decline in the importance of firn processes in the investigated region. In the second step, the accumulation parameter, i.e.* c_{prec} *in Eq. 1, is optimised in order to minimise the difference between modelled and observed SCAFs over the melt seasons 2015-2021, i.e. the value of* c_{prec} *that provides the lowest Root Mean Square Error (RMSE) between observed and modelled SCAFs is selected (for more details refer to Cremona et al. (2025)). In the third step, the temperature lapse rate is adjusted to further optimise the model's reproduction of the observed SCAF. This step is new with respect to the procedure presented in Cremona et al. (2025), and it addresses the significant impact of the temperature lapse rate on the altitudinal distribution of mass balance and, hence, on the modelled course of the SCAF during the melt season. The initial estimate of the temperature lapse rate is based on the gridded temperature data, deriving the gradient from the temperature and elevation of the cells next to the glacier. This provides insight into the lapse rate for a wider area, but neglects local patterns. We then refine this estimate by adjusting the temperature gradient in 5% increments, selecting the value that minimises the difference between modelled and observed SCAFs. This procedure is applied for every value of the ice volume change available from the swisstopo DEMs (GLAMOS, 2024) and the ones derived in this study from the SNFI DEMs, resulting in one parameter set for each value of a geodetic ice volume change inferred over various periods of at least five years between 2010 and 2024, whereby parameters remain constant over the study period. This calibration is applied to 87 glaciers with available SCAF observations, and across these 87 glaciers the number of parameter sets ranges from 1-20, with an average of 5 parameter sets per glacier. For each parameter set, the daily mass balance over 2010-2024 is calculated, i.e. in an ensemble approach, and the final estimate for the daily mass balance corresponds to the resulting mean of this ensemble. It is important to note that this calibration approach does not rely on direct field observations, which is a substantial advantage for regional-scale assessments.*

In our opinion, there's no circular reasoning when comparing with GLAMOS annual balances: The utilised DEMs do not enter GLAMOS glacier-wide mass balance evaluations. GLAMOS glacier-wide mass balance series are periodically validated with geodetic mass balances (see e.g. Huss et al., 2015) but this was not the case during the time period considered in the present study. Additionally, besides the annual balances, GLAMOS in-situ data are also used to assess the year-to-year and seasonal variability of mass balance, which is independent.

RC4: I was confused with the manuscript structure, especially the "study site and data" and "methods" section. The long description of the DEMs and their origin in section 2 is not well placed and should be in the same sub-section as current 3.1. In my opinion, the method section should start with the model description, then the calibration strategy and the geodetic data processing. This is more a personal opinion, but it might help differentiating this work from (Cremona et al., 2025).

AR4: We thank the reviewer for the comment, and we acknowledge that there are some methodological aspects in Sect. 2 "Study site and data". However, the scope of the study is focused on the evaluation of the model outputs, and not on the DEMs, which we use and treat here as data to calibrate the model. Thus, we wanted to highlight this difference and decided to put the short description for the DEM generation, which is necessary for context, in this section. As for the other proposed re-structuring, i.e. starting with the model description, then the calibration strategy and the geodetic data processing, we find it confusing to describe the calibration strategy before the data on which on which it relies.

RC5: Why not looking at summer mass balance? - ok not to do it, but it should be explicitly acknowledged

AR5: Actually, in Fig. 8 and 10 we conducted assessments including the summer balance, even if it is the anomaly of the summer balance and not the actual balance. In Fig. 7, we only presented winter and annual balances and not the summer balance because this is complementary information. By knowing the winter and annual balances, the summer balance does not add any new information as it is the difference between the two.

RC6: Many other "processes" could be explored with your simple mass balance model, in particular to explore further extremes as in Cremona et al. (2023). Can you quantify the effect of the lengthening of the melt season? The effect of early summer heatwaves? The intensification of melt? Your time series is rather short and might not allow to discuss these issues, but it would still be interesting to touch upon these questions.

AR6: We appreciate the reviewer's insightful suggestion. We addressed these points, however, we did not find a significant trend indicating an increase of the melt season length or an intensification of the melt. This is mainly because those factors are strongly influenced by the strong year-to-year fluctuations in meteorological conditions. Thus, the beginning and end of the melt season, as well as, for example, the average melt during the different summer months, vary strongly between years, and therefore prevented us from detecting any significant trend. Additionally, as the reviewer correctly states, the study period is rather short, which make such conclusions difficult to draw. We suggest adding the following to the revised manuscript:

Our daily mass balance time series provides the basis for further assessments related, for example, to changing characteristics of the melt seasons over the years. Here, we assess (i) changes in the length of the melt season, and (ii) changes in the melt rates during summer over the study period. We define the length of the melt season as the number of days between the date with the maximum cumulative mass balance and the date with the lowest cumulative mass balance (both expressed relative to 1 October). No significant trend indicating a lengthening of the melt season can be observed. We also compared the melt rates during the summer months (June, July, and August) to detect the potential intensification of the melt rates. However, also here no significant trend was found. Strong year-to-year fluctuations are present in the inferred variables. These variations override a potentially significant trend over the relatively short time period considered here.

Regarding the suggestion on the effect of summer heatwaves, we suggest rewording the manuscript as follows:

In contrast, the year 2022 shows the drastic effect of a year with extremely unfavorable conditions (Berthier et al., 2024; Menounos et al., 2025). This year was characterized by (i) among the lowest winter precipitation sums recorded over most of Switzerland during the study period, (ii) one of the warmest summers (Fig. 9) with the presence of several heatwaves (Cremona et al., 2023), and (iii) a very early start of the melt season (e.g. Voordendag et al., 2023). The combination of these factors led to a fast depletion of the snow cover, leaving large fractions of glacier ice exposed very early in the season, and led to unprecedented ice losses recorded in one year, i.e. with a mass balance of below $-2\,m\ w.e\,yr^{-1}$ *for all four basins (Fig. 11) and almost* $-2.5\,km^3$ *of ice wastage in the Swiss Alps (Fig. 6b).*

Specific Comments:

RC7: L3-4: "daily mass balance estimates" are not really presented in the manuscript text (higher validation resolution is seven days). Consider removing from the abstract.

AR7: Although the reviewer is right about the scale of the validation, i.e. results are indeed not validated at resolutions higher than seven days, daily mass balance estimates are presented at various instances throughout the paper, and the daily resolution of the results is a central aspect of the paper. The daily scale is not validated as no suitable large-scale data were available. The 7-day scale was simply chosen to make the validation meaningful because of uncertainties in daily observations. However, this does not mean that the method is unable to provide a daily resolution.

RC8: L32: "combine" is ambiguous. In practice geodetic estimates are used to calibrate a model.

AR8: We suggest rephrasing as follows:

A well-suited approach for bridging this temporal gap and, thus, deriving glacier mass balance over shorter time scales is to employ modelling approaches relying on geodetic mass balance estimates for calibration.

RC9: L121: filtering at one standard deviation is quite extreme as it removes more than thirty percent of observed data in the case of a Gaussian distribution. Classical hypsometric filtering are much less aggressive with three or five sigma thresholds (McNabb et al., 2019)

AR9: We acknowledge that filtering at one standard deviation might be more aggressive compared to other methods mentioned. Therefore, to test the choice of the filtering threshold and its effect on the results, we compared filtering at one standard deviation with filtering at 3- and 5-standard deviations. We did this comparison for 12 selected glaciers. The comparison indicates only a slight difference in the resulting geodetic ice volume change when filtering at 1- versus at 3- or 5-standard deviations. In fact, on average over the 12 selected glaciers, the difference in geodetic volume change to the reference case (filtering at 1 standard deviation) corresponds to 0.2% and 0.3% for the 3- and 5-standard deviation threshold, respectively. Thus, we expect the influence of the threshold on our regional results to be minimal.

RC10: L122: local or regional hypsometric interpolation? What happens to the elevation bands with no observation?

AR10: It is a local hypsometric interpolation, which is performed with the implementation of the xDEM package (xdem contributors, 2021). For elevation bands with no observations, the function interpolates based on the values of the remaining elevation bands (see. xDEM's documentation for more detail).

We suggest rephrasing as follows to account for this comment:

The, data gaps are then filled following a local hypsometric approach as implemented in the Python package xDEM (xdem contributors, 2021; Mannerfelt et al., 2022). This approach relies on the correlation between elevation and elevation change to capture the hypsometric signal of each glacier. Note that for elevation bands without any available data, an interpolation based on the elevation-dependent signal derived from the remaining elevation bands is performed (see xdem contributors (2021) for more details).

RC11: L130-140: a reference to Piermattei et al. (2024) should be added, as they face the same issue with the temporal stamping of SwissTopo DEMs.

AR11: Thanks for the insightful hint. We added the suggested reference.

RC12: L165: albedo should be unitless.

AR12: We apologize for the mistake and have corrected accordingly.

RC13: L166-167: what is the advantage of using two melt models and taking the average estimate?

AR13: It is more stable because it reduces the uncertainties that may come from the limitations of the individual melt models. We re-worded as follows:

We then combine the outputs of the two melt models by taking the average of the computed daily melt, promoting stability in the model outputs by reducing uncertainties that may derive from the limitations of the individual melt models.

RC14: L242: the “ensemble approach” is not well defined. What does it refer to? What is the ensemble size?

AR14: We apologize for the lack of clarity introduced by our wording. For every glacier, there are multiple geodetic ice volume change estimates available for calibration. For each individual one of these ice volume changes, a specific parameter set is calibrated. As a result, there are multiple parameter sets available for every glacier. Thus, the ensemble consists of the mass balance that is derived using each of these parameter sets. Thus, the ensemble size varies from glacier to glacier and corresponds to the number of geodetic ice volume changes available for calibration. The average ensemble size corresponds to 8 members (see Sect. 3.1). We suggest rephrasing as follows:

As before, because there are multiple ice volume changes available for every glacier (see Sect. 3.1), multiple parameter sets are available for every glacier. For each of these parameter sets, the daily mass balance is calculated, i.e. in an ensemble approach, whereby the size of the ensemble depends on the number of geodetic estimates used during calibration.

RC15: L264: NMAD has a precise definition which is different from the one given here (Höhle and Höhle, 2009). You should try another term like Standardized MAD.

AR15: We thank the reviewer and change accordingly to ”Standardized MAD”.

RC16: L268 and figure 5: can you colorcode data points with their elevation or whether they belong to the ablation or accumulation area? Why restraining to 90 days? Winter observations might be particularly useful to validate the model performance.

AR16: Regarding Figure 5, we color-coded the points above the median glacier elevation in blue, and the one below in red. See below for the new Figure with the respective new caption:

Furthermore, this validation was performed with the purpose of investigating the performance of the model for short-term changes. Therefore, we restricted to 90 days. In addition, for winter, normally very few short-term mass balance observations are available. Instead, the validation over longer scales, i.e. seasonal and annual, is done and shown with Fig. 4.

RC17: L284: start a new paragraph for annual values.

AR17: We thank the reviewer and changed accordingly.

RC18: Fig. 7: It would be better not to use a diverging colorbar for panel a, and to center the colorbar of panel b on zero (conventionally positive mass balance are in blue and negative ones in red).

AR18: We acknowledge that it might not be optimal to use diverging colorscales. However, with Fig. 7 we wanted to highlight spatial differences in average seasonal and annual mass balances. For this reason, we decided to use diverging colorscales centered around the mean values, even if it might be a bit counterintuitive at first.

We modified the figure caption to provide clarity:

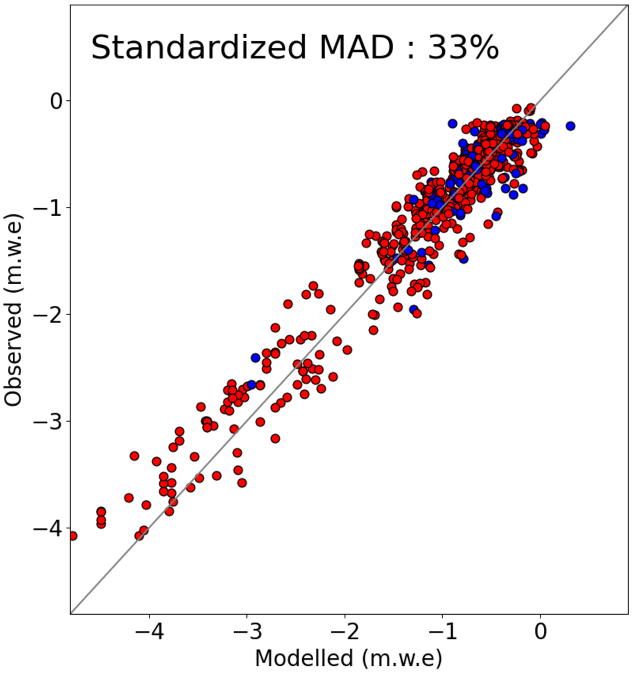

Figure 1: Validation of cumulated modelled daily mass balance against point mass balance measurements for periods spanning 7-90 days. Blue dots correspond to locations above the median glacier elevation, and red ones below it.

Average seasonal mass balance for the Swiss Alps over 2010-2024. (a) Average winter balance (30 April), and (b) average annual balance (30 Sept). Glacier-wide values are aggregated over a 20x20km grid with area-weighted averaging, whereby the size of the circles shows the glacierized area within each grid cell. Note that the color scales are centered around the mean to highlight regional differences.

RC19: L343-344: “which is within the respective uncertainty ranges” -¿ which uncertainty ranges are referred to? See my general comment.

AR19: In the new version of the manuscript it refers to the uncertainty that is calculated taking into account the four sources of uncertainty (see AR1).

We rephrased as follows to avoid confusion:

The Swiss-wide volume change over the study period agrees well for both approaches, with a difference in total volume change of approximately 10%, which is within the respective uncertainty ranges (see Sect. ”# of resp section” for the derivation of uncertainties).

RC20: L358-362: this interpretation could be developed further.

AR20: We reworded as follows under the consideration of the reviewer’s comment:

The second factor is associated with data availability. In the present study, we include newly-derived ice volume changes from a large set of digital elevation models covering the study period, while the statistical upscaling data product considered here (GLAMOS, 2024c; van Tiel et al., 2025) relied on longer-term and older geodetic mass balances (Fischer et al., 2015) for inferring glacier-specific mass change trends. This may be responsible for the systematic differences between the data sets as long-term glacier mass changes are given by different data. Additionally, the geodetic estimates from the two approaches cover different periods during which the conditions for glaciers changed considerably with potential consequences on the outputs of the respective approach.

RC21: Fig. 12: could you add a panel that shows the total annual mass balance for glaciers surveyed by GLAMOS only?

AR21: We thank the reviewer, but we think this goes beyond the scope of the Figure. The aim of the Figure is to discuss and compare the regional results with existing approaches, i.e. the statistical upscaling approach. Thus, adding data for individual GLAMOS glaciers would modify the purpose of the figure towards a more detailed glacier-by-glacier comparison, which we think would make the figure unnecessarily complex, and divert readers from the main message.

RC22: Conclusions: if I understand correctly your method does not need field measurements. Could you comment on its applicability to unsurveyed glaciers? What would be the limitations? Mereological forcings? DEM availability?

AR22: Correctly, our approach does not need field measurements for calibration. We are a bit confused by second part of the comment, as actually, it is already applied to unsurveyed glaciers in this study. If the reviewer means the applicability to other regions worldwide, yes, the main limitations would be the availability of the data, especially SCAF observations, but also DEMs and daily meteorological products. We suggest rewording the conclusion as follows:

In this study, we provide the daily mass balance for the $\sim$ *1400 glaciers in the Swiss Alps over the period 2010-2024. Relying on a modelling framework and exploiting machine learning techniques for spatial extrapolation, remote sensing observations of the snow-covered area fraction are combined with geodetic volume changes to separate the accumulation and ablation components, and thus better reproduce the seasonal variability of mass balance. We estimate a total ice volume change over 2010-2024 of* $-15.2 \pm 1.6\,km^3$*, i.e. almost 25% of the ice volume in 2010 was lost in 15 years. Swiss-wide seasonal average mass balances over the study period amount to 1.27 m w.e for winter and to* -0.95 *m w.e for the annual scale, respectively. The spatial variability of seasonal mass balance is further evaluated to capture regional differences, showing a clear pattern for winter accumulation with the highest amounts in central and western Switzerland ranging between 1.5-1.9 m w.e and lowest values are recorded in Valais with values of 0.9-1.2 m w.e. Winter accumulation exhibits high spatial correlation, revealing a strong dependence on precipitation patterns, whereas annual balances ranging between* -0.6 *and* -1.5 *m w.e are less spatially correlated, confirming the predominant effect of local topographical factors on annual balance as already documented in the literature. The high spatio-temporal resolution of our approach enabled us to better understand the relation between seasonal mass balance and the climatic drivers in the Swiss Alps, underlining the relevance of winter accumulation on the year-to-year spatial variability of mass balance. Furthermore, this study does not rely on direct field measurements for calibration and thus highlights the importance of remote sensing observations to provide reliable estimates of short-term mass balance of unmeasured glaciers at the scale of entire regions. The approach holds substantial potential, and its applicability to other regions worldwide mainly depends on the availability of the input data, i.e. SCAF observations, DEMs, and daily meteorological products. Overall, this suggests that observations of the snow-covered area fraction on glaciers may improve glacier-specific modelling at the large scale.*

References

E. Berthier, C. Vincent, and D. Six. Exceptional thinning through the entire altitudinal range of Mont-Blanc glaciers during the 2021/22 mass balance year. *Journal of Glaciology*, 70:e30, 2024. doi: 10.1017/jog.2023.100.

A. Cremona, M. Huss, J. M. Landmann, J. Borner, and D. Farinotti. European heat waves 2022: contribution to extreme glacier melt in Switzerland inferred from automated ablation readings. *The Cryosphere*, 17(5):1895–1912, 2023.

A. Cremona, M. Huss, J. M. Landmann, G. Schwaizer, F. Paul, and D. Farinotti. Constraining sub-seasonal glacier mass balance in the Swiss Alps using Sentinel-2-derived snow-cover observations. *Journal of Glaciology*, 71:e25, 2025. doi: 10.1017/jog.2025.1.

GLAMOS. Swiss glacier volume change, release 2024, glacier monitoring switzerland, 2024.

M. Huss, L. Dhulst, and A. Bauder. New long-term mass-balance series for the Swiss Alps. *Journal of Glaciology*, 61(227):551–562, 2015. doi: 10.3189/2015JoG15J015.

B. Menounos, M. Huss, S. Marshall, M. Ednie, C. Florentine, and L. Hartl. Glaciers in Western Canada-conterminous US and Switzerland experience unprecedented mass loss over the last four years (2021–2024). *Geophysical Research Letters*, 52(12):e2025GL115235, 2025. doi: 10.1029/2025GL115235.

A. Voordendag, R. Prinz, L. Schuster, and G. Kaser. Brief communication: The Glacier Loss Day as an indicator of a record-breaking negative glacier mass balance in 2022. *The Cryosphere*, 17(8):3661–3665, 2023. doi: 10.5194/tc-17-3661-2023.

xdem contributors. xdem, May 2021. URL `https://doi.org/10.5281/zenodo.4809698`.

Answer to Reviewer 2

In the following, we provide a point-by-point Author Response (AR) to any of the Reviewer Comments (RC) obtained for the manuscript that was under discussion. Suggestions for how the manuscript text could be revised are presented in *italics*.

Major Comments:

RC1: Temperature lapse rates: There are several issues regarding the treatment of temperature lapse rates in this study.

a) Initial lapse rate calculation: The authors calculate the initial lapse rates from the temperature and elevation of the cells next to the glacier (L170, L215). I am concerned that if the neighboring grid cells are not ice-covered but instead rock, forest, or other surfaces, their microclimate and hence surface temperature differs from that of glaciers. Consequently, the derived lapse rates may be inaccurate. I recommend calculating lapse rates directly from pressure levels at glacier grid cells, as for example done in Draeger et al. (2024) or Marzeion et al. (2012). Additionally, the manuscript does not report the initial lapse rates used. Please provide a range and compare them to commonly used lapse rates in global glacier modeling (e.g., the constant value of - 6.5 K km^{-1} in Schuster et al., 2023).

b) Lapse rate adjustments: The authors introduce a new third calibration step, where temperature lapse rates are adjusted to further optimize the model's reproduction of observed SCAF (L211–212), which is new compared to Cremona et al. (2025). It is unclear to me why the remaining bias in the third calibration step is attributed to the lapse rate rather than to near-surface temperature values. Previous glacier models have often applied a separate temperature bias correction in addition to lapse-rate correction (e.g., Rounce et al., 2020). To correctly separate the sources of bias, I recommend calculating more accurate lapse rates from pressure levels and framing the adjustment as an additive temperature bias correction instead of a temperature lapse rate adjustment. Conceptually, this distinction is important: lapse rate bias represents the bias in the vertical temperature gradient, while temperature bias reflects errors in near-surface temperature. This reframing would not change results, as the lapse rate correction is also applied additively (see minor comment on assumed typo in Equation (4)), but it would clearly disentangle the two effects and improve conceptual clarity.

AR1: We thank the Reviewer for the review of our manuscript and for the time invested in checking our submission. a) For what the specific comment is concerned, we entirely understand that this approach would be optimal and is often used in regional-scale glacier studies. However, in our case it is both not possible, and -– in our opinion -– not necessary. We elaborate our argumentation below in a new paragraph in the Discussion section to address the constraints and limitations with respect to the choice of air temperature lapse rates.

The calculation of the initial temperature lapse rate, based on the grid cells around the glacier's centroid (see Sect. 3.2), does not consider the surface type of the cells, which may lead to inaccurate estimates. Our temperature lapse rates range from –6.5 to –4.4 K km^{-1} with an average value of – 6 K km^{-1}, i.e. slightly lower than the value commonly used in global glacier modelling studies of –6.5 K km^{-1} (Schuster et al., 2023). While for the large glaciers, the surface type of most grid cells around the glacier centroid corresponds to ice, for small glaciers the surface type may be different, i.e. mostly rock. Thus, the different microclimate could affect the resulting temperature lapse rate. To overcome this challenge, other studies calculated lapse rates directly from pressure levels of gridded climate re-analysis datasets (e.g. Marzeion et al., 2012; Draeger et al., 2024). However, the high-resolution meteorological dataset for Switzerland does not provide the respective information. Also, gridded air temperature in the available data set is not derived from physically-informed modelling that distinguishes between glacier and non-glacier grid cells but is based solely on terrain elevation and interpolation of surrounding weather station data (MeteoSwiss, 2021). As a result, since most *meteorological stations are located below 2500 m a.s.l., air temperature lapse rates in high mountain areas are inherently poorly constrained. We thus adjust lapse rates during calibration, and we argue that applying a more sophisticated method to estimate their initial values would only have a minor impact.*

b) As described above, we now specifically discuss the constraints and limitations due to the choice of air temperature lapse rates in a new paragraph of the Discussion section. We also fully understand the concern raised by the reviewer here. Indeed, adjusting air temperature (in addition to the precipitation correction) is a common approach in global glacier modelling (e.g. Huss and Hock, 2015; Rounce et al., 2023; Schuster et al., 2023). We considered implementing a similar strategy here as well, however, we realized that it is not effective enough to improve model performance in both matching geodetic mass change rates and observations of the Snow-Covered Area Fraction (SCAF). The air temperature lapse rates in high-mountain areas are inherently poorly constrained. This is also mentioned by the review in the first comment. Although lapse rates are measured and well represented in the high-resolution meteorological data set used in our study below ca. 2500 m a.s.l. they are basically extrapolated above this elevation, which is most relevant for glaciers. We thus argue that air temperature lapse rates are actually a poorly constrained model parameter that can be used to optimize agreement with all available data sets. These explanations are now included in the additional Discussion paragraph (see above).

RC2: Extrapolation: The extrapolation relies only on topographic features and does not incorporate climatological variables, which could in my view introduce substantial uncertainty. L383–385 states: "Considering that the Swiss-wide average precipitation correction factor is 1.8 and the winter mass balance 1.27 m w.e., uncertainties in winter mass balance introduced during the extrapolation correspond to 20–25% on average." This is considerable, and it would be worth testing whether including climatological predictors per glacier (e.g., average mean summer temperature, average total winter precipitation, average temperature range, etc.) could improve the performance. Moreover, since SCAF observations are mostly available for larger glaciers (L59), it would be very valuable to include an analysis of extrapolation performance for smaller glaciers. As the training set is biased toward large glaciers, there is a risk that mass loss from smaller glaciers is not correctly modeled (underestimated?) when extrapolating parameters. For instance, you could compare the test-set performance between small and large glaciers.

AR2: This is a valid comment and we agree that testing additional variables can be beneficial. In response to this suggestion, we included more predictors in the machine-learning scheme for mass balance extrapolation and evaluated whether the results improve compared to those presented in the submitted version. In fact, we found that including factors such as the average mean summer temperature, average total winter precipitation and average temperature range does not deliver significant differences in the resulting mass balances. Therefore, we decided not to rely on these additional predictors to keep the number of variables entering the machine-learning approach as parsimonious as possible. This was added to the Method section as follows:

We also tested including additional predictors, namely the average mean summer temperature, the average total winter precipitation, and the average temperature range, which are related to the climatological conditions of each glacier and could thus potentially improve model performance. However, since these variables did not lead to a significant improvement of the results, we chose not to include them in order to maintain a parsimonious set of predictors for the machine-learning approach.

Regarding the second part of the comment, we added a paragraph to the Discussion section, discussing the limitations due to the small training data set which is biased towards large glaciers as follows:

Another limitation is associated with the machine-learning extrapolation approach. In fact, the *training data set for the machine-learning model is (i) relatively small, and (ii) biased towards large glaciers. This may lead to a bias in inferred mass balances, especially for small glaciers. Despite these limitations, the validation, which also includes small glaciers (about 40% of considered glaciers are smaller than 1 km^2) indicated that the extrapolation approach is also reliable for that size class (see Sect. 3.3). No bias in the resulting mass balance for small glaciers was detected. This confirms the method's reliability, but we suggest keeping in mind such limitations if the presented method is to be applied elsewhere.*

RC3: Evaluation: I am wondering why the authors have not compared their estimates against the seminal dataset by Hugonnet et al. (2021), which provides geodetic mass balance for all glaciers worldwide, including the Swiss Alps, for 2000–2019. This dataset is currently the most widely used for calibrating regional and global mass balance models. It would be valuable for the glacier modeling community to compare the mass balance results in this study with theirs and to discuss differences in the treatment of digital elevation models, etc.

AR3: We are aware of this global-scale data set, but we do not consider it directly comparable in the frame of the present study for the following reasons: (1) Time periods covered vary. A comparison may be possible using a sub-set of Hugonnet's pseudo-annual data over the common period, but it is known that the annual resolution of the product does not fully resolve actual mass balance variability. (2) There is a clear mismatch in the level of detail between the Hugonnet et al. (2021) data set and our digital elevation models and volume changes: Our data are based on high-resolution (25-50cm) orthophotos and locally referenced photogrammetry, whereas the global-scale data set is much coarser. This is especially relevant for the many small glaciers in the study region. (3) There is an inconsistency in the inventory data: Although Hugonnet relies on the Randolph Glacier Inventory from 2003 in the study region, we use the local Swiss inventory from 2016 (center of our study period) as a baseline. That inventory has a much higher level of detail and also resolves small and debris-covered glaciers.

For these reasons, a direct comparison would not help. We have nevertheless taken up this suggestion and now include a comparison of our annual resolution to the results presented by The GlaMBIE Team (2025) providing annual mass change at the level of the European Alps based on a combination of multiple methods, among others the Hugonnet et al. (2021) data set. The results are now described and indicate that:

Additionally, we compared our mass balance estimates with the one from The GLAMBIE Team (2025), derived for the European Alps. Despite the fact that the two datasets do not compare the exact same region, the results are consistent (Fig. 12b). The findings agree with the comparison with local data sets based on upscaling (GLAMOS, 2024; van Tiel et al., 2025), indicating that the mass balances generally compare well, with the main discrepancies observed in individual years, especially the extreme ones such as 2022 and 2023.

RC4: Area changes: It is unclear how exactly the area changes are calculated. The authors update the area-elevation distribution of their model domain "by attributing relative annual area changes equally to all bands below the mean equilibrium line altitude" (L201). Does this mean that each elevation band below the mean ELA changes by the same relative area? If so, this may not reflect reality, as lower-elevation bands near the terminus typically lose mass and area fastest. I am wondering how glacier retreat occurs in your model under this assumption? Applying uniform relative area changes could artificially retain lower bands, potentially exaggerating glacier-wide negative mass balance and contributing to the reported bias of -0.05 m w.e. (L253). Additionally, Cremona et al. (2025) report a 0.04 m w.e. difference in seasonal mass balance for a 0.5% yearly glacier area change (L377–379), which may explain part of the observed bias.

Regarding the "mean equilibrium line altitude" (L201), are you assuming a constant ELA throughout the study period? How was it calculated – for example, using median elevation as a proxy or derived from yearly mass balance values per band? Please clarify the rationale for assuming a constant ELA. I also suggest considering a sensitivity study to assess how changes in ELA would influence glacier-wide

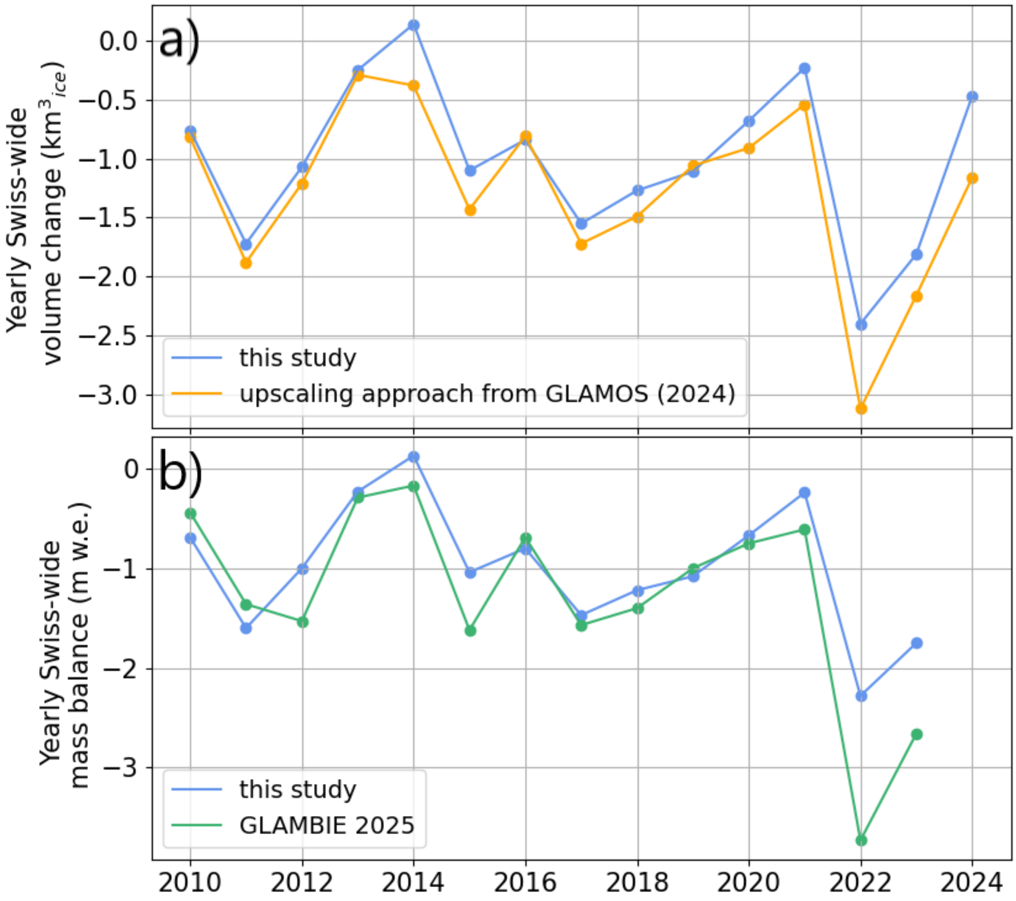

Figure 1: Comparison with existing approaches. a) Comparison of the yearly Swiss-wide ice volume change between this study (blue) and the upscaling of local mass balance measurements (GLAMOS, 2024c; van Tiel et al., 2025) (orange). b) Comparison of the yearly Swiss-wide mass balance from this study (blue) and the yearly mass balance for the European Alps from The Glambie Team (2025) (green).

mass balance.

AR4: We realize that the description of our procedure for estimating the changes in glacier area and their elevation-distribution was not clear enough and we have updated the text accordingly, thus also answering the concerns of the reviewer:

> *Since the area and volume for the inventory date are given, this year serves as our starting point for the glacier area updating forward (up to 2024) and backwards in time (to 2010). This is achieved by first calculating the cumulative yearly mass balance with Eqs. (1)-(3), which allows deriving the* ΔV_y*, and subsequently calculating the glacier area with Eq. (6) for the next time step (see also van Tiel et al., 2025). To update the area-elevation distribution of our model domain, it is assumed that all changes in glacier area occur in the ablation zone, i.e. below the median glacier elevation according to the geometry of the SGI 2016. We attribute computed relative annual area changes equally to all bands below this mean equilibrium line altitude, corresponding to the center of the study period. This means that each elevation band in the lower part of the glacier shows the same annual area changes in relative terms. Since this is a considerable simplification of actual retreat dynamics, the associated limitations are discussed in Sect. 4.4.*

As mentioned in the revised text above, we agree with the reviewer that the simple parameterizations to account for area changes and debris cover effects are first-order approximations that may not fully reproduce reality. However, they permit deriving reasonable regional estimates, which is the focus of the study. We added the following paragraph in the discussion section to address this point:

Another limitation is associated with the rather simplistic parameterizations employed to account for changes in glacier area and the effect of supraglacial debris. Both parameterizations are first-order approximations to account for these effects for the presented regional-scale assessment. We thus do not expect these approaches to exactly match actual processes everywhere, but assume them to be reliable on average. As a consequence, for example, elevation bands near the terminus are likely to lose mass and area with a somewhat lower rate than it is modelled. Nevertheless, since the study focuses on the regional scale, and bearing in mind the relatively short study period considered, we consider these first-order approximations to be reasonable assumptions.

RC5: Albedo parameterization: In general, net radiation, particularly shortwave radiation, is the dominant term in the surface energy balance of glaciers (Hock, 2005). Therefore, the albedo parameterization is critical. The authors use the parameterization by Brock et al. (2000), which is based on a single glacier in Switzerland with data from 1993–1994. However, albedo can vary significantly between glaciers, even within the same mountain range, and also shows strong temporal variability (Williamson et al., 2025). Moreover, the authors state in L309–311: “The high correlations [of winter precipitation anomaly with winter mass balance, and summer temperature anomaly with summer mass balance] confirm the predominant influence of temperature and precipitation on determining seasonal mass balance variations.” While temperature and precipitation are indeed important, the authors have not analyzed other drivers, such as albedo variations, even though they acknowledge their significance (L313). For example, Williamson et al. (2025) found that 31–41% of increased melt in Western North America and the Canadian Arctic is attributable to albedo decline. I therefore suggest: (1) Including a sensitivity study on the influence of the chosen albedo parameterization on mass balance and (2) analyzing correlations between albedo variations and mass balance to assess its potential impact. This would provide a more balanced assessment of the factors driving seasonal and annual glacier mass balance and avoid over-attributing effects to temperature and precipitation alone.

AR5: We fully agree with the reviewer that albedo variations are indeed important for glacier melt modelling. As suggested, we have now included a sensitivity study to assess the impact of the applied albedo parameterization on the results and found that the sensitivity is low. This is not surprising given that the model is tightly constrained with various observational data, hence leaving little room for the albedo parameterization to strongly affect trends in our results. Moreover, we also note that data for incorporating actual daily changes in regional-scale albedo (e.g. due to Saharan dust deposition) are not available and including such time-varying processes in our regional model would be beyond the scope of this study.

To test the sensitivity to the modelled glacier surface albedo, we tested two additional scenarios, i.e. by systematically varying the albedo by +25% and by −25%. The sensitivity of the resulting mass balance is low, i.e. with a standard deviation between the two scenarios of just 2% for both winter and annual mass balance. This is not surprising given that the model is tightly constrained with various observational data, thus leaving little room for the albedo parameterization to strongly affect trends in our results.

RC6: Description of data and methods: In the current draft, the data and methods sections are intermixed. For example, important methodological details, such as volume–area scaling, are presented in the Data section, which makes it difficult to follow and locate information. For better readability, I suggest clearly separating the Data and Methods sections.

AR6: We thank the reviewer for this comment and refer to AR4 of Reviewer Response #1.

Minor Comments:

RC7: Abstract: It would be good to mention that the authors are using a simplified surface energy balance melt model.

AR7: We added the information according to the reviewer's suggestion as follows:

> *To do so, we leverage two simplified surface energy-balance models and remote sensing observations, i.e. geodetic volume changes and observations of the snow-covered area fraction (SCAF) of glaciers during summer, together with machine-learning techniques for extrapolation purposes.*

RC8: L55: The authors state that "the last complete assessment of geodetic mass balance covering all glaciers of the Swiss Alps spans the period 1980–2010 (Fischer et al., 2015)". However, the dataset by Hugonnet et al. (2021) is not mentioned. I recommend that the authors acknowledge and discuss this dataset (see Major Comment 3).

AR8: Please also refer to AR3. Besides Hugonnet et al. (2021) there were also other regional-scale assessments of European glacier mass change (e.g. Sommer et al., 2020). However, these large-scale data sets are not comparable in terms of their resolution and the inventory baseline-data used. We have nevertheless added a note to refer to these additional studies:

> *Note that there are additional recent regional to global-scale assessments of the geodetic change of glacier mass on the scale of the European Alps (e.g., Hugonnet et al., 2021; Sommer et al., 2020), but these datasets are not comparable in terms of spatial resolution or the baseline inventory data used.*

RC9: L64: Typo

AR9: Thank you, we corrected accordingly.

RC10: L 68: Typo: should read "Fig. 2."

AR10: Thank you, we changed accordingly.

RC11: L 102: The manuscript uses the term "validation." Model validation usually refers to assessing model structure and assumptions during the development stage, whereas "evaluation" refers to assessing model performance afterwards. I argue that what is performed here is "evaluation," and suggest using this term consistently throughout the paper.

AR11: Thanks for this useful comment. This was changed accordingly.

RC12: L129-130: Typos

AR12: Thank you, we corrected the mistake.

RC13: L168: Equation (4) seems to contain a typo. I assume you mean T(z) = T + (z - z_{ref}) * deltaT/deltaz instead of multiplication.

AR13: Thanks for catching this! We corrected it.

RC14: L169-170: Please clarify how the temperature lapse rate was calculated, including the method used (linear regression?), the number of neighboring grid cells considered, the elevation range they cover, and whether extrapolation beyond this range was applied to higher or lower elevation bands.

AR14: Apologies for the missing information. We added it to the manuscript accordingly:

..., where T is the temperature at the reference elevation z_{ref} of the grid cell in the meteorological product closest to the glacier centroid, and dT/dz is the temperature lapse rate derived with linear regression from temperature values of the five surrounding cells.

RC15: L177-178: The manuscript states: "For glaciers larger than 2 km^2, we use the melt factor provided by Rounce et al. (2021)." It is unclear what is meant by "melt factor." Are you referring to the sub-debris melt enhancement factors from Rounce et al. (2021)? If so, please use the correct terminology consistently throughout the manuscript to avoid confusion with the melt factor (MF) in your model.

AR15: We apologize for the confusion caused by our wording. Yes, it is the sub-debris melt enhancement factor. We reworded as suggested.

RC16: L 179: "At each elevation band, the average debris melt factor is calculated." Please provide more detail on this calculation. The description of the methodology for dealing with debris cover is currently vague and would benefit from clarification.

AR16: We rephrased the subsection "Debris cover" as follows to promote clarity:

Supraglacial debris cover has a significant effect on melt enhancement or reduction for the ice underneath (Nicholson and Benn, 2012; Rounce et al., 2021). To account for this effect, the modelled melt resulting from Eqs. (2) and (3), corresponding to those of clean ice, is scaled with a sub-debris melt enhancement factor. A sub-debris melt enhancement factor below 1 indicates that the debris protects the ice, leading to reduced melt compared to clean ice. This is often the case for a thick layer of debris with a significant isolation effect on the underlying ice. Vice versa, a sub-debris melt enhancement factor above 1 indicates enhanced melt compared to clean ice, which is typical for a thin layer of debris that offers virtually no isolation effect but decreases the surface albedo. For glaciers larger than 2 km^2, we use the distributed sub-debris melt enhancement factor provided by Rounce et al. (2021), which is derived by relying on Landsat-8 images spanning 2013-2018 combined with modelling. At each elevation band, the average sub-debris melt enhancement factor from Rounce et al. (2021) ($\bar{E}_d$) is multiplied with the modelled melt from Eqs. (2) and (3) as follows:

$$a_{sfc,deb} = a_{sfc} * \bar{E}_d \tag{1}$$

Note that this occurs only for snow-free elevation bands, i.e. bands with a modelled snow depth larger than zero (see also Sect. 3.2). For glaciers smaller than 2 km^2, Rounce et al. (2021) did not calculate the sub-debris melt enhancement factor. For these glaciers, we derive the sub-debris melt enhancement factor for each elevation band by using the debris-covered area provided by the SGI2016 and assuming a debris melt factor of 0.63 over this area, i.e. the mean value found by Rounce et al. (2021). In this approach, the debris factor remains constant over time for each elevation band, i.e. debris thickness and characteristics are assumed to remain unchanged over the study period, under consideration of the geometry corresponding to the year of the SGI2016.

RC17: L180: "Only if the elevation band is snow-free." Please explain how you distinguish between snow-covered and snow-free conditions.

AR17: We refer to AR19, where we clarify how snow cover on the glacier ice is modelled.

RC18: L195: Consider rephrasing "define c" as "calculate c".

AR18: We re-worded as suggested.

RC19: L 203-226: It is unclear how SCAF modeling is performed. Please provide a brief overview of the methodology in addition to referencing Cremona et al. (2025).

AR19: Thanks for raising the important aspect. We added a brief description of how we model snow cover as follows:

To model the evolution of snow cover on top of the glacier ice, we assume an initial snow depth of zero across the entire glacier at the start of the study period. The snow water equivalent of each elevation band is subsequently updated according to the modelled climatic mass balance, i.e. the daily accumulation and melt. At each time step, the modelled snow-covered area fraction is derived by summing the area of all elevation bands with a snow depth greater than zero with respect to the beginning of each hydrological year, and dividing this total by the glacier area.

RC20: L208: Please provide a reference for this statement.

AR20: We added the following reference as requested to support the statement: Menounos et al. (2025).

RC21: L239: In addition to absolute error measures, please also provide relative error metrics for the machine learning model (e.g., mean absolute percentage error). Also, please specify the sizes of the training, validation, and test sets.

AR21: We agree with the reviewer that the manuscript would benefit from this information. For changes to the text, we refer to AR2 of Reviewer Response #1.

RC22: L247: The phrase "To address the uncertainty" would be better phrased as "To evaluate the model," since the uncertainty itself is not being addressed here.

AR22: We agree with the reviewer and re-worded accordingly.

RC23: L384: The authors mention the average precipitation correction factor. It would be helpful to also report the mean and ranges of the other calibrated parameters in the manuscript.

AR23: We agree with the reviewer and added the ranges of the parameters and their average of the investigated period as follows:

Across Swiss glaciers, MF ranges from 0 to 9 $\mathrm{m\,w.e\,K^{-1}\,d^{-1}}$*, with a median of 1.79* $\mathrm{m\,w.e\,K^{-1}\,d^{-1}}$*,* $a_{snow/ice}$ *ranges from 0 to 0.6* $\mathrm{m\,w.e\,m^{-2}d^{-1}\,W^{-1}\,K^{-1}}$*, with a median of 0.016* $\mathrm{m\,w.e\,m^{-2}d^{-1}\,W^{-1}\,K^{-1}}$ *for the temperature-index model. For the simplified energy-balance models, TF ranges from 0.4 to 8* $\mathrm{m\,w.e\,K^{-1}\,d^{-1}}$*, with a median of 2.90* $\mathrm{m\,w.e\,K^{-1}\,d^{-1}}$*, and SRF ranges from 0 to 2.4* $\mathrm{m^3\,d^{-1}\,W^{-1}}$*, with a median of 0.06* $\mathrm{m^3\,d^{-1}\,W^{-1}}$*.*

RC24: Figure 4: Please explain all abbreviations (ALE, RHO, etc.), as these have not been introduced in the manuscript.

AR24: We agree with the reviewer and now use the glacier names in the figure instead of the abbreviations (see below).

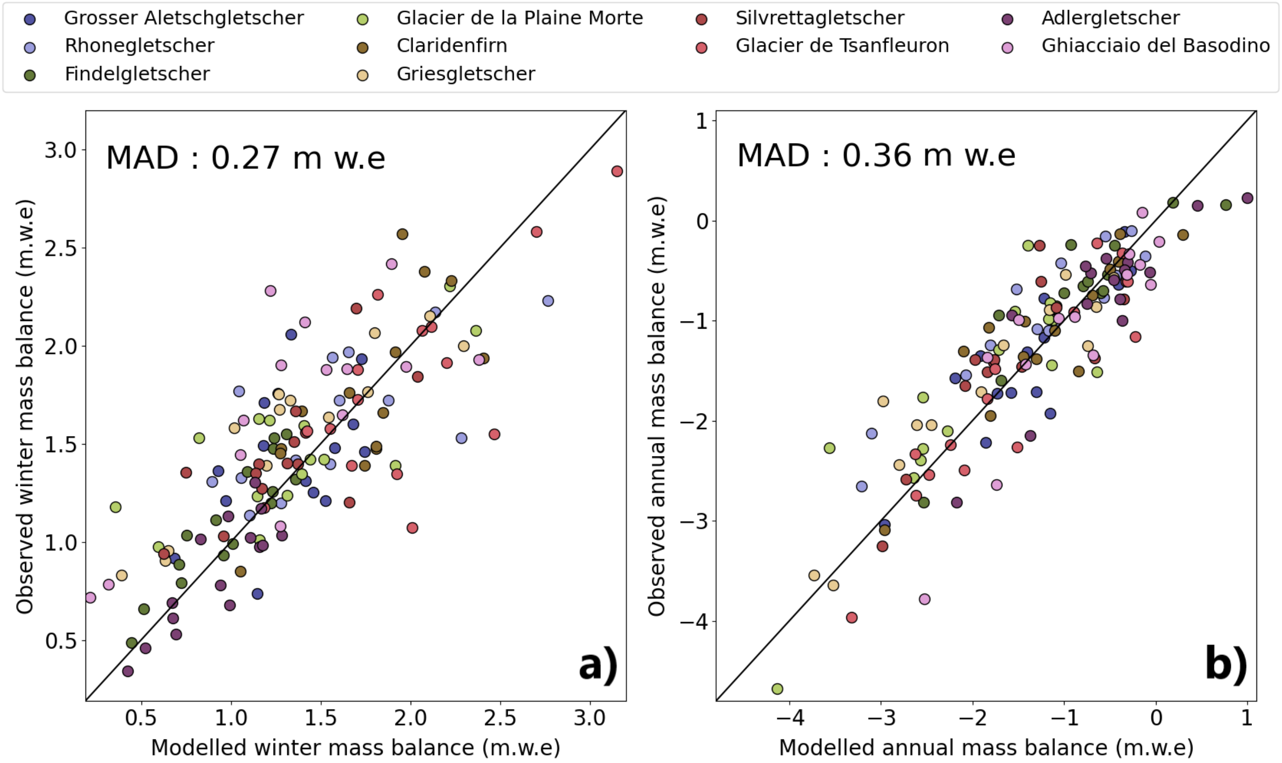

Figure 2: Validation of seasonal glacier-wide mass balance. a) Comparison between modelled and observed winter mass balance. b) Comparison between modelled and observed annual mass balance.

References

C. Draeger, V. Radić, R. H. White, and M. A. Tessema. Evaluation of reanalysis data and dynamical downscaling for surface energy balance modeling at mountain glaciers in western canada. *The Cryosphere*, 18(1):17–42, 2024. doi: 10.5194/tc-18-17-2024.

GLAMOS. Swisswide glacier mass balance, release 2024, glacier monitoring switzerland, 2024.

R. Hugonnet, R. McNabb, E. Berthier, B. Menounos, C. Nuth, L. Girod, D. Farinotti, M. Huss, I. Dussaillant, F. Brun, et al. Accelerated global glacier mass loss in the early twenty-first century. *Nature*, 592(7856):726–731, 2021. doi: 10.1038/s41586-021-03436-z.

M. Huss and R. Hock. A new model for global glacier change and sea-level rise. *Frontiers in Earth Science*, 3:54, 2015. doi: 10.3389/feart.2015.00054.

B. Marzeion, A. Jarosch, and M. Hofer. Past and future sea-level change from the surface mass balance of glaciers. *The Cryosphere*, 6(6):1295–1322, 2012. doi: 10.5194/tc-6-1295-2012.

B. Menounos, M. Huss, S. Marshall, M. Ednie, C. Florentine, and L. Hartl. Glaciers in Western Canada-conterminous US and Switzerland experience unprecedented mass loss over the last four years (2021–2024). *Geophysical Research Letters*, 52(12):e2025GL115235, 2025. doi: 10.1029/2025GL115235.

MeteoSwiss. Documentation of MeteoSwiss Grid-Data Products: Daily Mean, Minimum and Maximum Temperature: TabsD, TminD, TmaxD. Available at: https://www.meteoswiss.admin.ch/content/dam/meteoswiss/de/service-und-publikationen/produkt/raeumliche-daten-temperatur/doc/ProdDoc_TabsD.pdf, 2021. Last accessed: September 2021.

L. Nicholson and D. I. Benn. Properties of natural supraglacial debris in relation to modelling sub-debris ice ablation. *Earth Surface Processes and Landforms*, 38(5):490–501, 2012. doi: 10.1002/esp.3299.

D. R. Rounce, R. Hock, R. McNabb, R. Millan, C. Sommer, M. Braun, P. Malz, F. Maussion, J. Mouginot, T. Seehaus, et al. Distributed global debris thickness estimates reveal debris significantly impacts glacier mass balance. *Geophysical Research Letters*, 48(8):e2020GL091311, 2021. doi: 10.1029/2020GL091311.

D. R. Rounce, R. Hock, F. Maussion, R. Hugonnet, W. Kochtitzky, M. Huss, E. Berthier, D. Brinkerhoff, L. Compagno, L. Copland, D. Farinotti, B. Menounos, and R. R. McNabb. Global glacier change in the 21st century: Every increase in temperature matters. *Science*, 379(6627):78–83, 2023. doi: 10.1126/science.abo1324.

L. Schuster, D. R. Rounce, and F. Maussion. Glacier projections sensitivity to temperature-index model choices and calibration strategies. *Annals of Glaciology*, pages 1–16, 2023. doi: 10.1017/aog.2023.57.

C. Sommer, P. Malz, T. C. Seehaus, S. Lippl, M. Zemp, and M. H. Braun. Rapid glacier retreat and downwasting throughout the european alps in the early 21st century. *Nature communications*, 11(1):3209, 2020. doi: 10.1038/s41467-020-16818-0.

The GLAMBIE Team. Community estimate of global glacier mass changes from 2000 to 2023. *Nature*, 639:382–388, 2025. doi: 10.1038/s41586-024-08545-z.

M. van Tiel, M. Huss, M. Zappa, T. Jonas, and D. Farinotti. Swiss glacier mass loss during the 2022 drought: persistent streamflow contributions amid declining melt water volumes. *EGUsphere*, 2025:1–29, 2025. doi: 10.5194/egusphere-2025-404.

Answer to Community Comment

In the following, we provide a point-by-point Author Response (AR) to any of the Community Comments (CC) obtained for the manuscript that was under discussion. When presenting suggestions for how the manuscript text could be revised, text is presented in *italics*.

CC1: Sect 3.1: The part of the third paragraph on assigning a date to each DEM is hard to follow. In the example discussed and in Fig 2, why do you omit the DoDs of 2009-2018 and 2008-2018 (blue) and 2008-2015 (red)? The three blue DEMs (2009, 2018 and 2023) seem to be producing inconsistent thinning estimates: '09-'18 $\sim$ 0.9, '18-'24 $\sim$ 0.9, but '09-'24 $\sim$ 1.8?

AR1: We rephrased the mentioned paragraph as follows aiming for more clarity:

However, for some of these DEMs the actual acquisition dates were not traceable and only the year the DEM was known, thus the exact period of the corresponding ice volume change observations could not be resolved. To address this drawback, we used the consistently reported dates of the SNFI DEMs to assign the missing dates to the swisstopo DEMs, since both datasets are based on the same aerial image acquisition flights (see Section 2). Therefore, when only one DEM is available for a given year, the date of the SNFI DEM is also assigned to the swisstopo DEM. In years with multiple DEMs with different acquisition dates, the mean acquisition date of the SNFI DEMs for that year is taken and assigned to all swisstopo DEMs in the respective year. The standard deviation in the acquisition date indicates the temporal spread of the DEMs within that year. Ice volume changes with lower standard deviations in acquisition dates are favoured to reduce uncertainties and improve the temporal constraint on the time period of ice volume changes based on swisstopo DEMs (GLAMOS, 2024d).

The ice volume changes mentioned in the comment are missing because either the DoD had a void ratio larger than 50% or because the spread of DEMs within one year, which determines the uncertainty in the date to be assigned, was too high.

Fig. 2b shows the geodetic ice volume change and not mass balance. Therefore, we understand that it could be source of confusion. Consequently, we now change Fig. 2b and show geodetic mass balance instead of volume change. This allows for better comparability and should promote better clarity.

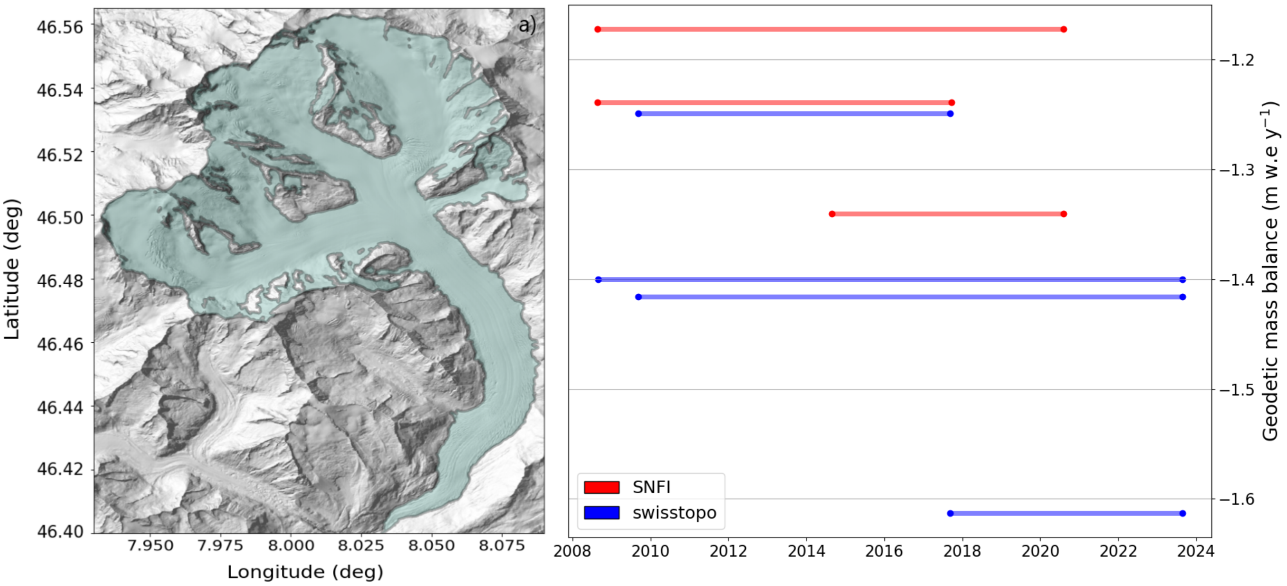

Figure 1: Temporal coverage of the geodetic mass balance estimates available for Grosser Aletschgletscher. a) Overview of Grosser Aletschgletscher with glacier outline from the SGI2016. b) Geodetic mass balance based on the SNFI DEMs (red) and swisstopo DEMs (blue).

CC2: Sect 3.2.1: The melt factor is not defined with an equation. What is “debris-coverage” - the fraction? the total area? How is it kept constant as the hypsometry evolved? How do you, if at all, validate the performance of the subdebris-melt module?

AR2: We rephrased the subsection ”Debris cover” as follows to promote more clarity. For details, we refer to AR16 of Reviewer Response #2

CC3: Sect 3.2.2: Why not validate the performance of Eqs. 5 and 6, along with a glacier-specific c, using the observed geodetic balance and area-change data, wherever these are available? Please see the limitation of these method pointed out in the comments in the attached pdf. How are the issues of a different scaling exponent, stagnation of tongue, etc., known for the debris-covered glaciers taken care of in this evolution model?

AR3: In our opinion, the above-mentioned validation approach would be beyond the scope of the study. After all, we clearly acknowledge in the manuscript that this approach is rather simplistic and allows us to roughly estimate a glacier-specific area correction for every considered glacier for the few years after the inventory date. This is not a study about a glacier evolution model. The estimated area changes have a minor effect on the results, as is now shown with the added uncertainty assessment but we still wanted to integrate this factor relying on a simple and straight-forward approach. In the revised manuscript, this was expanded though in response to other reviewer comments. For changes in the manuscript we refer to AR4 of Reviewer Response #2

CC4: Sect. 3.2.3: Why do you consider only the difference in various geodetic estimates as the only source of error? What happens to the glaciers with only one or two geodetic data points? Is the final prediction uncertainty comparable with the spread in the geodetic data, or is it much less than that? If it is the latter, then how can you explain that?

AR4: We performed a new uncertainty assessment, including various sources of uncertainty. We refer to AR1 in the answer to review #1 for what this comment is concerned, and for the respective changes to the revised manuscript.

CC5: Sect 3.2.4: You seem to extrapolate only cprec . However, it is not apparent from the heading or the first several sentences of the section. Also, the rationale for doing this is not given. The best-fit values/range of cprec remained hidden for me even as I searched both the papers (this and the JoG one). Therefore, I do not have any clue if an RMSE of 0.37 is good or bad. Moreover, the consequent uncertainty in the computed winter and annual balance is not discussed. There is a fair chance that this could be very significant.

AR5: We apologize for the lack of clarity caused by our chosen section heading. We suggest rewording the heading of the section to *Extrapolation of the precipitation correction factor* to avoid confusion. The precipitation correction factor has to be extrapolated because of lacking SCAF observations for the remaining glaciers (see also Sect. 2 for the availability of SCAF observations). We suggest rewording as follows to avoid any misunderstanding:

Therefore, the parameters derived for the 87 glaciers must be extrapolated to the remaining Swiss glaciers that lack SCAF observations.

We also added the following sentence, giving the range and average values for c_{prec}, so that the RMSE of 0.37 is better put into context.

This model is employed to extrapolate the precipitation correction factor c_{prec} *to approximately 1300 glaciers. The resulting* c_{prec} *values range from 1.1 to 3.5, with a Swiss-wide area-weighted average of 1.8.*

Regarding the last part of the comment, the uncertainty introduced by the extrapolation is accounted for the new version of the uncertainty assessment. We refer to AR1 in the answer to review #1.

CC6: Sect 3.3: The logic behind the choice of the 10 glaciers is not explained (L250).

AR6: We apologize for not providing full insights. The selection of the ten glaciers is motivated by the intersection of glaciers that are part of the GLAMOS monitoring programme (and thus have detailed in situ seasonal mass balance observations available) and that have been calibrated with SCAF observations (i.e. not extrapolated c_{prec}). In the revised manuscript, we reworded as follows:

Seasonal glacier-wide mass balance is validated on a subset of ten glaciers (Fig. 4) to evaluate the performance of the calibration relying on geodetic ice volume changes and SCAF observations (see Sect. 3.2.3). These ten glaciers were selected as they are part of the GLAMOS long-term monitoring programme and have SCAF observations available for calibration.

CC7: There is no validation attempted for the daily-scale product. It is hard to imagine that constants like c_{prec}, which are derived using a seasonal-scale mass balance, will capture the variability on a daily scale. While the model resolution is daily, robust estimates may require some averaging.

Strangely enough, the MAD is given as %-age for shorter time scales, while the actual values are given in the seasonal and annual scales. The scatter in fig 5 does not scale with the value, and the noise appears additive (as opposed to multiplicative). The biases are omitted for the subseasonal case as well. In fact, In all three cases, the scatter plots show that there is a significant variation in MAD and bias from one glacier to another. Its consequence for regional to catchment scale estimates remains to be assessed.

It is unclear if the subdebris melt estimates were included in the validation. If so, what is the corresponding MAD?

AR7: Regarding the temporal scale of the results, i.e. whether the daily-scale model output can be validated, we refer to AR7 in the answer to review #1, where the same point was raised. Regarding the MAD of the daily scale validation, it is given in percentage simply because we compare different period lengths and thus different magnitudes of mass balance. For example, a same error of 0.1 m w.e. over a 10-day period could otherwise not be confronted to a 100-day period. Therefore, it is difficult to put such a bias into perspective, the reason because it is omitted. For the annual and seasonal scale, however, the temporal scale is fixed, and we therefore provided absolute values as suggested.

As already mentioned, we cannot expect that our method will have the same performance for all glaciers. Since a similar point was mentioned in CC3, we refer to AR3 for details and related changes to the manuscript.

Yes, the validation includes the effect of supraglacial debris. We added the following sentence:

Note that for model validation, the effect of supraglacial debris and changes in glacier area were included as described in the respective sections.

CC8: The limitation of the uncertainty analysis is another factor that, in my view, weakens the paper. Only using the spread of the geodetic balance values – that too without taking into account their uncertainty – is done without any serious justification. The reference to the previous JoG paper actually increases the confusion – I did not find any clear computation of the uncertainty in mass balance there (see comment in attached pdf). The MAD is not the same as the uncertainty.

The Fig. 5 of the JoG paper may suggest SCAF has less scatter for the larger glaciers like Aletsch, which is not unexpected. However, it also had the largest MAD for annual balance, which was about twice as large as the quoted mean. The Fig. 4 of the present paper also suggests a possibly larger scatter of winter balance for some of the smaller glaciers. However, the scatter for some of the smaller and larger glaciers is similar for the annual balance. Trends like this need to be investigated carefully to avoid uncertainties and biases present in the fitted model.

AR8: We acknowledge the incomplete uncertainty estimation in the submitted version (see also replies to both reviewers), and we invested considerable effort to perform an improved uncertainty assessment. Again, we refer to AR1 in the answer to review #1 for details on this. Of course, it is expected that the method has different performances on different glaciers, and as discussed in Cremona et al. (2025), the method is likely to be challenged more on small glaciers than on large glaciers. However, after checking the individual MADs and biases, no significant trend attributable to glacier size became evident. We refer to AR2 of Reviewer Response #2 for changes to the manuscript.

CC9: In fact, the biases on individual glaciers, including some of the large ones, may be significantly higher than the mean quoted in the text (L257). The biases are not discussed much except giving a mean bias, which is going to be small because of the oscillating sign. Could you check the mean absolute bias? A clear systematic bias also shows up in Fig5 , with the model overestimating melt wherever the observed melt is less (more negative) than -2 m/y. This bias and its effect on the model output, which are completely ignored, are likely to be significant, particularly on the extreme years, and demand a thorough analysis.

AR9: Unfortunately, we do not exactly understand the reviewer's concern here. In the text both the bias and the mean absolute difference (i.e. accounting for the oscillating sign) are shown and discussed. Furthermore, both the results shown in Figs 4 and 5 do not indicate any systematic skew towards higher model misfits at very negative mass balances in our opinion. As all results are clearly presented for the inspection of the reader, we would not see what additional analysis could be performed here. Of course, we acknowledge that for some glaciers, the disagreement can be substantial. For example, for Plaine Morte and Basodino, there's a model error of more than 1 m w.e. in one individual year (see Fig. 4b). But we feel that it is not possible to specifically discuss every individual data point, given that the oveall statistics are shown and are favourable.

CC10: It was not demonstrated if that the subdebris melt and glacier-geometry evolution, two pieces that were added to the model here, actually improves/changes the estimates. How much they increase the uncertainties needs to be considered as well, given the known limitations of the specific schemes used (see attached pdf).

AR10: Although the subdebris melt and the glacier-geometry evolution are methods used for first-order approximation, they are conceptually important components of the system. Even though they may not be expected to fully resolve all physical processes in full detail for every glacier, they will provide better estimates than not considering those effects at all. We however agree that some clarification and better description of those approaches was needed and have implemented this in the revised manuscript. Please refer to AR4 of Reviewer Response #2.

CC11: It may be true that the set of independent geodetic mass balances for a glacier may have a large spread. However, that may not help in getting an accurate measure of uncertainty, as these data sets may have different error bars, and sometimes there may not be enough measurements. Another alternate approach, which we had taken in Banerjee et al., 2022, JoG, is to add appropriate noise in a Monte Carlo and compute the mass balance for each case to generate a large ensemble to obtain the error bar.

AR11: We agree and have therefore updated our uncertainty assessment. We refer again to AR1 in the answer to review #1.

CC12: Another related question: could you add the modelled mean balance for Aletsch in Fig. 2b, with an uncertainty band, and see how it compares with the spread of the input geodetic values, which vary by almost 100%. Since this is the basic input calculation are based on, the uncertainties in output should be comparable to this spread. If the procedure yields a lesser spread, the questions would be why and how. That's where the uncertainty band in Fig6 may require a careful revisiting.

AR12: We appreciate the suggestion, however, we believe this addition would go beyond the intended scope of Fig. 2. The purpose of this figure is primarily to illustrate the availability of the input data rather than to perform a quantitative comparison between modelled and observed balances. Such comparisons are already presented and discussed in the validation section, where model outputs are compared against mass balance observations. Also, we revised Fig. 2 as it was source of confusion, and we thus refer to AR1 for the changes.

CC13: The model uncertainty as a function of time scale of prediction, starting from days to decades, would be a good addition to the paper.

AR13: We agree that this would be a good addition to the manuscript and accordingly added this. We refer to AR1 in the answer to review #1 for changes to the manuscript.

CC14: The discussions can be more mindful of the model assumptions, and the inferences have to be substantiated using the model output. For example, to describe the modelled variability, you refer to things like avalanching, wind-driven redistribution, Saharan dust, etc., none of which are present in the model! Also, while you consider the role of P and T, you do not look into radiation and SCAF to understand the variability of observed mass balance, particularly that of the annual and summer balance. According to me, some really interesting features in our output (e.g. see the comments to Fig. 10, in the annotated PDF) are overlooked. I get the feeling the powerful, detailed data set of forcing and response that you produce can be exploited a bit more.

AR14: We agree and adjusted and/or added several paragraphs in the Discussion section touching upon these points. We refer to the two Reviewer Responses for specific changes to the manuscript.

CC15: The writing and referencing leave scope for improvements. I have pointed out several instances of complicated sentences or incomplete information in the annotated PDF attached, which should help illustrate the point.

AR15: Many thanks for this effort - highly appreciated. Considering these suggestions, the text was rephrased and adjusted throughout the paper.

CC16: The authors may want to look beyond a set of papers familiar to them and look harder for the most appropriate references. While I could not check all the references, I did find a few surprising inclusions. Here are a few examples:

L19 van Tiel et al. (2025), which deals with buffering of runoff in an extreme year, may not be the most suitable reference for general properties of runoff from glacierised catchments. I remember referring to the excellent paper by Hock, Jansson, and Braun (2005) while discussing runoff variability.

L22 From the titles of the three references cited, while not totally unrelated, they do not seem to be the most appropriate ones on the topic of " increasing interest in accurate monitoring of glacier mass balance and runoff on the regional scale".

L25 "Dussaillant et al., 2018; Denzinger et al., 2021; ... " how were these papers chosen, when you want to introduce something as basic as geodetic mass balance? While these are interesting global and local scale studies, they may not serve as a basic introduction to the method and are also not the most relevant ones for your study area, as far as my limited understanding of the topic goes.

L225 Cremona 2025 JoG does not really show that as far as I can tell. See the comment in the annotated PDF.

AR16: Thanks for this critical re-assessment of the chosen references. We have carefully gone through the text again and added/exchanged references as suggested.

CC17: Sect 4.4 A quantitative comparison with only one of the previous reconstructions (where some of you are coauthors) is a missed opportunity I feel. Can you not compare with the other existing studies? Are there other reconstructions available for the Swiss glaciers – at least at annual scale, or for specific glaciers or regions – beyond van Tiel 2025? Apart from van Tiel 2025, you only mention Dussaillant 2025, but do not compare with their results. As it stands, the claim made in L355 is unsubstantiated. Are there not other reconstructions, say, based on remote-sensing proxies like snowline etc.? They are many in the Himalayan literature that I am more familiar with. I am half sure they may be there Swiss Alps. In general, there may be a need to connect better with the existing literature and work by other groups, through improved referencing, and by incorporating results from such studies.

AR17: No, to our knowledge, there are no other comparable reconstructions at the mountain-range scale in the region and at a (sub-)annual resolution. Of course, several studies purely based on the geodetic method are available providing long-term mass changes (e.g. Sommer et al., 2020; Hugonnet et al., 2021) but these do not add useful information in the present context. We now, however, explicitly include a comparison with The GlaMBIE Team (2025), relying on a similar approach as discussed in Dusaillant et al. (2025) and van Tiel et al. (2025). We refer to AR3 of Reviewer Response #2 for changes to the manuscript.

References