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:*

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

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 $a_{snow/ice}$

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:

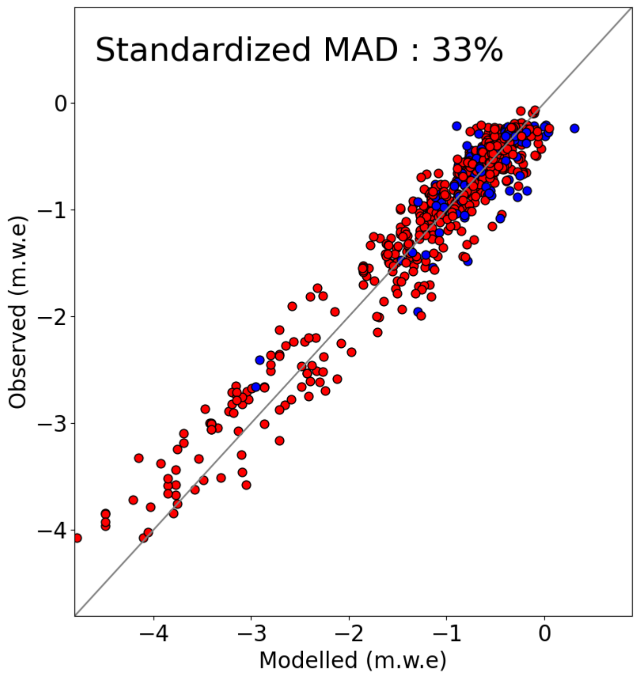

Figure 1: Validation of cumulated modelled daily mass balance against point mass balance measurements for periods spanning 7-90 days. Blue dots correspond to locations above the median glacier elevation, and red ones below it.

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`.