

# Reply on RC2

The paper applies FROST (Herrmann et al., 2025), which combines an ensemble Kalman filter with IGM, to infer SMB/ELA-related parameters from elevation change data. Compared to the original FROST study on a single glacier, the current study extends the framework to a regional Alpine application. A useful additional contribution is the snowline altitude (SLA) product derived from satellite optical imagery, which should be stated more clearly in the abstract as part of the paper's contributions. Overall, the manuscript presents a relevant regional-scale demonstration of the framework. I do not have major concerns, only a couple of minor-to-moderate comments and questions that would help clarify a few points and limitations.

Thank you for your review and the positive feedback. We would like to clarify that the SLA product is not a contribution of this paper, but is a product of a parallel study, which is now available as a preprint (Sommer et al., 2026). We have added this reference in the manuscript and refer to it for details on the SLA dataset. The other points are addressed individually below.

Sommer, C., Groos, A. R., Fürst, J., and Braun, M.: Transient snow line altitudes of glaciers in the European Alps from multi-mission remote sensing data (2000–2025), *Earth Syst. Sci. Data Discuss.* [preprint], <https://doi.org/10.5194/essd-2026-35>, in review, 2026.

## General comments

### 1. Uncertainties in the evaluation against GLAMOS:

- "We select glaciers with more than five years": is this enough to neglect ELA / MB-gradient variability over time? I would look (maybe in the supplement) at the variability of the average ELA over a certain window (>5 years) relative to the 20-year average, using glaciers that have full temporal coverage (assuming there are some).

Thank you for this valuable suggestion. We implemented your idea and added Figure S9 in the supplement, showing the difference between the mean of a reduced sample set and the full 20-year mean. We find that using five years results on average in a deviation of about 50 m from the 20-year mean, with some outliers reaching differences of up to 200 m. Based on this analysis, we increased the threshold for glaciers included in the evaluation to 10 years. This removes only Vorabgletscher from the GLAMOS dataset, and Venedigerkees, Zettalunitzkees, Seekarlesferner, and Vedretta Occidentale di Ries from the WGMS dataset, all of which have between 5 and 10 years of mass balance measurements. This reduces the mean sampling error to roughly 25 m and limits outliers to about 100 m. We additionally include the sampling error in the method and refer to it in the discussion.

- Second, regarding the MB gradients, are the actual profiles piecewise linear (with only two segments)?

Thank you for this question. The extraction of MB gradients from the glaciological measurements was indeed not sufficiently explained in the Methods section. GLAMOS and WGMS provide mass balance values per elevation bin over multiple years. We compute the 20-year mean for each elevation bin and fit a bilinear function that is split at the ELA.

We are aware that this approach reduces the information contained in the glaciological measurements, and that a comparison to direct point measurements would provide a more direct link to observed mass balance. However, our focus is on evaluating the calibration framework rather than the surface mass balance model itself. By aggregating the glaciological measurements to quantities that are directly comparable to the SMB model parameters, we enable a more consistent evaluation of the calibration results and reduce the influence of the simplified SMB formulation on the evaluation.

We added an exemplary figure S10 to illustrate how well the piecewise linear function represents the mass balance profiles provided by WGMS.

2. Uncertainties in the SLA: was the methodology evaluated in a prior work? How accurate should the estimated SLAs be? And the temporal-coverage concern I raised in the previous point could also be mentioned here. I think I would at least look at how well the SLAs align with GLAMOS ELAs for glaciers with a good temporal overlap.

We apologize for the confusion. The SLA product is developed as part of a parallel, ongoing study, which is now available as a preprint (Sommer et al. 2026). This work includes an extensive evaluation of the SLA retrieval, including a comparison with glaciological ELA data.

3. In the FROST paper I see that there is a bin-size parameter: how was it chosen here (per glacier?), and how does it influence the precision of the inverted ELA?

Thank you for this question. The elevation bin size is set to 50 m for all glaciers. The influence of this parameter on the precision of the inferred ELA is evaluated in our previous study (Herrmann et al., 2025, Fig. 7e-f). There, we show that the resulting error remains below ~20 m and does not decrease substantially for smaller bin sizes, while computational costs increase.

We therefore chose 50 m as a trade-off between computational efficiency and accuracy and added this information to the Methods section.

4. Since it's a Brief Communication, I would add a few more "so what" statements (including in the abstract). Otherwise, it sounds mainly like an application paper where FROST is further validated.

Thank you for this suggestion. The study serves as a simplified test case in which we reduce the temporal resolution of the observations, while applying a method that is inherently designed for truly transient data assimilation, where observations are incorporated at their respective times of acquisition. We have clarified the broader goal of our work and the context of this study in the abstract, introduction, and conclusion.

Line-by-line comments

- 19: Surface Mass Balance (SMB) is first used here

Corrected

- 23: I would cite also OGGM v1.6 (<https://doi.org/10.5281/zenodo.7718476>), since only starting from 1.6 the geodetic MBs are used, as far as I know. Since you rely on OGGM for data processing as well, it's good to refer to a specific version.

Thank you for specifying the version of OGGM which uses geodetic MBs. We included it in the revised manuscript.

- 32: See AGILE v0.1 (Schmitt et al. 2026) --- very recently published but fits quite well to this discussion

Thank you for your comment. We are very familiar with this work and agree that this relevant study was missing from our introduction. While the AGILE appears similar to our approach, there are three key differences. First, AGILE is based on OGGM as a flowline model, whereas we use the 3D IGM. Second, the primary objective differs: AGILE focuses on deriving ice thickness, while our framework aims to infer SMB parameters. Third, the optimization algorithm differs, as AGILE relies on gradient-based optimization, whereas we use an ensemble-based method and statistically update the parameters. The gradient-based approach enables more efficient optimization but requires differentiable model components. In this sense, AGILE is more comparable to the IGM inversion step than to the SMB calibration framework itself.

- 42: I would briefly justify the 1 km<sup>2</sup> choice.

Thank you for this comment. We added a brief justification for the 1 km<sup>2</sup> threshold in the manuscript. The threshold is chosen to ensure that glacier areas are sufficiently large to resolve spatial patterns in elevation change and surface mass balance relative to observational noise.

- 47: "more robust" compared to what?

Thank you for the question. We clarified this sentence in the manuscript. Here, "more robust" refers to the comparison between the 20-year elevation-change product and shorter time intervals (e.g., four consecutive 5-year periods). While the shorter intervals provide a temporal resolution, they come with higher uncertainty, making the 20-year aggregate more robust compared to the shorter period.

- 49: I would also add the GLAMOS citation here.

Done.

- 58: Also for the accumulation area?

Yes, the density is constant over the entire glacier, including the accumulation area. We agree that the effective density in the accumulation area is likely lower, potentially by up to ~50%. This could influence the estimated accumulation gradient. However, in our case, the accumulation area is relatively small in most years, which limits the impact on the overall mass balance, and we do not observe a systematic shift in the inferred gradients. Nevertheless, this assumption may introduce larger biases for glaciers with more extensive accumulation areas.

- 75-77: Was the method applied on both bands separately, or on some index?

We used a band ratio of near-infrared (NIR) and shortwave infrared (SWIR) as proposed by Li et al. (2022) to differentiate between snow and glacier ice:  $\text{Ratio} = \text{NIR} * (\text{NIR}/\text{SWIR})$ . The main advantage of this band ratio is that it provides a more precise spectral separation between the two classes than using NIR reflectance alone. The full methodology of the SLA mapping is described in the recently published preprint Sommer et al. (2026), alongside a number of sensitivity studies.

We apologise for the relatively imprecise description of the snow line measurement here. This was due to the length constraints of this manuscript and the planned separate publication in Earth System Science Data. The dataset and processing code are now available here (<https://doi.org/10.5281/zenodo.18223929>) and here (<https://github.com/Reply on RC2cr-610>).

- 77-79: How exactly was the SLA estimated from the snow masked DEM?

As the transition between snow and bare ice is typically not a distinct line, but rather an area of mixed snow and ice, we did not use the lowest snow-covered DEM pixels to determine the elevation of the snow line. Instead, we estimated the SLA using the lowest 10% percentile of snow-covered pixels, which is a similar approach to that used by Loibl et al. (2025).

- 80-81: Does this mean that if >3 images are present, the estimated SLA is always retained for that particular year & glacier?

Yes, the end-of-summer snow line altitude (SLA) was estimated in all cases where at least three valid satellite images were available between 15 August and 30 September. However, unfortunately, this does not guarantee that one of the images actually covers the highest position of the snow line, which could partially explain the differences observed between the modelled ELA and the estimated end-of-summer SLA, as mentioned in Section 3.2.

- 87: Maybe define mean absolute error (MAE) here and use MAE everywhere else.

Done.

- Fig 1:

- It is quite difficult to see any correlation. I would be tempted to discard the glacier-area encoding to reduce overlap (or try another visualization). Otherwise, I am not sure the climatic gradients mentioned in the text can actually be observed. On top of this, I see in Fig. S1 that there is no significant correlation between area and ELA.

Thank you for this comment. The choice to include glacier area was intended to highlight large glaciers and not to show a correlation with ELA. Additionally, it helps to provide an overview of all glaciers. Using a constant marker size leads to a trade-off between small markers, where the color is not clearly visible, and larger markers that strongly overlap. We agree that potential climatic gradients can be visually distorted by the varying marker sizes. We therefore reduced the scaling factor between glacier area and marker size to limit distortions and improve the visibility of the gradients.

- I would move the text labels so they do not overlap with the points.

Done.

- Fig. 2:

- I am not sure what "Mean error" means here, but I would report the MAE before and after bias correction. From the text, I understand that "Mean error" = MAE.

Thank you for pointing out the imprecise labeling. We replaced "mean error" with MAE and MAE\* (bias-corrected MAE) and added a clarification in the figure caption.

- Is the comparison between the modeled  $dh/dt$  and the observed one telling us anything critical, since the observations themselves were used to calibrate the model? Perhaps this is better suited for the supplement, and the space could be used for something more interesting. For instance, it would be interesting to see a comparison between the calculated SLAs and GLAMOS ELAs, or another type of validation for the SLAs, to help interpret the metrics from panel (b).

We understand this point and appreciate the suggested changes. The correlation in panel a) is expected, as it represents the calibration target. The actual calibration target is spatially distributed, and this plot illustrates how well the calibration reproduces it. While good agreement is expected, it is still important to verify that the calibration performs as intended. We observe a strong, though not perfect, correlation, which already provides a first indication of model performance. The intended interpretation of this figure is therefore that calibrating against elevation change (a) yields an estimate of the ELA (b). We also understand the suggestion to compare SLA and GLAMOS. The comparison to glaciological measurements and other analysis of the SLA product has been documented in the preprint of Sommer et al. (2026), which we are now able to reference.

- Maybe add the sample sizes in both panels.

Thank you for this suggestion. We added the sample size (409) in the figure caption.

- I would force integer tick labels for panel (a).

We are not entirely sure what you mean by this.

- The title of panel (b) is confusing.

We agree. We change the title to "ELA vs "EoS SLA"

- Fig 3:

- What do the ellipses represent?

Thank you for this comment. In the first three panels, the ellipses represent the ensemble spread in the y-direction, which can be interpreted as the calibration uncertainty, and the annual variability of the GLAMOS dataset in the x-direction. In the remaining panels, the ellipses indicate either uncertainty (d,e) or the standard deviation (f,g,h) of the displayed mean.

We agree that this representation is not fully consistent and confusing, as the ellipses reflect different quantities (uncertainty, annual variability, standard deviation). Together with the addition of more WGMS data points we therefore decided to omit the ellipses from the revised figure to allow better visual inspection of individual glaciers.

- I see that Aletsch is an outlier also in (g) & (h). Any idea why? The ice thickness looks better (this is the one I expected to be an outlier following the discussion from lines 128–130).

Thank you for your interest. We added a figure in the supplement showing the results of the ice thickness inversion for Aletsch. The figure reveals spatial inconsistencies, with thicker ice in regions where GPR measurements are available and systematically thinner ice in between. Since panel (f) only includes locations with direct measurements, this issue is not visible there. However, the overall tendency towards thinner ice in unobserved regions affects the modeled ice flux and is reflected in the velocity fields shown in panels (g) and (h). This behavior is a known limitation of the current IGM inversion setup. In principle, it can be mitigated by tuning inversion parameters, but future IGM versions are expected to improve the inversion of the ice thickness and match the observed velocity field.

#### Citations

Schmitt, P., Maussion, F., Goldberg, D.N., Gregor, P., 2026. AGILE v0.1: The Open Global Glacier Data Assimilation Framework. *Geoscientific Model Development* 19, 1301–1319. <https://doi.org/10.5194/gmd-19-1301-2026>

Sommer, C., Groos, A. R., Fürst, J., and Braun, M.: Transient snow line altitudes of glaciers in the European Alps from multi-mission remote sensing data (2000–2025), *Earth Syst. Sci. Data Discuss.* [preprint], <https://doi.org/10.5194/essd-2026-35>, in review, 2026.