Reply on RC1

The manuscript 'Inferring Glacier Equilibrium Line Altitudes in Central Europe with FROST' by Herrmann and co-authors presents an application of the newly developed IGM assimilation framework FROST to all glaciers in the European Alps. The framework assimilates surface velocity, ice thickness and elevation change data to calibrate a simple mass balance model, returning ELAs and altitudinal gradients of ablation and accumulation. The authors validate these results against satellite derived end-of-season snowlines and in-situ measurements, showing generally good agreement.

I enjoyed reading this paper and seeing a concrete regional application of the FROST framework. It shows promising results and definitely represents a new stepping stone for data assimilation in glacier models. There were several instances where I would have liked more details on the method and its underlying limitations although I also appreciate that the short format makes it difficult to expand on these – potentially merging the result/discussion or discussion/conclusion may help remove repetitions and gain space to expand on some of the discussion of the limitations? Beyond this, I still had a number of concerns that I believe should be considered.

Thank you for your review of our manuscript, for the thorough engagement with our work, and for the constructive feedback and valuable questions. We have revised the structure by moving the interpretation and discussion elements from the Results section into the Discussion section to improve clarity and reduce redundancy. The remaining concerns are addressed individually below.

General Comments

Description of the model and its limitations: There are a number of limitations/assumptions to the approach/model used that are barely mentioned and that end up being listed in the discussion but with very little explanation of their effects on the results. A few examples:

- The temporal mismatch between the Millan et al. (2022) and the RGI outlines, and the effect of the uncaptured glacier slow-down in the 21st I suspect this would have a strong effect on the initial thickness inversion?

  Thank you for raising this point. We agree that a temporal mismatch between the velocity product of Millan et al. (2022) and the RGI outlines can have a significant impact on the initial thickness inversion, even though the general velocity pattern is not expected to change drastically. We became aware of this issue after submission, as the modelled ice thickness near the glacier front appears too low compared to the observed elevation change. This supports your hypothesis and suggests that the effect is particularly pronounced at the glacier front, where the slowdown may even transition absent ice flow. We added these to the discussion.

Future work will aim to explore alternative velocity products (e.g. Rabatel et al. 2023 or ITS_LIVE) and explicitly address temporal mismatches within a truly transient data assimilation approach.

- The limitations of the simple SMB model are touched upon when it comes to horizontal variability, but less so regarding the non-linearity of the altitudinal mass balance patterns.

  Thank you for pointing out the limitations of the SMB model. We elaborated on the non-linear pattern of mass balance and included this as a limitation of the SMB model in the Discussion.

  At the same time, we would like to emphasize that the main focus of this study is on the calibration framework rather than on the specific choice of the SMB model. The current implementation uses a simplified elevation-dependent parameterization, which is consistent with many regional-scale applications but does not capture the full non-linearity of altitudinal mass balance patterns. Such simplifications are common in large-scale glacier modeling, where computational efficiency and data availability often require reduced-complexity representations .

  In parallel work, we have already implemented a temperature-index model and plan to further extend the framework towards more physically based approaches, including energy-balance models. We therefore acknowledge the simplicity of the current SMB representation and agree that this aspect was not sufficiently elaborated in the original discussion.

Furthermore, I find the half-results, half-discussion paragraph (L114-123) not well structured and hard to follow.

Thank you for the comment. We moved the interpretation of the differences between calibrated values and those derived from glaciological measurements to the Discussion section. The paragraph now focuses purely on describing the results, with the aim of improving readability and overall flow.

Manual tuning of the IGM optimization parameters: this part is very vague but actually quite crucial for the application of the method and potentially strongly influences the results. More details are needed on how this was actually done and based on what criteria. It's also unclear whether GlaThiDa was used in the optimization or just the validation. Presumably the wealth of thickness data in the Alps (or at least in Switzerland, cf. Grab et al. (2021)) would allow for a cross-validation experiment.

Thank you for this important comment. We agree that the description of the IGM inversion setup was too vague, although it is a crucial component of the workflow. We have revised the manuscript to provide more details.

We started from the default parameter settings provided in the IGM repository. These settings are themselves the result of prior manual tuning and served as a baseline. A first qualitative assessment of the inversion performance was based on visual inspection (e.g. Figure S4), which includes comparisons of observed and modelled surface velocities using summary statistics such as the mean, quantiles, and extremes of the velocity distribution.

We then tested several alternative parameter sets. However, none of the tested configurations provided consistently satisfactory results across all glaciers. Given the computational cost of inverting 409 glaciers, we restricted the analysis to a small number of parameter sets and selected the one that yielded the best overall agreement with observed surface velocities. The final parameter configuration used in this study is documented in the GitHub repository.

We note that limitations of the inversion are a known issue of the IGM v3.0.0 and are actively addressed by the development team. During the course of this work, the model underwent continuous development, and we evaluated multiple updates. However, due to time constraints, we fixed the model to a version to 3.0.0, acknowledging that the inversion performance is not optimal in all cases.

We agree that a more systematic evaluation, for example using cross-validation with independent thickness data, would be highly valuable. However, a comprehensive hyperparameter search for the current version of IGM is unlikely to yield substantially improved results.

We clarify that GlaThiDa observations were used in the inversion where available. This information was missing in the original manuscript and has now been added to the Methods and Results sections.

Limitation to GLAMOS: While GLAMOS is an exemplary dataset, I find it disappointing that measurements from Austria, Italy or France were not included.

We fully agree with this point and have extended the evaluation to additional countries. The WGMS dataset provides 9 additional glaciers from Austria and Italy, that have observations in at least 10 years between 2000 and 2019 and larger than 1km2 in the RGI7 outline. Unfortunately, no elevation-dependent mass balance data are available in WGMS for glaciers in France.

Snow line mapping: I find the definition of snowlines using a simple Otsu threshold overly simplistic. One simple Otsu threshold is likely to get the snowline wrong in the presence of debris. Even without debris, if the snowline goes beyond the firn transition, how can one be sure that the depicted snowline doesn't correspond to the ice-firn transition, while the snowline would be higher? This makes me wonder if that could not explain the ELA bias obtained from the EoS SLAs. There are a number of significantly more robust approaches to map snowlines out there (Loibl et al., 2025; Roussel et al., 2025; Kim et al., 2025 to name but a few recent ones), including some actual EoS SLA datasets for the Alps if I recall. I therefore wonder why go with such a simplistic approach.

We agree with the reviewer that selecting an unconstrained Otsu threshold from all reflectance values across a glacier may be affected by various factors, such as the presence of debris or cloud cover, or an indistinct transition between bare ice, firn and snow. Instead, we apply a multi-stage filtering approach to Landsat and Sentinel-2 time series images, accounting for rocks, debris, clouds, terrain shadows and sensor saturation over bright accumulation areas. This forces the reflectance of the glacier surface into a bimodal distribution by excluding pixels that are most likely neither snow nor ice. This results in much more robust automated thresholding of the glacier surface. The full methodology is described in the recently published preprint Sommer et al. (2026), alongside a number of sensitivity studies, as well as comparisons with contemporaneous manually drawn snowlines and published ELAs across the Alps.

Figure A2 in Sommer et al. (2026) shows the ELA values available from WGMS compared to the end-of-summer (EoS) snowlines, as estimated by us. This comparison reveals similar discrepancies to those found here, with the estimated EoS snowlines sometimes being lower than the reported ELA values. One possible reason for this is the absence of a cloud-free snowline measurement representing the highest snowline altitude at the end of the ablation period. Another potential issue with this approach, as mentioned by the reviewer, is the classification of firn as snow cover, given that their reflectance is relatively similar. This can result in the snow line elevation being underestimated in cases where the glacier ice is completely exposed (discussion section of Sommer et al. (2026)). On the other hand, this method automatically accounts for differences between sensors (e.g. different spectral bandwidths and other technical characteristics), differences in image illumination and sun angles, and the varying characteristics of snow surfaces (e.g. fresh vs. perennial snow). These factors can present challenges for estimates based on an empirically selected constant threshold (or set of thresholds), as employed in the studies of Loibl et al. (2025) or Roussel et al. (2025). As far as we are aware, there is currently no snowline dataset available that covers a large number of glaciers and an extended period of years in the Alps. Unfortunately, the snowline data by Rastner et al. (2019) is no longer available, as the first author of this study personally informed us.

We therefore apologise for the relatively imprecise description of the snow line measurement. 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/cr-610)."

Alternative metrics of glacier state: There is a focus on ELA in this study, but I believe that this is sometimes not the most telling metric when it comes to characterizing glacier health. For this I would suggest using the Accumulation Area Ratio, which may lead to stronger correlations in Supplementary figure 1.

We appreciate this suggestion and agree that the accumulation area ratio (AAR) can provide a more informative measure of glacier state than ELA alone, as it accounts for the glacier-specific elevation range. However, the primary aim of this study is not to characterize glacier health, but to demonstrate the calibration framework. In this context, ELA serves as a direct and interpretable parameter of the elevation-dependent SMB model.

Incorporating AAR into the evaluation would mainly introduce a transformation of both the calibrated results and the evaluation datasets. For this reason, we focus on ELA as the primary metric, while acknowledging that AAR could be a useful complementary measure in future analyses.

Line-by-line comments

L1: I find this first two sentences strange and not very convincing as a justification.

Thank you for pointing out this weak justification. We changed the sentences in the revised manuscript.
“The current pace of glacier retreat in the European Alps is unprecedented in the observational record and has significant implications for water resources and downstream ecosystems. Quantifying the future evolution of these systems requires physically based glacier models that are calibrated against observational data.”

L5: the Instructed Glacier Model

Done

L6: ‘correlations’ is vague, be more specific about the actual metric used

Done

L7: The term ‘glacier equilibrium conditions’ is not standard and slightly confusing. Maybe rather mention how this information could be used?

We agree and changed the wording to

“… demonstrating that FROST enables satellite-based calibration of SMB parameters.”

L12: or have already disappeared…

Done

L13-14: ‘these glaciers are vital … for hazard management’? Doesn’t really make sense (one would argue that there’d be no glacier-hazard management without glaciers), should be reformulated

Thank you for pointing out this incorrect formulation. We removed the misleading statement and instead clarified that it is the consequences of glacier retreat that are relevant for hazard assessment and risk reduction at the end of this paragraph.

L16: Not convinced that the Birchgletscher collapse is a result of glacier retreat.

We agree that the retreat of Birchgletscher was not the primary cause of the collapse and have therefore removed the reference.

L17: Syntax error with the sentence split in 2

Corrected.

L20-23: it is apparent here that you’re referring to regional-scale models which up to now have basically only used Hugonnet for calibration (although this is changing – see Cremona et al. (2025) with snowlines). The reference to these types of models should be made clearer, because at the smaller glacier/catchment scale, Hugonnet is only one of the many options to calibrate an energy-balance model.

Thank you for pointing out the missing distinction between local and regional model approaches. We have revised the paragraph accordingly to clarify the differences in available datasets.

L23: reduces -> reduce

Corrected.

L24: more correct to specify that these are spatially distributed datasets

Done.

L30: I see what is meant but have to disagree here. IGM does invert for a so-called ‘sliding’ term, but this is just a black box with many hidden processes. In no way does it say anything realistic about the actual internal glacier structure/basal conditions.

We agree that the original formulation was misleading. We have revised the sentence in the manuscript to:

“IGM also enables the inversion for ice thickness and an additional sliding coefficient field that represents unresolved processes influencing surface velocity that are not explained by ice thickness or bed slope alone.”

L36: Space to be removed after ‘model internals’

Corrected.

L41: ‘building on… Hermann et al. (2025), we apply it…’ Something wrong with the syntax here.

Corrected.

L42: So here you refer to 2000 as the year of RGI 6.0 for the European Alps, but this is an approximation, right? Would be better to specify the year range.

Thank you for pointing this out. We clarified that the glacier selection is based on the RGI7 outlines, which are intended to represent conditions as close as possible to the year 2000 (although in fact its range of dates can still be substantial in some regions).

L57: I can’t help wondering why is a Kalman filter actually needed here. There are only 2 observations, so a more basic Monte Carlo approach would do just as well?

Thank you for this comment. We agree that for the specific task of estimating SMB parameters from elevation-change observations over a single period, approaches such as Monte Carlo methods would be sufficient and more efficient.

The SMB calibration performed with FROST should be interpreted as a simplified test scenario for the broader concept of data assimilation in glacier modeling. The primary strength of the Ensemble Kalman Filter lies in its ability to assimilate observations sequentially through time, allowing the model state to evolve jointly as new data become available. This becomes particularly advantageous in settings with high frequency of observations and higher-dimensional state spaces.

The present setup represents a reduced case to demonstrate the framework, which forms the basis of more complex and truly transient data assimilation scenarios.

L57: What was the range of parameters used?

We use the same parameter ranges as in our previous publication (Hermann et al. ,2025, Figure 6). The prior distributions are derived from the GLAMOS dataset by computing the mean and standard deviation. The initial ensemble spread is set to three times the standard deviation. The resulting prior parameters are: ELA = 3144 ± 498 m, ablation gradient = 7.55 ± 8.04, and accumulation gradient = 3.09 ± 3.39.We agree that this is relevant information in this publication and added the initial distribution in the revised manuscript.

L61: specify that you're using RGI 7. This makes me wonder how you recalculated the geodetic mass balance with these outlines since Hugonnet et al. (2021) use RGI 6.

Thank you for pointing this out. We now explicitly state that we use RGI7.

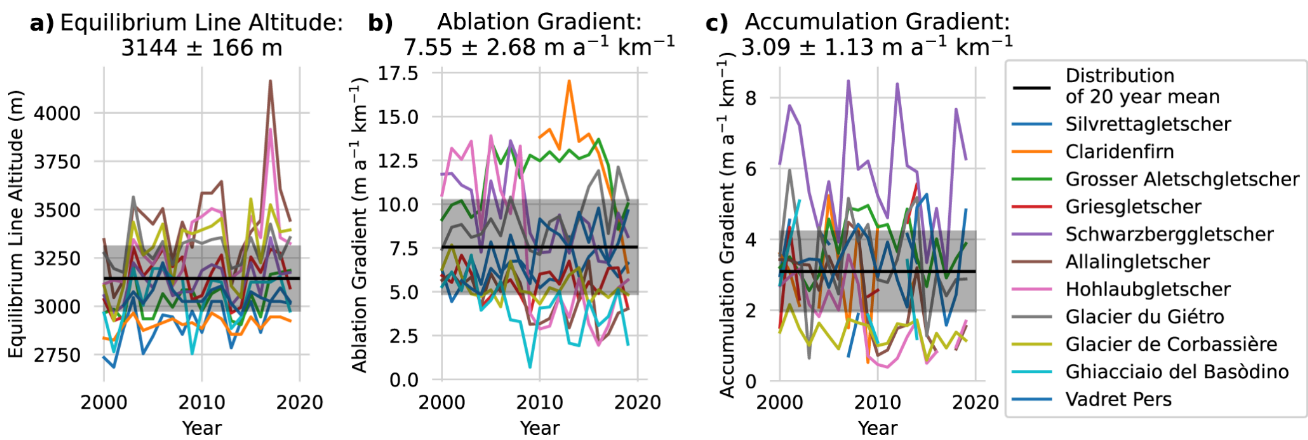
We would like to clarify that we do not use the geodetic mass balance from Hugonnet et al. (2021), but rather the elevation change fields. These are provided as spatially continuous datasets and are not constrained to glacier outlines. During the preprocessing within FROST, the elevation change maps are first cropped to the glacier area (maximum glacier extent plus a varying margin), and subsequently masked using the RGI7 outlines.

L67: In principle yes although I see some strong limitations there, the main one being the relative simplicity of the mass balance model. One particular issue comes with the presence of debris, which strongly affects the mass balance gradient in the ablation zone. Even in the Alps there are extensively debris-covered glaciers (Unteraar, Zmutt, Glacier Noir…) and it would have been interesting to know how these affect the ELA patterns.

Thank you for this comment. We agree that the original statement was too broad and have revised the text to reflect the limitations of the SMB model, in particular regarding debris cover. We note that the FROST framework is not tied to this simple SMB model and can be extended to more complex formulations, including debris-cover of glaciers. The effect of debris on the calibrated ELA is an interesting question; however, it is likely smaller than the current uncertainty in the calibration and therefore difficult to assess within this study.

L71: 5 years sounds small, especially if these are the most recent years, overall much warmer, making their representativity questionable.

We agree that a 5-year period may appear short and potentially affected by recent warming. However, the GLAMOS data show that the interannual variability of ELA for individual glaciers is relatively high, while the trend over the 20-year period is comparatively small. This can be seen in a figure from our previous publication (Hermann et al. ,2025 , Figure 6):



However we agree that the small sample size introduces a sampling error, which we now analysed in Figure S9. The mean sampling error of 5 years is around ~50m with outliers reaching 200m difference. Increasing the number of years required for our evaluation reduces the sampling error but also reduces the number of glaciers for the evaluation. With this new information we increased the threshold to 10 years. Reducing the sample error to mean ~ 25m and extremes of 100m difference. The new threshold 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 between 2000-2019.

L73: I don't think that the acronym EoS has been defined

Done

L76: (Otsu, 1979)

Done

L80-81: Not sure what is meant here

We agree that the original sentence was unclear and have rephrased it for clarity. Estimates based on fewer than three suitable satellite scenes were subjected to outlier detection, as a minimum of three scenes is required to ensure sufficient temporal coverage of the late ablation period. If more than three scenes are available the likelihood increases to capture the true maximum retreat of the snowline.

L88: Pearson's correlation is standard and I do not see the need to include the formula here.

Done

L96: 'of from' – one of these can be removed

Done

L100: The other figures of the SI are not referenced in the main text.

Done

Figure 1: Indicate what the numbers in parenthesis mean

Done

L115: extent

We meant extant as a synonym for existence.

L114: I understand what is meant, but the sentence is unclear and would deserve to be expanded upon.

Thanks for the comment we moved this part into the Discussion and elaborated on the effect of a small accumulation area on the total mass balance and gradient calibration.

L119-120: I do not understand where this affects the results.

We agree that the explanation in the original text was too brief and did not clearly describe the effect on the evaluation. What we intended to highlight is that the SMB is calibrated against elevation change data, while the evaluation is performed using GLAMOS. If differences exist between the calibration target and the evaluation datasets at the glacier-wide scale, these can contribute to mismatches between the calibrated SMB parameters and those derived from the evaluation data. To assess the consistency between these datasets, we compare them in Fig.~3e. In this comparison, one dataset represents volume change, while the other represents mass change. For the correlation analysis, we chose not to convert volume change to mass change.

L123: I disagree, ice thicknesses are quite extensive in the Alps.

We agree that, compared to other regions, the Alps likely have the highest ratio of ice thickness measurements relative to glacier area. However, our point is that these observations are typically limited to profiles rather than being fully spatially distributed, as is the case for elevation or surface velocity fields. While unobserved areas can often be reasonably interpolated by glacier models, this version of the IGM inversion appears to have a weakness, where modeled ice thickness can deviate from plausible values in the absence of direct observations (see Figure A1, A4).

Figure 2: Specify ‘glacier-wide’ mean elevation change, and ‘EoS’ Snow Line Altitude

Done.

L127-129: This should be explained in detail in the methods – how was this done exactly? Was this validated?

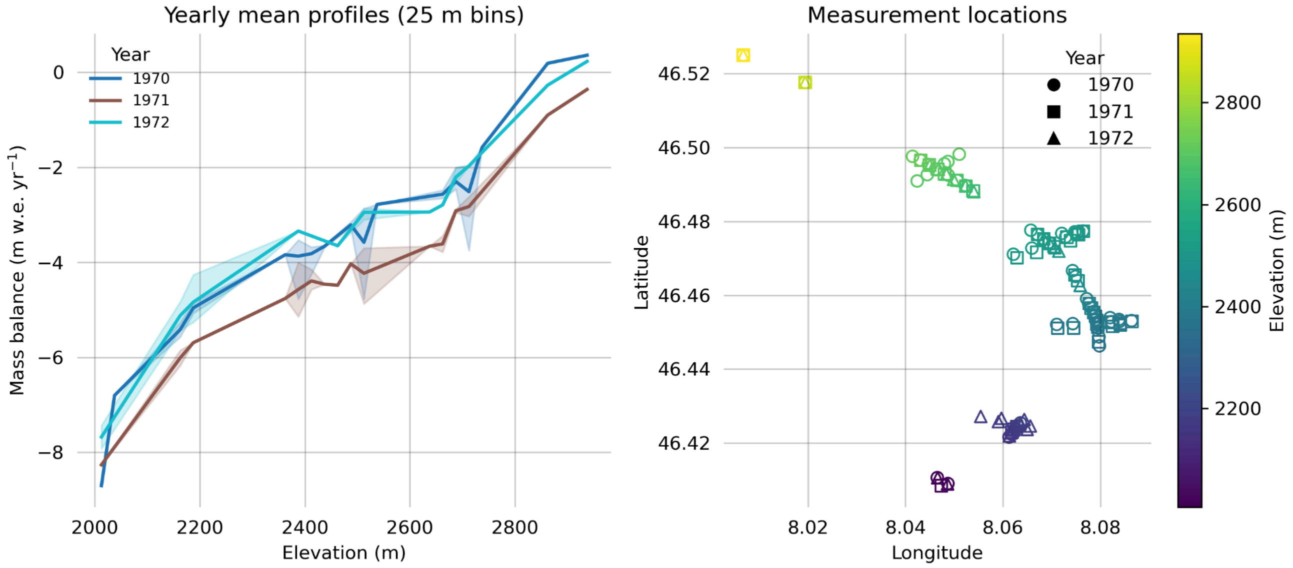
We agree that this is a weak point of the method, as there is no systematic procedure for selecting the inversion parameters and the choice is based on visual inspection (e.g. Fig. S4, including mean, quantiles, and extremes of the velocity distribution). As noted in our response above, we have now added a more detailed description of this procedure to the Methods section.

L136: Unclear to me what is meant by 'regional biases'. Is this a reference to the product uncertainties?

Thank you for pointing out this incorrect wording. We meant local biases in the accumulation area due to low contrast and steep areas where small horizontal offsets lead to high vertical offsets.

L143: More details would be welcome here. Has this effect been investigated to explain the poor results obtained for Aletsch?

Thank you for this question. We investigated the available point measurements from the WGMS dataset for Aletsch. Unfortunately, there are no measurements covering different branches of the glacier. At lower elevations (~2500 m), multiple transects are available for the years 1970–1972, showing a variability of up to ~2 m at the same elevation, though typically lower. While this dataset is limited, it provides some indication of horizontal variability in SMB. Differences between branches with varying aspects are likely larger, but cannot be quantified with the WGMS data. The poor results for Aletsch are unlikely to be explained by this effect alone and are more likely primarily related to poor thickness inversion.



L151: What do you have in mind more specifically? This is the first mention of transient data assimilation, why is this coming up only here? More generally, it's frustrating to see the application of this mass balance model without any explanations on how this can be used for future model improvement. This is barely mentioned at the end of the conclusion.

Thank you for the comment. We have extended the motivation and context of this study in the Introduction to more clearly present the Ensemble Kalman Filter as a method for transient data assimilation. In this work, we consider a simplified test setup using only a single observation period. We have also expanded the Conclusions to better outline potential future applications, including the extension toward fully transient data assimilation.

L163: I would have been interested to see more discussions on these velocity artefacts, especially where they come from.

We agree that it would be valuable to better understand the origin of these artefacts, but we can only provide speculations. Low surface contrast in snow-covered areas can reduce the performance of feature tracking, while cloud cover may further degrade velocity estimates. Small glaciers are likely more affected than large ones, as they provide fewer trackable features for reliable velocity retrieval. We have included four additional figures in the Supplement (Fig. S6, S7) to illustrate these artefacts and their impact on the inferred ice thickness and SMB fields.

L175: Missing references to the velocity & ice thickness datasets

Thank you for pointing out the missing reference. We added the reference for velocity and ice thickness in the data availability section.

References:

Gardner, A. S., Greene, C. A., Kennedy, J. H., Fahnestock, M. A., Liukis, M., López, L. A., Lei, Y., Scambos, T. A., and Dehecq, A.: ITS_LIVE global glacier velocity data in near-real time, The Cryosphere, 19, 3517–3533, https://doi.org/10.5194/tc-19-3517-2025, 2025.

Loibl, D., Richter, N., and Grünberg, I.: Remote sensing-derived time series of transient glacier snowline altitudes for High Mountain Asia, 1985–2021, Scientific Data, 12, 103, https://www.nature.com/articles/s41597-024-04309-6, 2025.

Rabatel, A., Ducasse, E., Millan, R., and Mouginot, J.: Satellite-Derived Annual Glacier Surface Flow Velocity Products for the European Alps, 2015–2021, Data, 8, 66, https://doi.org/10.3390/data8040066, 2023.

Rastner, P.; Prinz, R.; Notarnicola, C.; Nicholson, L.; Sailer, R.; Schwaizer, G.; Paul, F. On the Automated Mapping of Snow Cover on Glaciers and Calculation of Snow Line Altitudes from Multi-Temporal Landsat Data. *Remote Sens.* 2019, *11*, 1410. https://doi.org/10.3390/rs11121410

Roussel, L., Dumont, M., Réveillet, M., Six, D., Kneib, M., Nabat, P., Fourteau, K., Monteiro, D., Gascoin, S., Thibert, E., Rabatel, A., Sicart, J.-E., Bonnefoy, M., Piard, L., Laarman, O., Jourdain, B., Fructus, M., Vernay, M., and Lafaysse, M.: Saharan dust impacts on the surface mass balance of Argentière Glacier (French Alps), The Cryosphere, 19, 5201–5230, https://doi.org/10.5194/tc-19-5201-2025, 2025.

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.

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^2 choice.

Thank you for this comment. We added a brief justification for the 1 km^2 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: Ratio = NIR * (NIR/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.