Reviewer 1

Note: Reviewers' comments are shown in blue, **our responses** in bold black, and *revised manuscript text quoted in the responses* in black italics.

Summary

The authors describe a new modelling framework to quantify mass balance and frontal ablation with a join calibration method of parameters. The model is applied to 71 glaciers in Svalbard, and the model output is compared to decadal observational data. The results show good accordance with observations with frontal ablation being slightly overestimated. I commence the authors on their thorough description of the methodology, an overall well written paper, and the development of a very useful tool for the community. However, I do have one major and some minor comments that should be addressed.

Response: We thank the reviewer for the positive assessment and constructive comments. We have revised the manuscript accordingly, including substantially revising the discussion of surge-type glaciers to address the major comment. Detailed responses to each comment are provided below.

Major Comments

In the limitations you state that the majority of the glaciers in your study are surge- type glaciers, yet your model is unable to capture surge-type behavior. This makes me question why surge-type glaciers are included in the study and how this inclusion impacts the results on a regional scale. I suggest to either exclude surge- type glaciers from the study and only use the remaining 35 glaciers or to differentiate surge-type/non-surge-type glaciers in the results.

Response: We agree that differences in model performance between glaciers classified as surge-type in RGIv6 and those with no surge evidence warrant examination. However, whether these differences arise from active surge processes, and whether and how explicitly representing those processes would improve model performance, remain open questions. This model-description study presents the integration of the SERMeQ frontal-ablation formulation into the PyGEM–OGGM framework, together with joint Bayesian calibration, and demonstrates the coupled framework's application using Svalbard as a test region. Incorporating explicit surge dynamics and reproducing surge events or cycles lie beyond the scope of this study.

Nevertheless, we consider it important to retain surge-classified glaciers because they constitute a substantial part of the selected sample (36 of 71 glaciers). Their inclusion allows us to assess the framework's performance and limitations across the glacier types encountered in this regional application. Such an assessment is relevant to the development of glacier models intended for regional and global applications.

Following the reviewers' suggestions, we have identified surge-classified glaciers in Table S1 and reported the convergence proportions for both groups in Section 6.3: 30/36 surge-classified glaciers (83.3%) and 32/35 glaciers with no surge evidence (91.4%) converged within 10 iterations under the prescribed hyperparameter configuration. We have also added a comparison of observation–model mismatches by surge classification in Section 4.2 and revised Section 6.3 to discuss these findings and clarify the model limitations. In particular, retaining surge-classified glaciers does not imply that surge dynamics are adequately represented; the effects of omitting

surge processes on regional projections remain unquantified.

The following text has been added to Section 4.2 (**Lines 273-276**):

"Additionally, glaciers classified as surge-type in RGIv6 (RGI Consortium, 2017) were more frequently represented near the top of the mismatch rankings in several comparisons, although the patterns varied among variables and periods. Potential contributors to the mismatch are discussed in Section 6.1, and the implications of surge classification are considered in Section 6.3."

The revised Section 6.3 reads as follows (**Lines 367-381**):

"A high proportion of Svalbard's marine-terminating glaciers are classified as surge-type in RGIv6 (RGI Consortium, 2017). Glacier surges are typically viewed as internal instabilities (Meier and Post, 1969), characterized by cyclical flow between extended quiescent periods and brief surge events. Thirty-six out of 71 glaciers in this study were classified as surge-type in RGIv6 (RGI Consortium, 2017) and are identified in Table S1. Under the prescribed hyperparameter configuration, 30/36 surge classified glaciers (83.3%) and 32/35 glaciers with no surge evidence (91.4%) converged within 10 iterations (Table S1). Among these converged cases, the mismatches reported for surge-type glaciers in Section 4.2 do not establish that active surging caused the discrepancies.

Our frontal-ablation model uses seasonally to annually averaged stress fields without explicit subglacial hydrology, and the coupled framework does not explicitly represent surge events or cycles. It can nevertheless simulate terminus change and frontal ablation for surge-type glaciers. Surge-related terminus movements may influence calibration through the length-change observations, potentially causing parameter adjustments to compensate for processes not explicitly represented by the framework. However, RGIv6 classifications alone do not provide the event timing and magnitude needed to assess these effects. The results therefore illustrate the framework's application under the selected settings, while the need for additional treatment of surge-classified glaciers remains open. Future projections should be interpreted with the caveat that unrepresented surge cycles may alter terminus positions, and their consequences for regional projections have not been quantified."

Minor Comments

Line 73: Reads slightly odd. Rephrase to "PyGEM simulates the climatic mass balance monthly for every 10 m of elevation…".

Response: Revised as suggested in **Line 73**.

Line 73: Remove space at the end of the sentence.

Response: Revised as suggested in **Line 73**.

Lines 126 – 128: Maybe a workflow graph would be helpful here so that the reader can easily understand the integration of PyGEM and OGGM rather than having to follow this in the text.

Response: we have added a workflow diagram illustrating the coupling between PyGEM, OGGM, and SERMeQ (Figure 1; reproduced below), and cited it in the main text (**Line 130**) to provide a clear visual reference: *(Figure 1)*.

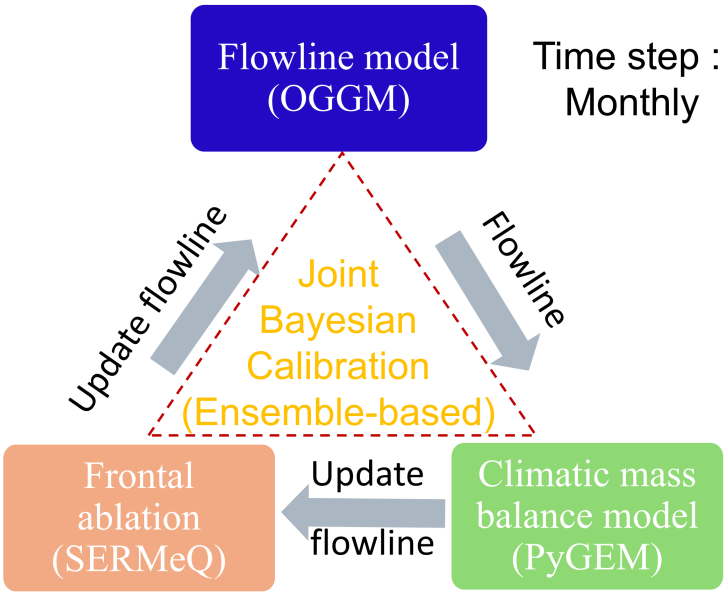


Figure 1. Workflow of the coupled PyGEM–OGGM–SERMeQ model. SERMeQ is called within the OGGM flowline model.

Line 134: Replace "the 4 x1 vector" with "a 4 x 1 vector"

Response: Revised as suggested in **Line 137**.

Line 203: You use ERA5 temperature data – Do you correct for the elevation difference between ERA5 gird cell height and glacier surface height? This might not change temperatures much, but it should be considered.

Response: Yes, the model accounts for this elevation difference using monthly temperature lapse rates. We have clarified this in the revised manuscript (**Lines 207-209**):

"The model was forced with monthly temperature and precipitation from ERA5 (Hersbach et al., 2020), with temperature adjusted for elevation differences between the centroid of each glacier and the closest ERA5 grid cell and then distributed as a function of elevation for each glacier elevation bin using monthly lapse rates (Rounce et al., 2020)"

Line 218: Remove space between parenthesis and 1.5%

Response: Revised as suggested in **Line 224**.

Lines 222 – 224: This paragraph should be integrated in section 4.1. It is directly applicable to section 4.1. but not to section 4.2.

Response: To introduce both subsections and clarify their distinct purposes, we have restructured and revised the opening paragraph as follows (**Lines 228-231**):

"To illustrate the modeling workflow and evaluate hindcast performance, we present two complementary applications in Svalbard. An illustrative hindcast for a single glacier, Sabinebreen, provides methodological clarity and aids interpretation of model outputs and their temporal evolution. A regional application to 71 glaciers provides a quantitative evaluation of hindcast performance."

Line 226: Remove space between Svalbard and parenthesis

Response: Revised as suggested in **Line 233**.

Line 226: If the previous paragraph is integrated into this section, it is not necessary to state that this is an illustrative case and it should therefore be removed.

Response: Following revision of the preceding paragraph, we have removed the repeated phrase "as an illustrative case" and retained the the rationale for selecting Sabinebreen. The sentence now is revised as follows (**Lines 233-234**):

"We selected Sabinebreen ($\sim$ 67 km^2, on Nordaustlandet in northern Svalbard), for the hindcast because it benefits from relatively direct"

Lines 228 – 229: I understand that the authors want to show the best case scenario, but I think it would be good to also show results that are affected by heavy data infilling (at least in the supplementary).

Response: As suggested, we have added a figure showing model outputs for a glacier with heavy data infilling to the Supplement (Figure S1; reproduced below), providing a comparison with the Sabinebreen example. We have also cited this figure in the main text immediately after explaining why Sabinebreen was selected (**Line 236**):

(see Figure S1 for an example with substantial data infilling)

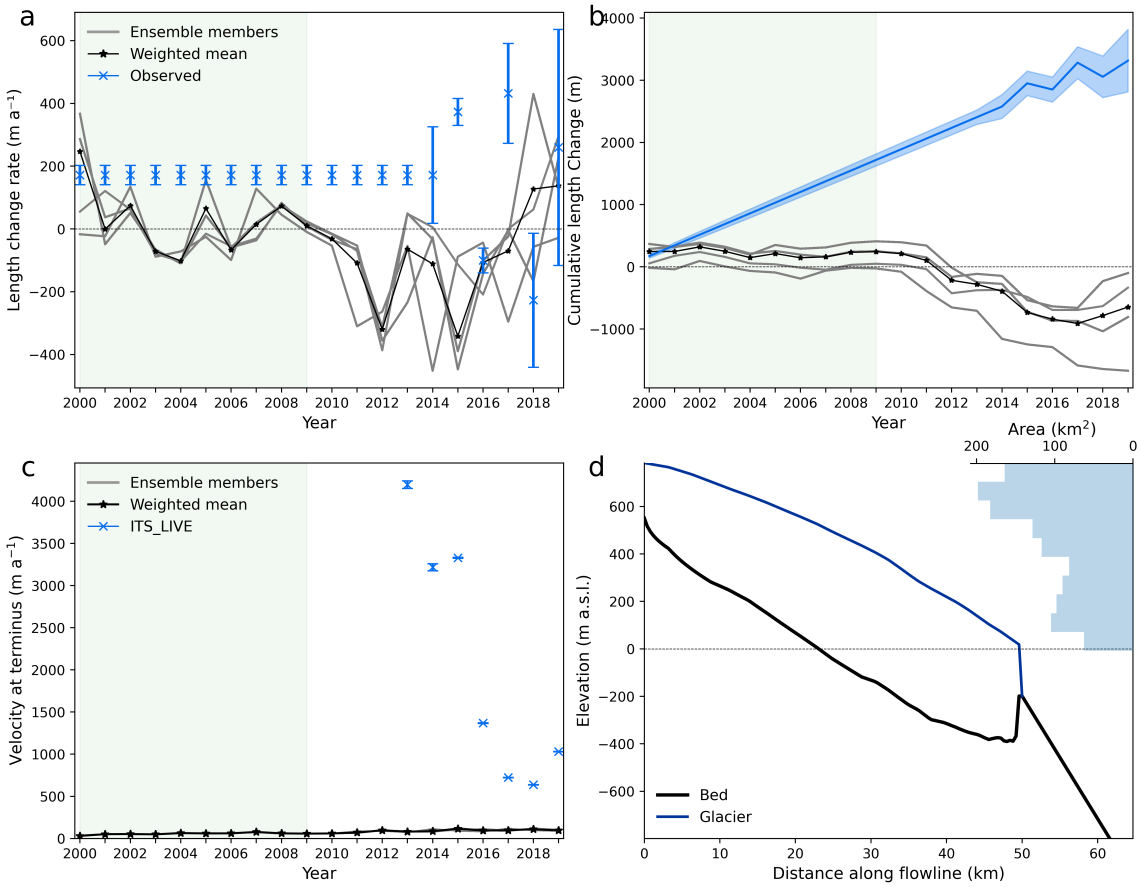


Figure S1. Comparison of modeled results with observations for a glacier requiring heavy data infilling (RGI60-07.00027;calibration did not converge within 10 iterations) from 2000 to 2019 (a) annual length change rate, (b) cumulative length change, and (c) annual terminus surface velocity. Panel (d) shows the initial glacier bed and surface elevation as well as the area-altitude distribution at the year 2000. The observed values in blue include the associated uncertainties. The light green shading refers to the calibration period. The gray lines represent individual members of the Bayesian ensemble.

Figure 1d: It is not quite clear why this panel is relevant as it is not referred to in the text. Either make clear why it is important or remove.

Response: We have added an explicit reference to Figure 2d and clarified the relevance of the geometry shown to hindcast initialisation **Lines 238-239**:

"*Figure 2d shows the glacier geometry in 2000, represented by the elevation-band flowline used to initialise the hindcast.*".

Figure 2: Here would be a good place to compare the results of a single glacier with observational data as is stated in Lines 210–211.

Response: Figure 3 presents monthly time series for the Sabinebreen hindcast to illustrate the simulated temporal variability. Whereas the observations cited in lines 216-217 provide only 10-year mean frontal ablation. Figures 4e and S2e already compare modelled and observed frontal ablation over matching 10-year periods for each of the 71 glaciers individually, including Sabinebreen. We have therefore retained Figure 3's focus on monthly variability to avoid duplicating this comparison.

Line 257: Interesting. Can you explain what drives the per-glacier mismatches? Is there a common factor e.g. glacier geometry that causes the mismatches?

Response: As suggested by both reviewers, we examined whether the largest mismatches occurred for the same glaciers across length change, climatic mass balance, and frontal ablation, and whether these mismatches were associated with glacier geometry and RGIv6 surge classification. This analysis covered the 62 glaciers with converged calibration and considered each variable and period separately. We did not identify a consistent geometry-related pattern or establish a common cause of the mismatches. We now report these exploratory findings in Section 4.2 and discuss potential contributors and limitations in Sections 6.1 and 6.3.

We have added the following clarification in Section 4.2 (**Lines 268-276**):

"Among glaciers with converged calibration, mismatches were ranked separately for each variable and period by the separation between the modeled ensemble range and the observation interval, defined as the observed value plus or minus its reported uncertainty. The glaciers with the largest mismatches were partly shared across length change, climatic mass balance, and frontal ablation, but their rank order differed among variables. An exploratory comparison with glacier geometry attributes from RGIv6 (area, maximum length, mean surface slope, and median elevation) did not reveal a consistent geometry-related pattern across all three variables and both the calibration and validation periods. Additionally, glaciers classified as surge-type in RGIv6 (RGI Consortium, 2017) were more frequently represented near the top of the mismatch rankings in several comparisons, although the patterns varied among variables and periods. Potential contributors to the mismatches are discussed in Section 6.1, and the implications of surge classification are considered in Section 6.3."

We have also expanded the discussion of potential contributors to these mismatches in Section 6.1 (**Lines 335-342**):

"Despite these strengths, limitations remain at the glacier scale. Frontal ablation is systematically overestimated, highlighting shortcomings in how terminus dynamics are represented and constrained. This bias may arise from compounded uncertainties in the initial flowline geometry (e.g., bed, thickness, and width), errors in derived velocities, and the fact that velocity is only indirectly constrained by observations. Frontal ablation observations are also not directly used as calibration targets. In addition, the annual length change records for some glaciers are extensively infilled by interpolation, limiting the amount of directly observed information available to constrain year-to-year variability. Together, these limitations point to the need for more accurate

bed inversions, better constrained velocity estimates, and more complete observational coverage of annual length change to improve glacier-scale model performance."

Lines 257 – 258: I understand why you chose to only show one period (2010-2019), but it would be good to show both in the main manuscript.

Response: We agree that both periods are relevant to assessing model performance. We have retained the validation-period results (2010 - 2019) in the main manuscript to emphasise performance beyond the calibration period (2000 - 2009). The calibration-period results are provided in Figure S2 and Table S2 and explicitly compared with the validation results in the main text, (**lines 261-265**). We have retained this arrangement to keep the main presentation concise while providing both periods for comparison.

Figure 4/ Figure 5: I was a bit surprised to see that total length change is shown for each individual glacier, but frontal ablation is only shown cumulative for all glaciers. I think it would be good to show total length change, mass balance and frontal ablation for all glaciers individually as well as cumulatively.

Response: As suggested, we have added figures showing cumulative climatic mass balance and frontal ablation over 2000-2100 for each of the 67 glaciers in Svalbard as **Figure S3 and Figure S4**, respectively, in the **Supplementary** file (reproduced below). These figures follow the map and boxplot format of Figure 5. We have also added the following description and figure references to the **main** text (**Lines 289-292**):

"For the same glaciers, the median cumulative climatic mass balance over 2000-2100 is generally lower under SSP585 than under SSP126, reflecting smaller gains where the climatic mass balance is positive and greater losses where it is negative (Figure S3). Scenario differences in median cumulative frontal ablation are less uniform, although most glaciers show slightly higher values under SSP585 (Figure S4)."

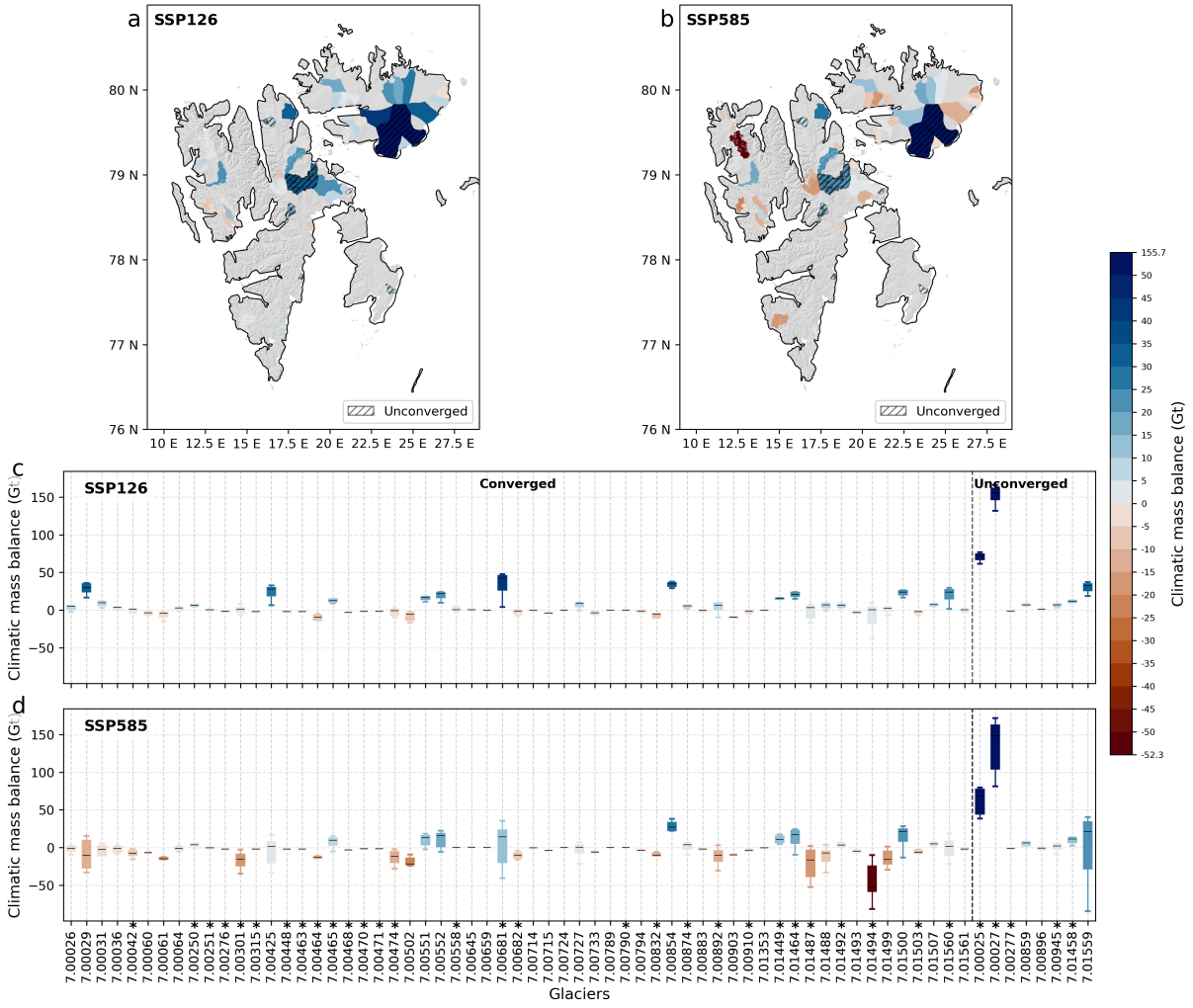


Figure S3. Projected cumulative climatic mass balance over 2000-2100 for each of the 67 glaciers in Svalbard. (**a-b**) Spatial distribution of median cumulative climatic mass balance under SSP126 and SSP585 scenarios, with hashed areas indicating unconverged calibrations. (**c, d**) Distribution of projected cumulative climatic mass balance of each glacier under scenarios SSP126 and SSP585, grouped by glaciers with converged calibrations (left) and unconverged (right). Each boxplot summarizes the projected total length change across the ensemble of 12 GCMs for a given glacier and scenario. Box elements: the center line indicates the median, the box spans the 25th–75th percentiles (interquartile ranges, IQR), and the whiskers extend to 1.5×IQR. Color scale represents magnitude of climatic mass balance, facilitating comparison between spatial and statistical representations. Glaciers are identified by their RGIv6 IDs listed without the common "RGI60-0" prefix, and surge-type glaciers are marked with an asterisk (*).

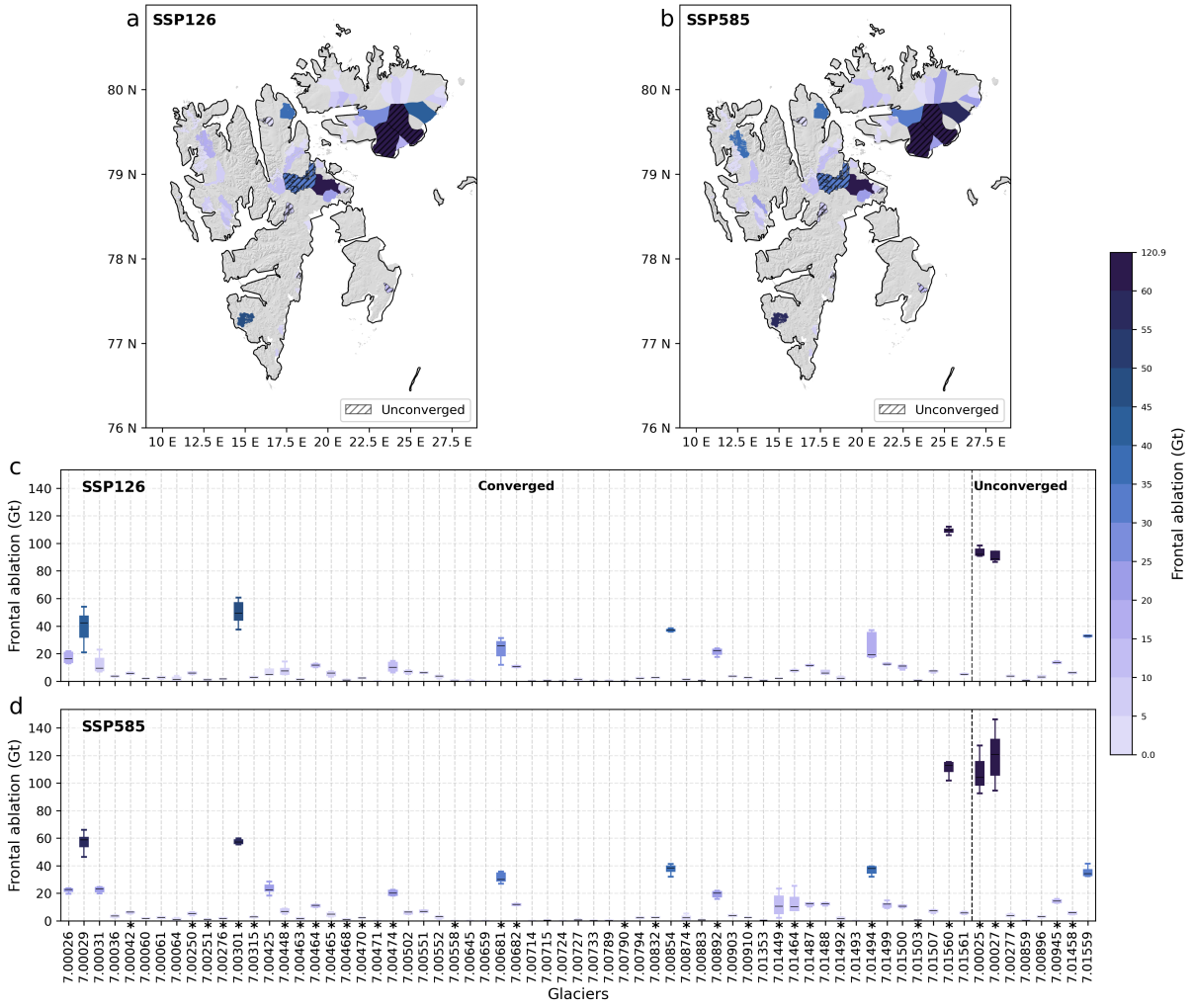


Figure S4. Projected cumulative frontal ablation over 2000-2100 for each of the 67 glaciers in Svalbard. (**a-b**) Spatial distribution of median cumulative frontal ablation under SSP126 and SSP585 scenarios, with hashed areas indicating unconverged calibrations. (**c, d**) Distribution of projected cumulative frontal ablation of each glacier under scenarios SSP126 and SSP585, grouped by glaciers with converged calibrations (left) and unconverged (right). Each boxplot summarizes the projected total length change across the ensemble of 12 GCMs for a given glacier and scenario. Box elements: the center line indicates the median, the box spans the 25th–75th percentiles (interquartile ranges, IQR), and the whiskers extend to 1.5×IQR. Color scale represents magnitude of length change, facilitating comparison between spatial and statistical representations. Glaciers are identified by their RGIv6 IDs listed without the common "RGI60-0" prefix, and surge-type glaciers are marked with an asterisk (*).

Lines 299 – 300: Given that SERMeQ only outputs terminus positions along the flow line, it also means that you do not account for changes in glacier front geometry i.e. if the glacier retreats faster on the lateral boundaries. Could you implement a different methodology in SERMeQ e.g. the curvilinear box method? Do you think this makes a difference in your frontal ablation estimates?

Response: We agree that our coupled framework's one-dimensional flowline representation does not explicitly resolve changes in glacier-front shape, including faster retreat at the lateral margins. Different two-dimensional front configurations can correspond to the same modeled terminus position. The curvilinear box method cannot be directly applied to the outputs of our

one-dimensional flowline framework, which represents terminus position along the flowline rather than resolving the full glacier-front geometry. Its application would require additional geometric information or assumptions beyond the current model representation. Unresolved front-shape changes may affect frontal-ablation estimates at individual glaciers. We have not quantified this effect and cannot establish whether it cancels or produces a systematic bias at the regional scale. This remains a limitation of the one-dimensional representation. We have added the following clarification in the manuscript (**Lines 331-334**):

"However, the one-dimensional framework does not explicitly resolve changes in glacier-front shape, such as faster retreat at the lateral margins. This simplification may affect glacier-scale frontal-ablation estimates, but the magnitude of this effect and its implications for regional projections have not been quantified."

Line 313: Do you exclude these three glaciers in addition to the 9 glaciers that did not converge? Please clarify.

Response: No. The nine glaciers that did not converge within the prescribed 10 iterations were not automatically excluded from future projections. The three glaciers discussed here were excluded because of physical inconsistencies in the projection simulations. Together with one separate exclusion due to persistent ensemble collapse during calibration, four glaciers were excluded from the original 71. The projection analysis therefore includes 67 glaciers: 59 with converged calibration and eight with unconverged calibration. We have clarified this distinction in the section 5 and Section 6.2 and added a reference to Table S1.

The revised projection section reads as follows (**Lines 302-305**):

"The projection results presented in this section and in Figures 5, 6, S3, and S4 are based on 67 of the total 71 glaciers. Four glaciers (Table S1) were excluded before calculating the projection statistics because of the physical instabilities and model limitations discussed in Section 6.2. The retained sample comprises 59 glaciers with converged calibration and 8 with unconverged calibration. All 67 retained glaciers contribute to the regional totals shown in Figure 6."

The corresponding clarification in Section 6.2 reads as follows :

(**Lines 345-347**): *"Three glaciers were excluded from our future projection analysis because of physical inconsistencies in the projection simulations. These exclusions are distinct from the calibration-convergence classification: failure to converge within the prescribed 10 iterations did not automatically exclude a glacier from projections."*

and (**Lines 362-363**): *"Together, these exclusions leave 67 glaciers in the projection analysis: 59 with converged calibration and eight with unconverged calibration, as classified under the prescribed 10-iteration limit (Table S1)."*

Lines 319 – 320: Dashes appear quite long. Replace with normal dash (–).

Response: Revised as suggested in **Line 355-356**.

Line 322: Again, is that glacier removed in addition to the nine glaciers that did not converge? Please clarify.

Response: The nine glaciers that did not converge within the prescribed 10 iterations were not automatically excluded from future projections. The glacier discussed here is one of the four exclusions from the original 71 glaciers. It was excluded because ensemble collapse persisted even when the iteration limit was increased to 100. Please see our response to the comment on Line 313 for clarification of the retained projection sample.

References

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, 2020.

Meier, M. F. and Post, A.: What are glacier surges?, Canadian Journal of Earth Sciences, https://doi.org/10.1139/e69-081, publisher: NRC Research Press Ottawa, Canada, 1969.

RGI Consortium: Randolph Glacier Inventory – A Dataset of Global Glacier Outlines: Version 6.0, Technical Report, Global Land Ice Measurements from Space, 2017.

Rounce, D. R., Hock, R., and Shean, D. E.: Glacier Mass Change in High Mountain Asia Through 2100 Using the Open-Source Python Glacier Evolution Model (PyGEM), Frontiers in Earth Science, 7, https://doi.org/10.3389/feart.2019.00331, 2020.