

Response to review

Manuscript ID: ESSD-2025-304

Title: Calving front positions for Greenland outlet glaciers (2002–2021): a spatially extensive seasonal record and benchmark dataset for algorithm validation

Dear Erik Loebel,

Thank you very much for your recognition and constructive comments. We fully accept all the suggestions and are incorporating them into both the manuscript and the dataset, and your feedback has been extremely helpful in guiding our ongoing revisions. Below is a summary of the main revisions followed by the point-to-point responses:

1. Clarified dataset coverage to include marine-, land-, lake-terminating, and floating-tongue glaciers, incorporated additional digitization to enhance spatial coverage, and added GlacierName and GlacierType fields.
2. Conduct a comprehensive cross-dataset Average Minimum Distance (AMD) assessment by identifying/interpolating at least one overlapping delineation per glacier and comparing it against TermPicks, AutoTerm, CALFIN, and Black & Joughin (2023).
3. Revised overly assertive wording, clarified the relevant limitations, and refined the dataset descriptions to improve clarity and consistency throughout the manuscript.

We hope that these revisions adequately address all your comments.

Best regards,
Xi Lu and all co-authors

General comments

This paper presents a large dataset of manually delineated glacier calving fronts. These glacier fronts are valuable to the glaciology community, facilitating a better understanding of glacier dynamics and serving as constraints for ice dynamic modelling. They are also useful reference data for future machine learning-based delineation efforts. If I am not mistaken, this is currently the third largest manually delineated calving front dataset for Greenland (after the TermPicks repository and the data from Black and Joughin, 2023).

According to the manuscript, this product has (1) complete spatial coverage, (2) consistent and (3) high temporal resolution, as well as (4) homogeneous and (5) high accuracy delineation protocols. After reading the abstract and introduction, I anticipated a product that would meet all these criteria. However, upon reviewing the rest of the paper and examining the actual product, I was left quite disappointed, as only point 1 and, with some limitations, point 3 were actually met. I still believe this is a valuable

product that will benefit the glaciology community. As it stands, however, in its current form, the manuscript does not effectively describe the product, particularly its limitations. Significant revisions are required.

At various points in the paper (e.g. lines 19 and 78), it is emphasised that this data product has complete and consistent temporal sampling. This would be a major advantage, particularly for diversifying reference data for machine learning approaches, as most other manually delineated products have heterogeneous sampling, as they are usually by-products of focused glaciology studies. However, Figure 9 shows that the sampling is inconsistent. At L317, it is also stated that glaciers of higher scientific interest are captured at a higher temporal frequency, which seems to contradict the emphasis on temporal consistency somewhat. I am also not quite convinced by the definition of 'scientifically interesting'. Nioghalvfjærdsbræ only has 15 entries and Humboldt Glacier is not included at all. I don't see this as a major issue. I would just like these statements to be more representative of the actual product. Also, as this dataset is intended to serve as a benchmark, this should be mentioned as a limitation.

Response: We sincerely appreciate the reviewer's positive recognition of the value of our dataset to the glaciology community. Regarding the comments on the five key characteristics listed in the manuscript, we acknowledge that some statements in the original submission were overly positive and did not clearly articulate the limitations of our dataset. The corresponding explanations are provided below, and, in addition to supplementing the dataset where necessary, the revised manuscript presents the relevant limitations with greater accuracy and balance.

(1) Complete spatial coverage:

To support automated calving-front detection, our broader objective was to develop a dataset suitable for large-scale investigations of the seasonal behaviour of Greenland outlet glaciers, including joint analyses with seasonal velocity products derived from Landsat imagery over the past few decades. Accordingly, we used the spatial extent of glaciers covered by Landsat-based seasonal velocity fields as our selection criterion (Fig.1, Rosenau et al. (2015)). Glaciers outside this extent were assumed to exhibit limited frontal variability and were therefore not included. Within this defined spatial domain, our dataset effectively covers the glaciers (~290 compared to the 302 included in Rosenau et al. (2015), with some low-quality cases removed) for which seasonal velocity variations can be reliably detected.

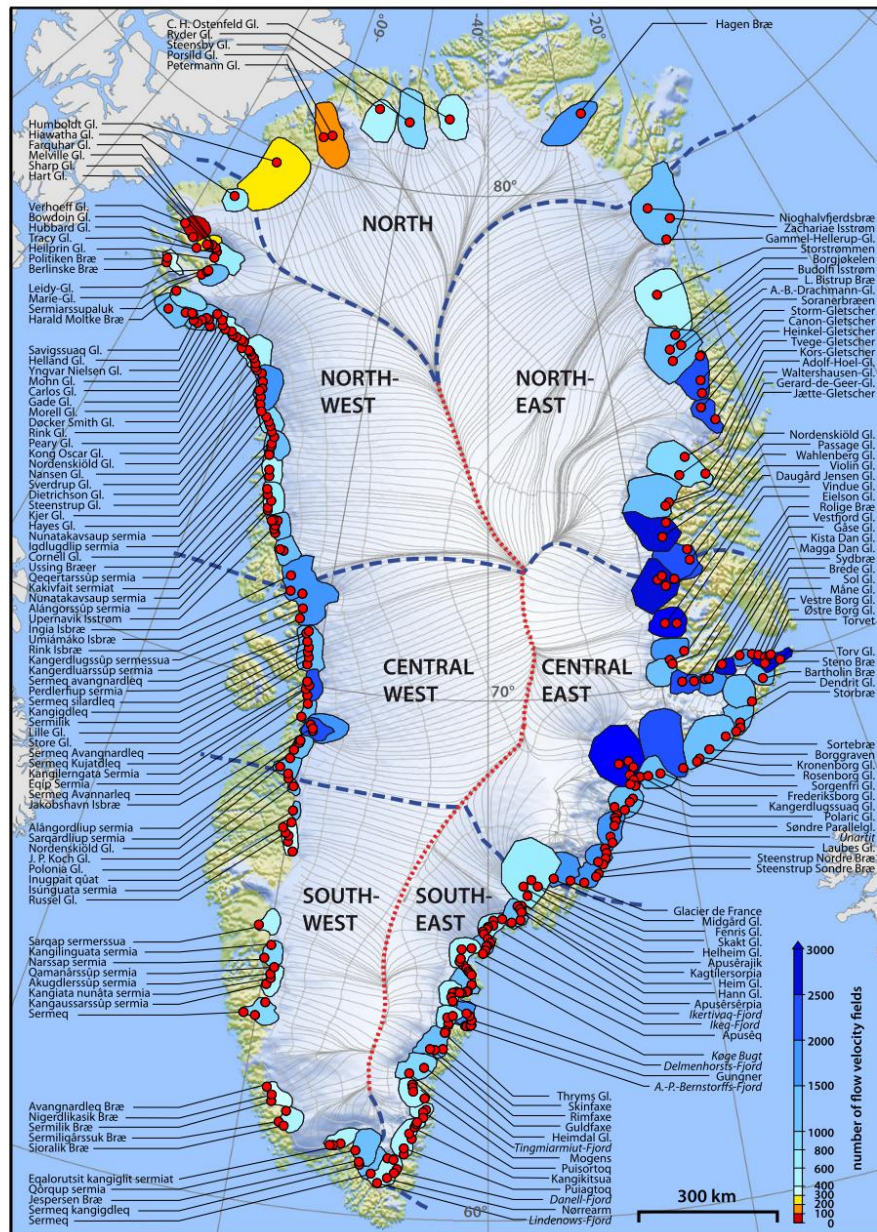


Fig. 9. 302 locations (red circles) and names of all analysed outlet glaciers. The official spelling of the glacier names was chosen according to NunaGIS (2015). Unofficial glacier names are marked italic. The grey flow lines are determined by Gebler (2009). The hill-shading was performed with the elevation data from Ekholm (1996). The ice divide is highlighted with the red dotted line. The dashed blue lines indicate the regional boundaries.

Fig.1 Selection basis of the research area, defined by the spatial extent of outlet glaciers covered by Landsat-derived seasonal velocity fields (Rosenau et al., 2015).

Considering the availability of existing datasets and the suggestions from the other reviewer, we have refined the dataset's definition of coverage to explicitly include marine-terminating, land-terminating, lake-terminating and floating-tongue glaciers, rather than limiting it to those defined by long-term seasonal velocity coverage. We are currently extending the dataset to incorporate additional glaciers of each type, which will be clearly labelled in the metadata to enhance accessibility and usability for the wider cryosphere research community.

(2) Consistency and (3) High temporal resolution:

The dataset was designed to achieve seasonal resolution (at least two scenes per year). For glaciers where GEEDiT imagery did not achieve this, additional images were manually acquired to maintain seasonal completeness. Compared with other published products, this dataset provides more complete seasonal coverage, especially during winter periods, by supplementing with Sentinel-1, ERS-1/2 and Envisat SAR imagery. The term “consistent” here was intended to describe this seasonal-scale fulfilment rather than perfectly uniform temporal sampling. Nevertheless, we agree that the temporal sampling still varies among glaciers due to image availability, and we have clarified this limitation in the revised manuscript.

(4) Consistent and (5) High-accuracy delineation protocols:

These terms were used to summarise the visual inspection and manual delineation procedures, which provide more reliable results than products with contributions from different authors and automated methods, particularly for glaciers with complex fronts. In the initial submission, these characteristics were supported by the six comparison results presented in the manuscript. In the revised version, we will provide additional methodological details and include expanded comparison results to further substantiate these two characteristics.

In addition to supplementing the dataset, we have softened the use of the terms *complete* and *consistent*, replacing them with *extensive spatial coverage* and *seasonally targeted sampling* to more accurately describe the dataset’s characteristics. The revised description now defines the product as a high-precision, long-term seasonal record of calving front positions with broad coverage across diverse glacier types. The scope and limitations associated with each characteristic have been explicitly clarified throughout the manuscript.

Data quality and validation. This is the most critical point, in my opinion. The authors report an accuracy range of 40 to 100 metres for this data product. This value is derived by comparing the calving fronts in 15 satellite images across five glaciers to delineations from other data products. This is supplemented by visual comparisons against Planet imagery for another six calving fronts, as well as visual comparisons of five calving front change time series with other products.

The data quality of this product varies significantly from glacier to glacier (and also from delineation to delineation), but this is not picked up by the validation in the manuscript. It even feels somewhat disingenuous to show how well the glacier front’s teeth-like structure is delineated in Fig. 4 when the quality of most other fronts is significantly worse. For example, many of the calving fronts of Nioghalvfjærdsbræ consist of fewer than 50 vertices for the entire 40 km-long glacier front, with some vertices being more than 5 km apart. These calving fronts clearly exceed 100 m in accuracy and are likely not suited for validating ML models. Similar issues are present for other glaciers. Clearly, validating using just six representative glaciers (with a total of 15 calving fronts assessed computationally) is insufficient to capture the

characteristics of the data product.

My suggestion is the following. Use the data from the larger, existing calving front repositories (I would recommend at least the four larger than this one, TermPick, AutoTerm, CALFIN and Black and Joughin, 2023). Identify pairs of calving fronts where there is an entry in both this and other data sets at the same date. Calculate the average minimum distance error between these two same-day entries. The results could then be shown as a data-product-to-data-product difference overall (e.g. as mean and median), and also per glacier (perhaps in a large table or histogram in the supplementary material). In conjunction with the validation results of the other data products, such an analysis would provide a much more complete picture of data quality.

This is only a recommendation, but I think this paper would benefit from a brief discussion of how it compares with other products. This could particularly cover other manual delineated repositories, such as TermPicks; other benchmark datasets, such as the one from Gourmelon et al. (10.5194/essd-14-4287-2022); and automation products, such as AutoTerm.

Response: Thanks for this important and constructive suggestion. In the initial submission, same-day delineations were difficult to identify for specific glaciers. Therefore, we selected one well-studied glacier from each basin to enable visual and quantitative comparisons. However, as both reviewers correctly noted, this approach was not sufficiently representative of the full dataset. For example, in Figure 4 the central section of the glacier front shows close agreement among all datasets and aligns well with the underlying imagery, while the lateral margins display larger differences, where the manual delineation appears to capture the visible shape more consistently than the automated products. Nevertheless, this single example does not capture the overall variability in delineation quality across Greenland.

Following the reviewer's comments, we are currently implementing a systematic cross-dataset validation. Our dataset is compared against the four largest existing calving-front products—TermPicks, AutoTerm, CALFIN, and Black & Joughin (2023)—using at least one same-day pair for each glacier included across the datasets. Where exact temporal matches are not available, the external products are interpolated to the closest available date to enable a consistent comparison. For overlapping calving fronts, we calculate the Average Minimum Distance (AMD) between corresponding front traces and summarize the results using the mean, median, and per-glacier distributions illustrated through histograms. Furthermore, in order to demonstrate the high precision of the data, we will also conduct repeat digitation on the same calving front and compare the results.

To further characterize delineation quality, we have also introduced a new vertex density (vertices per km) quality flag, enabling users to filter out sparsely sampled traces that may be unsuitable for machine-learning training or validation. While sparse

vertex distributions often reflect genuinely smooth and regular calving fronts rather than coarse delineation, this metric provides a transparent way to assess geometric detail.

In summary, the revised manuscript will incorporate a comprehensive cross-validation with all major existing datasets to provide a clearer and more representative assessment of data quality. Upon completion of the expanded comparison, the full distribution of inter-dataset differences will be presented using histograms across all glaciers. In addition, Section 2.3 will be updated to present a comprehensive comparison of our product against both manual and automated delineation datasets.

Glacier names and type (marine-terminating, land-terminating) should be included in the metadata. The writing in many places is below standard and requires revision. There are missing references (most notably in Table 1), places with citations where none is expected (e.g. L106, L117), mistakes (e.g. L140, Table 2) and inconsistencies (e.g. use of outlet and marine-terminating glacier). Please refer to the list of specific comments below.

Response: We appreciate the reviewer's careful reading and detailed corrections. The metadata will be expanded to include the fields GlacierName, defined with reference to the map from NunaGIS, and GlacierType, classified as marine-terminating, land-terminating, lake-terminating, or floating tongue. The manuscript has also been thoroughly revised to correct misplaced citations, add missing references (particularly in Table 1), and ensure consistency in terminology (e.g., "outlet glacier" vs. "marine-terminating glacier"). Furthermore, overall readability and language quality will be improved throughout. Detailed point-by-point responses to these specific issues are provided below.

Response to Specific Comments

1. L23 / L87 –Please provide an exact number.

Response: The exact numbers will be updated and reported in the revised manuscript after supplementing with additional glaciers (see response to major comment above).

2. L29 / L281 / L286 –I don't think the analysis in this manuscript fully justifies this statement. / This statement is not justified based on the analysis carried out here. / I don't see any advantage in terms of spatial or temporal coverage. TermPicks has similar spatial coverage and much longer temporal coverage.

Response: We agree that basing the analysis on just six representative glaciers in the initial submission was insufficient to fully justify the reported accuracy range. This limitation primarily arose from the difficulty of identifying overlapping same-day observations for the same glacier across datasets. The originally reported accuracy range of 40–100 m was therefore derived from a

small number of available sites and will be updated— and made more robust— using our expanded dataset and comprehensive cross-validation. Nevertheless, the initial results already indicate that our manual delineations are generally more reliable than automated products, particularly along complex glacier margins.

Our dataset and TermPicks differ in scope and objectives, but they are complementary in terms of their spatial representation and temporal characteristics. Regarding TermPicks, we agree that it provides excellent long-term temporal coverage for marine-terminating glaciers. Our dataset, however, has a broader scope, encompassing marine-, land-, and lake-terminating glaciers, and this coverage will be further expanded in the revised version. In addition, our product is specifically designed to capture seasonal variability in calving-front positions. In the current version, most glaciers achieve at least two observations per year across the study period, and additional scenes are being incorporated to further enhance seasonal completeness. By contrast, TermPicks integrates results from multiple studies with varying temporal sampling. Its seasonal sampling reflects the availability of the underlying source datasets and therefore covers only a subset of glaciers and time periods, with variability among individual glaciers (Fig.2, (Goliber et al., 2022)).

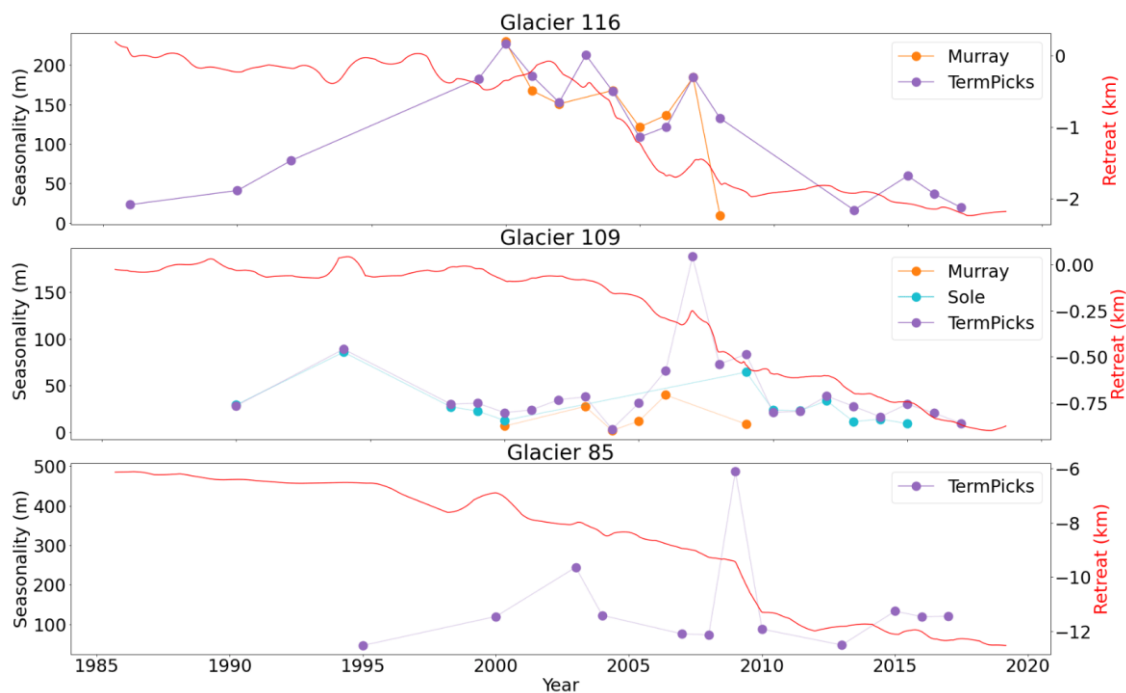


Fig.2 Seasonal sampling examples from the TermPicks dataset for Graae Gletscher (116), Heinkel Gletscher (109), and Humboldt Gletsjer (85) (Goliber et al., 2022).

3. L37 –Consider referring to the more recent IMBIE assessment (10.5194/essd-15-1597-2023)

Response: The reference has been updated to include the more recent IMBIE (2023) assessment (doi:10.5194/essd-15-1597-2023).

4. L43 / L89 –Why specifically retreat?

Response: Our original intention was to refer to the overall retreating trend observed and projected for most Greenland outlet glaciers in previous studies. However, we agree that using “retreat” alone may imply a preconceived bias and is not fully objective. We have therefore rephrased it as “calving front migration” in the revised manuscript.

5. L65 –Black and Joughin (2023) did not use deep learning in this publication.

Response: We thank the reviewer for the correction. The reference to Black and Joughin (2023) has been removed from this sentence in the revised manuscript.

6. L70 –When speaking of 'current algorithms', why use a reference from 2011?
Also, this statement is not entirely true.

Response: Thank you for pointing this out. We have updated the sentence to cite more recent studies and have revised the wording to provide a more accurate and balanced description of the capabilities and limitations of current delineation algorithms.

Text from the revised version (L70)

Recent studies show that automated calving-front algorithms would benefit from training data that better span diverse glacier geometries, have denser temporal sampling, include more complex surface conditions, and incorporate data from multiple sensors, thereby improving their applicability across regions (Cheng et al., 2021; Herrmann et al., 2023; Loebel et al., 2024).

7. L71 –This statement is misleading. I guess it only refers to the CALFIN product. The AutoTerm product has 278239 entries.

Response: We thank the reviewer for this clarification. In the initial submission, both CALFIN and AutoTerm (Zhang et al., 2023) were cited together because the latter also discussed similar limitations while proposing improvements. We have corrected the statement accordingly and now distinguish clearly between CALFIN and AutoTerm. The revised text now reflects the scale of existing products.

Text from the revised version (L71)

Despite being trained on more than 1,500 labelled fronts, advanced methods have extracted only ~22,000 Greenland calving-front positions (Cheng et al., 2021), a small fraction of the >400,000 available scenes. By incorporating the TermPicks dataset, Zhang et al. (2023) increased this to 278,239, demonstrating

that larger and higher-quality training data significantly improve model generalization, while indicating that intensive and more extensively sampled training sets are still required.

8. L61–84 –I feel like this paragraph needs to be completely restructured and given a common thread. The citations are all over the place. For example, why is Goliber et al. (2022, TermPicks) only mentioned in relation to the final statement and not for the product itself?

Response: We thank the reviewer for this valuable suggestion. The purpose of the original paragraph was to emphasise the need for reliable manually delineated datasets by outlining the limitations of existing automated approaches. As a result, most of the discussion focused on automated methods, and TermPicks, as a benchmark manual compilation, was only briefly mentioned at the end to highlight the importance of manual delineation. We agree that this structure weakened the narrative coherence and underrepresented the relevance of TermPicks.

In the revised manuscript, this paragraph will be substantially reorganized to improve its logical flow. TermPicks will be introduced earlier as a representative manual delineation dataset, discussed alongside automated products to provide direct comparison and context. This restructuring establishes a clearer link between manual and automated approaches and more effectively motivates the need for the new delineated product.

9. L81: We have products like this.

Response: We have reviewed and revised this sentence to more clearly highlight the distinctive features of this dataset, including its broad coverage across multiple glacier types and its seasonal temporal resolution.

Text from the revised version (L81)

Although several important products are now available, large-scale datasets that span diverse glacier types and offer broad seasonal sampling are still limited. This constrains the diversity of available training and verification data, and therefore reduces the opportunity to fully assess and further improve the generalization of emerging automated front-detection methods.

10. Table 1 –There are products missing: AutoTerm, ESA-CCI, <https://doi.org/10.5194/essd-14-4287-2022>, <https://doi.org/10.5194/tc-17-1-2023>, <https://doi.org/10.5194/tc-18-3315-2024>, <https://doi.org/10.18739/A2W93G>, <https://doi.org/10.22008/FK2/UNZUJF>, <https://doi.org/10.5194/tc-12-3813-2018>. Consider including a count of calving fronts. Isn't GEEDiT and ArcGIS manual as well?

Response: We thank the reviewer for these suggestions. In the revised version, we will include the missing datasets (AutoTerm, ESA-CCI, Gourmelen et al., 2022, Choi et al., 2023, Ehrenfeucht et al., 2024, Brough et al., 2018, Miller et al., 2023, and Shen et al., 2018) in Table 1, and add an additional column showing the total number of calving fronts reported in each product.

Regarding the question about GEEDiT and ArcGIS, we clarify that both are used for manual delineation, but in slightly different ways. Manual refers to calving fronts digitized entirely by the user after independently downloading and preprocessing satellite imagery. GEEDiT, in contrast, streamlines this process by providing an interface for direct delineation without requiring prior image download, making it a more efficient semi-manual tool. We therefore distinguish between manual and GEEDiT approaches for clarity in the table.

11. L106 / L117 –Places with citations where none is expected (e.g. L106, L117), check these citations.

Response: We thank the reviewer for pointing this out. The misplaced citations at L106 and L117 have been removed in the revised manuscript.

12. Table 2 –Landsat 5 has an image resolution of 30 metres and no panchromatic band.

Response: Thank you for the correction. The Landsat 5 entry has been updated to reflect its 30 m resolution and lack of a panchromatic band. This was a table-preparation oversight and has now been corrected.

13. L140 –The NDWI is usually calculated using green and near-infrared. Why was red used instead?

Response: Thank you for pointing this out. We used the MOD09GQ product, which provides only the red and near-infrared bands at 250 m resolution. MODIS products that include the green band have much coarser spatial (~500–1000 m) or temporal (8-day) resolution and were therefore not suitable for our application.

14. L147 –Why grounded ice? What about floating glacier tongues?

Response: We thank the reviewer for this comment. In the initial submission, we referred to “grounded ice” because almost all glaciers in the dataset terminate in grounded fronts directly interacting with the ocean, with only two cases involving floating ice tongues. To ensure accuracy and inclusiveness, this has been revised to ice front, which also encompasses floating glacier tongues.

15. L149 –I guess only for Landsat-7 and 8?

Response: The text has been clarified to specify that the true-color composites generated within GEEDiT were used to provide clear visual contrast for manual delineation for optical imagery.

Text from the revised version (L219-L221)

For optical imagery, true-color composites generated in GEEDiT were used to enhance visual contrast for manual delineation.

16. L155 –Please refer to which glacier ID has been used.

Response: We appreciate the reviewer’s suggestion. Following the comments from another reviewer, the updated dataset no longer relies on glacier IDs. Instead, each glacier is identified by its official name according to NunaGIS, as the update dataset will includes some non–marine-terminating glaciers not covered by existing ID.

17. L161 –I suspect this refers to the smallest possible error.

Response: We thank the reviewer for the clarification. The sentence has been revised to explicitly indicate that this refers to the resolution-limited lower bound, representing the smallest possible error.

18. L188 –From the satellite image in Figure 1, it looks like a land-terminating glacier. If that's true, that's probably why it hasn't been included in most other products. If that's the case, remove the subsequent statement.

Response: We thank the reviewer for this comment. As noted above, we have clarified the scope of the dataset, which includes multiple types of glaciers and is currently being expanded. Accordingly, Kangilinguata has been classified as a land-terminating glacier, and the corresponding description in Section 2.3 will be updated to reflect this adjustment.

19. L234 –Isstrøm is missing the "ø". Also in the figures. Please check.

Response: We thank the reviewer for pointing this out. All glacier names have been checked and corrected according to their official spellings based on map from NunaGIS and Bjørk et al. (2015), including the proper use of “ø” in Isstrøm and other figures.

20. L304 –This refers to the .shp files, whereas the product itself is a Geopackage (.gpkg).

Response: We appreciate the reviewer’s correction. The product was originally submitted in shapefile (.shp) format and later converted to GeoPackage (.gpkg)

following the editor's comments. This text was overlooked during the format update and has now been corrected.

21. Figure 10: How is this calculated? What was done with Humboldt Glacier, which is not included in the product?

Response: The values shown in Figure 10 were derived by averaging the 20-year calving-front change along the central flowline for all glaciers within each major drainage basin. Prior to submission, we removed several preliminary or lower-confidence delineations to ensure consistency and reliability across the full dataset. Humboldt Glacier was among the exclusions because its exceptionally wide front and highly fragmented northern margin generated irregular lateral endpoints that could not be standardized to a dataset-wide consistency level in the initial release. In the revised version, we will reprocess and include Humboldt Glacier to improve the spatial completeness of the dataset.

22. L350–359 –This does not really fit in this section, which should focus on the product and usage notes.

Response: We thank the reviewer for the comment. The paragraph describing the ISSM modeling application has been removed to keep this section focused on the data product and usage notes.

We hope the changes we have made address all points raised and bring the submission in line with ESSD's technical and editorial standards. Please let us know if further clarification or modification is needed.

References

- Björk, A. A., Kruse, L. M., and Michaelsen, P. B.: Brief communication: Getting Greenland's glaciers right – a new data set of all official Greenlandic glacier names, *The Cryosphere*, 9, 2215–2218, 10.5194/tc-9-2215-2015, 2015.
- Goliber, S., Black, T., Catania, G., Lea, J. M., Olsen, H., Cheng, D., Bevan, S., Bjork, A., Bunce, C., Brough, S., Carr, J. R., Cowton, T., Gardner, A., Fahrner, D., Hill, E., Joughin, I., Korsgaard, N. J., Luckman, A., Moon, T., Murray, T., Sole, A., Wood, M., and Zhang, E. Z.: TermPicks: a century of Greenland glacier terminus data for use in scientific and machine learning applications, *Cryosphere*, 16, 3215–3233, 10.5194/tc-16-3215-2022, 2022.
- Rosenau, R., Scheinert, M., and Dietrich, R.: A processing system to monitor Greenland outlet glacier velocity variations at decadal and seasonal time scales utilizing the Landsat imagery, *Remote Sensing of Environment*, 169, 1–19, 10.1016/j.rse.2015.07.012, 2015.