

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 reviewer,

We thank the reviewer for the careful and detailed evaluation. The original submission did not clearly demonstrate the dataset's scope and reliability, and we are undertaking substantial revisions to address all points raised. Your constructive suggestions have provided clear guidance, and we have revised the manuscript and dataset to further clarify their quality, coverage, and applicability.

1. Clarified dataset coverage and scope by explicitly including marine-, land-, lake-terminating, and floating-tongue glaciers; expanded digitization to improve spatial completeness; and added *GlacierName* and *GlacierType* fields to the dataset.
2. Implemented a comprehensive cross-dataset Average Minimum Distance (AMD) assessment by identifying or interpolating at least one overlapping delineation per glacier and comparing our dataset with TermPicks, AutoTerm, CALFIN, and Black & Joughin (2023).
3. Revised overly assertive wording, clarified the relevant limitations, and refined and expanded the method section.

We hope that the revisions fully address all of your comments, and we appreciate the detailed critique, which has significantly improved the clarity and quality of both the manuscript and the dataset.

Best regards,
Xi Lu and all co-authors

General comments

The authors describe the creation of a dataset of terminus positions from multi-source satellite imagery for ~290 marine-terminating glaciers containing ~12,000 individual traces. The terminus positions have been manually delineated, and I commend the authors for the effort that goes into that. However, I fail to see the novelty of the dataset.

The authors base the majority of terminus delineations on Landsat and Sentinel imagery and use existing methods such as GEEDiT, so that I find it difficult to believe that the dataset presented here has a higher accuracy than currently existing datasets. The presented dataset contains a variety of glaciers (land-, lake-, marine-terminating) so that a) the statement that the dataset contains delineations for 290 marine-terminating glaciers is incorrect, and b) the analysis of the data is skewed and needs to be revised. There are inconsistencies and a lack of detail throughout the methodology section that make it difficult to assess how the underlying satellite was processed. The validation of the data is

insufficient with one of the two methods applied not being available for the majority of the time period that the dataset covers, and the other being a comparison to existing data that does not provide a measure of accuracy of the dataset presented.

Response: We thank the reviewer for this thoughtful assessment. We agree that the dataset itself does not introduce methodological novelty, as it employs established tools such as GEEDiT and existing satellite archives. However, while recent novel deep-learning approaches have achieved notable progress, they remain highly dependent on large, reliable manual datasets for training and validation (e.g., Cheng et al., 2021; Goliber et al., 2022). Such manually curated data remain indispensable, particularly in complex terminus environments where automated methods can exhibit local errors of 1–2 km. In addition, the TermPicks compilation, which aggregates previously published manually delineated fronts, focuses solely on marine-terminating glaciers and lacks consistent seasonal resolution due to the heterogeneity of its source studies. For example, only 221 glaciers in TermPicks have seasonal resolution, and coverage varies by glacier and through time. Therefore, although not algorithmically innovative, our dataset provides a critical dataset for training, validation, and benchmarking, complementing ongoing advances in automated calving-front detection and supplementing existing manual datasets. The main contribution of this work lies in the creation of a spatially extensive, long-term, high-precision, and manually-verified dataset covering multiple glacier types, which offers essential reference material for both glaciological analyses and machine-learning development.

Regarding the dataset's scope and glacier classification, our initial goal—beyond supporting automated calving-front detection—was to create a dataset suitable for large-scale analyses of the seasonal behavior of Greenland outlet glaciers, such as joint studies with seasonal velocity products derived from Landsat imagery. Accordingly, we used the spatial extent of glaciers covered by Landsat-based seasonal velocity fields as the selection reference (Fig. 1; Rosenau et al. (2015)). Glaciers outside this extent were assumed to show minimal frontal change and were therefore excluded. Within this framework, the dataset effectively covers all glaciers exhibiting detectable seasonal velocity variations. Because marine-terminating glaciers constitute the majority of these outlets, the initial version erroneously referred to them collectively as “marine-terminating glaciers” without explicitly distinguishing glacier types.

Considering the availability of existing datasets and your suggestions, we have now refined the definition of coverage to explicitly include marine-, land-, lake-terminating and floating-tongue glaciers, rather than restricting it to those defined by long-term seasonal velocity coverage. The dataset is being extended to incorporate additional glaciers of each type, which will be clearly labelled in the metadata including GlacierType and GlacierName fields to enhance accessibility and usability for the wider cryosphere community. The additional glacier-front delineations were identified using a combination of existing inventories map from NunaGIS, glacier name from Bjørk et al. (2015), ice-marginal lake maps (Dømgaard et al. (2024)), and satellite imagery. Peripheral glaciers with negligible dynamic activity will be retained but not expanded.

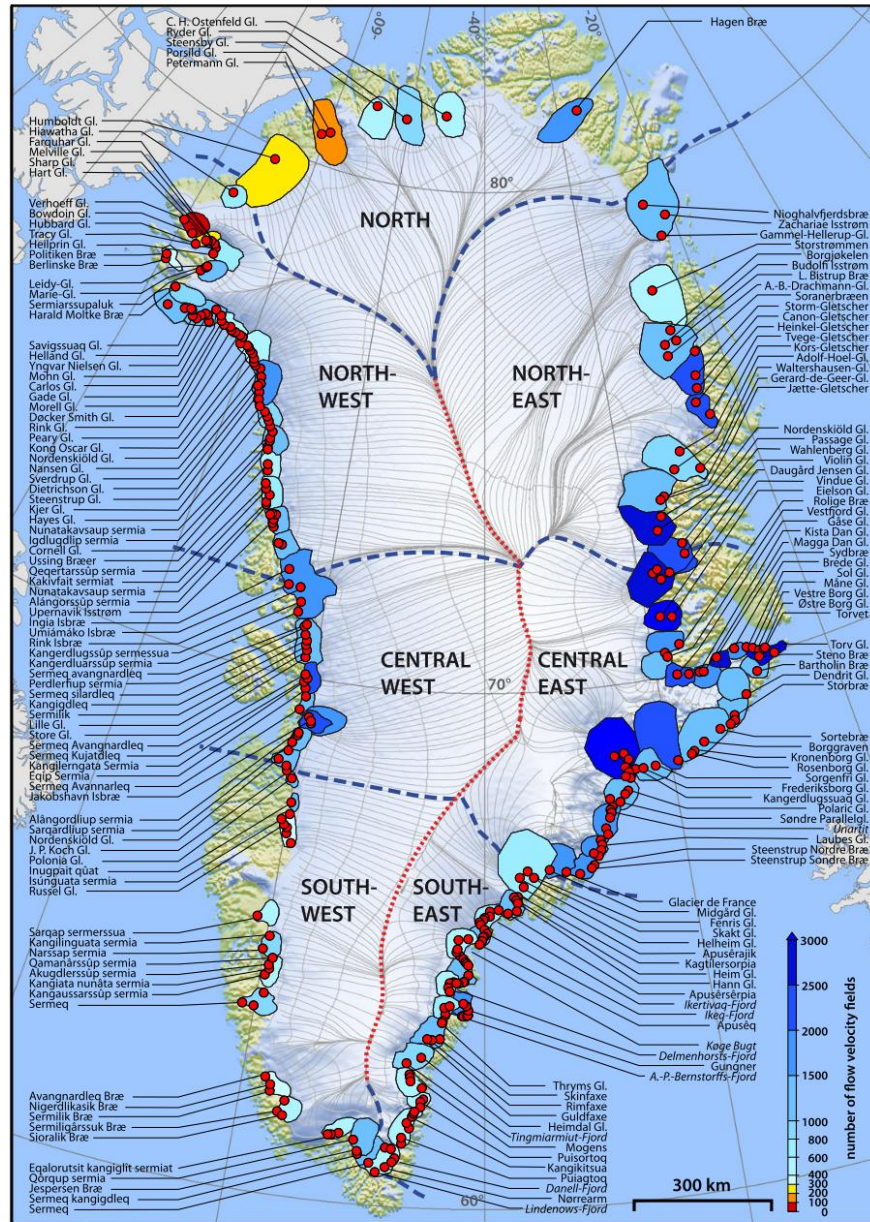


Fig. 9. 302 locations (red circles) and names of all analysed outlet glaciers. The official spelling of the glacier names was chosen according to NunaGIS (2015). Unofficial glacier names are marked italic. The grey flow lines are determined by Gebler (2009). The hill-shading was performed with the elevation 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).

To address concerns regarding validation and data reliability, we agree that the initial evaluation—based on six representative glaciers—was not sufficient to capture dataset-wide variability, mainly due to the difficulty of finding overlapping delineations for the same glacier on the same date. As also noted in response to another reviewer, we are now conducting an expanded validation that systematically compares our dataset with four major products (TermPicks, AutoTerm, CALFIN, and Black & Joughin, 2023) using Average Minimum Distance (AMD) metrics. To overcome the challenge of missing same-

day overlaps, we apply interpolation between available records from existing datasets to improve temporal matching. For each overlapping case, we calculate AMD statistics and summarize the results by glacier and by region (mean and median).

The methodology section is being revised to include detailed descriptions of satellite data preprocessing, delineation procedures, and validation steps. It will also clarify glacier selection criteria, specify the sensors and temporal intervals used, and explicitly distinguish between PlanetScope-based validation and inter-dataset comparisons.

In summary, while this dataset does not introduce methodological innovations, its value lies in providing the most spatially extensive, seasonally resolved, and manually verified record of calving front positions across Greenland. The forthcoming revised version will incorporate additional glacier types, extended validation, and detailed methodological clarification to strengthen its robustness and community relevance.

Response to major comments

1. The manuscript switches between the terms “marine-terminating outlet glaciers” and “outlet glaciers” making it unclear what the focus of this study is. When comparing the dataset presented to other existing datasets, the terminology makes a big difference as e.g. TermPicks (Goliber and Black, 2021; Goliber et al., 2022) contains only delineations for marine-terminating glaciers.

Response: Thank you for pointing this out. The dataset was initially defined by the extent of outlet glaciers with detectable seasonal velocity variations, without explicit distinction by glacier type. As marine-terminating glaciers make up the majority, the manuscript at times erroneously used “outlet glaciers” and “marine-terminating glaciers” interchangeably.

Following your suggestion, we are expanding the dataset to include more marine-, lake- and land-terminating glaciers, which are underrepresented in existing datasets but increasingly important for recent studies. In the revised version, we will use the general term “glaciers” for the dataset overall, while retaining type-specific terms where relevant. All terminology has been standardized, and glacier-type classification will be clearly defined in the methods. Each record now includes a GlacierType attribute in the metadata for clarity.

2. The validation of the terminus traces needs to be improved significantly.
 - a. Currently, visual validation is carried out using PlanetScope imagery which is only available from 2014 onwards and can therefore not be applied to the majority of the dataset. The authors show only 6 examples of how their delineations match with PlanetScope imagery and present no further quantification of errors. This is insufficient as a validation method.
 - b. A plausibility check is mentioned but not explained any further.
 - c. The only quantifiable validation is presented by an Average Minimum Distance (AMD) which is calculated between this dataset and TermPicks and/or MEaSUREs data. The authors do not explain which delineation is used as reference line (TermPicks or MEaSUREs) and present a single value as validation. The AMD

also does not provide a measure of accuracy but rather states the difference between two delineations with no indication as to which one is more accurate.

Response: We fully agree that validation is a critical component of this dataset and appreciate the reviewer's detailed comments.

(a) PlanetScope-based validation:

PlanetScope imagery (3 m resolution) was used for qualitative validation where such data are available (2014–2021). The initial purpose was to visually confirm the geometric fidelity of the delineated ice–ocean boundary, particularly in complex environments such as mélange-filled fjords, heavily shadowed termini, or regions with strong illumination gradients. Because the availability of such high-resolution imagery is limited, six representative glaciers from different sectors of Greenland were selected to illustrate how the manual delineations correspond to fine-scale features visible in the optical data.

In the revised manuscript, PlanetScope scenes are used not only for visual assessment but also for multiple repeated measurements of calving fronts. For these cases, AMD values are computed directly to provide a quantitative estimate of the precision of the repeated manual delineations and, where available, to support quantitative comparisons with contemporaneous entries from other datasets. We acknowledge that the restricted temporal availability of PlanetScope imagery does not represent the full 2002–2021 period. A broader quantitative cross-dataset validation (detailed below) is implemented to achieve comprehensive spatial coverage across the entire dataset, extending beyond the limited temporal window where PlanetScope observations exist.

(b) Plausibility check:

This step involves a post-processing consistency check, including a visual review of delineations based on time series of calving frontal changes extracted along each glacier's central flowline. Inconsistent fronts—such as those exhibiting abrupt, temporally unrealistic jumps—are removed based on this temporal consistency assessment. We will elaborate on this procedure in the revised methodology section and provide an example to illustrate its implementation.

(c) Quantitative validation and AMD metrics:

In the initial submission, AMD values were calculated between our dataset and existing manual reference datasets (TermPicks or MEaSURES) for six glaciers. Manual delineations are generally regarded as the most reliable reference, as they are based on visual interpretation of remote sensing imagery on a scene-by-scene basis. Consequently, automated products are typically validated against manual datasets, while manual compilations, such as TermPicks, assess internal consistency through comparisons of overlapping delineations from different authors (e.g., Goliber et al., 2022; TermPicks, median error = 107 m, range = 58.6–7,350 m). As a new manually delineated dataset, our product was initially validated quantitatively against overlapping entries from manual datasets (e.g., TermPicks) and qualitatively using both PlanetScope imagery and automated products to assess geometric fidelity, highlight the limitations of existing methods, and demonstrate the potential contribution of our dataset. This intercomparison-based strategy aligns with standard validation approaches used in existing calving-front products.

To further enhance reliability, we are now conducting an expanded cross-dataset validation against TermPicks, AutoTerm, CALFIN, and Black & Joughin (2023) using AMD statistics. Because exact same-day overlaps are rare (e.g., TermPicks manually traces only ~5.8 % of available Landsat scenes per glacier), we apply temporal interpolation between adjacent delineations in existing datasets to obtain near-coincident comparisons. Each comparison will include mean and median of AMD values, summarized by glacier and region, providing a transparent and quantitative measure of inter-dataset differences. Together, these enhancements establish a comprehensive validation framework combining high-resolution qualitative checks, inter-dataset quantitative analysis, and internal consistency screening. The full details will be provided in the revised methodology and validation sections.

3. The authors state that panchromatically sharpened Landsat-5 imagery is used to delineate glacier fronts, which is not possible as Landsat-5 does not have a panchromatic band. In addition, there is little detail on the processing of satellite imagery outside of GEEDiT (e.g. terrain corrections, filtering). The limited detail and inconsistencies in the description of the methods make it difficult to trust the resulting data.

Response: Thank you for pointing this out. This error was due to a table-editing oversight where the Landsat-7 entry was inadvertently duplicated. The text and table have been corrected to specify that panchromatic images apply only to Landsat-7 and 8. Additional details on image preprocessing—including radiometric and terrain correction, cloud and shadow masking, and contrast enhancement—will be added to ensure methodological transparency and consistency across all image sources.

4. The dataset is not consistent and needs thorough revisions. In its current state, the dataset contains five different types of glaciers:
 - a. Marine terminating glaciers (227)
 - b. Land-terminating glaciers (21)
 - c. Lake terminating glacier (5)
 - d. Peripheral glaciers (22)
 - e. Others including ice that I would not classify as a glacier (6), sections that are part of a larger glacier (1) or double delineations (3).

While the majority of the data is on marine-terminating glaciers, the inclusion of other glaciers skews the data and makes the presented analysis (Figures 10, 11) redundant. It should be ensured that the dataset matches the scope of the study i.e. if the study is aimed at marine-terminating glaciers, the dataset should only contain marine-terminating glaciers; if the study is aimed at outlet glaciers, the dataset should include all lake- and land-terminating glaciers. It should be made clear whether the dataset is focused on glaciers that are directly connected to the Greenland Ice Sheet or also includes peripheral glaciers. The authors further state the dataset contains comprehensive metadata, yet it contains less than comparable datasets such as TermPicks (Goliber and Black, 2021).

Response: Thank you for this valuable comment. As noted in our general response, the dataset's initial scope—beyond supporting automated calving-front detection—was designed to enable large-scale digitization of the seasonal calving front behavior of Greenland outlet glaciers. Accordingly, we used the spatial extent of glaciers covered by Landsat-based seasonal velocity fields as the selection reference (Fig. 1; Rosenau et al., 2015). Glaciers outside this extent were assumed to exhibit minimal frontal change and were therefore excluded. Within this framework, the dataset effectively includes all glaciers with detectable seasonal velocity variations without explicitly distinguishing glacier types.

Following your suggestion and considering the coverage of existing datasets, we have refined the definition of coverage to explicitly include marine-, land-, lake-terminating, and floating-tongue glaciers, rather than limiting it to velocity-defined outlets. The dataset is being extended to incorporate additional glaciers of each type, clearly labelled in the metadata to improve accessibility and usability for the cryosphere community.

5. The authors state that the presented dataset has a larger spatial coverage than existing dataset, which is misleading as it is only true when including non-marine-terminating glaciers.

Response: Thank you for this clarification. The original statement on spatial coverage did not distinguish glacier types and was based on the total number of delineated glaciers (~290 outlet glaciers), which—even without classification—already exceeds the coverage of existing products. As noted above, we are refining the dataset's classification and expanding it to include clearly defined marine-, land-, lake-terminating, and floating-tongue glaciers. Once the updated dataset is completed, the total number of glaciers will further increase. Nevertheless, in the revised manuscript, we will moderate this statement to avoid overly strong claims.

6. There is no consistency in glacier names throughout the manuscript with the example glaciers being spelled wrong and differently numerous times. I suggest picking either the official name or Greenlandic name for the example glaciers and use it consistently throughout the manuscript

Response: All glacier names will be corrected and standardized to their official or Greenlandic forms based on Bjørk et al. (2015) and map from NunaGIS to ensure consistency throughout the manuscript and figures.

Response to minor comments

1. **L19–20** –Maybe rephrase. It is not quite clear for what application it would be insufficient detail for. Also, did you test the performance and transferability of current datasets in automated calving front detection? If not, I would leave this out.

Response: Thank you for this suggestion. Our intention was to highlight that existing datasets provide limited glacier coverage and temporal sampling, which constrains their usefulness for training, validating, and assessing the generalization ability of automated detection algorithms. To avoid misunderstanding, we have rephrased the sentence to clarify that the limitation refers to spatial and temporal

consistency rather than any performance testing.

Text from the revised version (L19-L20)

Existing manual calving-front datasets vary in spatial and temporal coverage, limiting their ability to support the training, validation, and generalization of automated detection approaches.

2. **L21** –I fail to see how your dataset can be high accuracy (or higher than other datasets) given that you use Landsat and Sentinel imagery for the majority of the delineations. Similarly, when comparing your data to TermPicks I cannot see a higher spatial extent than other datasets. Consider omitting this sentence or rephrasing.

Response: As noted above, the reference to “high accuracy” was intended in comparison with automated products, rather than other manually derived datasets. From the six representative glaciers evaluated in the initial submission, our manually delineated fronts consistently performed better in complex regions—particularly along lateral margins—where automated methods often show large local errors. The earlier statement on spatial coverage referred to the total number of glaciers represented in the dataset. We will clarify this wording in the revised manuscript, especially as the ongoing expansion effort aims to include additional marine-terminating as well as other glacier types.

Given that these advantages are context-dependent, we have revised the sentence in the updated manuscript to more accurately describe the dataset’s strengths and limitation.

Text from the revised version (L21)

This dataset provides spatially extensive, seasonally resolved, and internally consistent delineations across multiple glacier types, complementing existing calving-front records.

3. **L23** –Why do you say approximately 12,000 traces for approximately 290 glaciers. You should state the accurate numbers, which are easily determined.

Response: Thank you for the suggestion. The approximation is due to the fact that some very close-by glaciers were digitized in the same file. The exact numbers will be reported in the revised manuscript once the ongoing dataset expansion is completed.

4. **L28** –It is unclear what the offsets of 40-100m relate to in this context. Consider rephrasing.

Response: The offsets of 40–100 m refers to the typical positional differences between our manually delineated fronts and reference manually datasets (TermPicks and MEaSURES) at overlapping dates. This has been clarified in the revised text.

5. **L37–40** –I would combine these two sentences e.g. Mass loss from the Greenland Ice Sheet contributes significantly to global sea level rise (Sheperd et al., 2018; Frederikse et al., 2020), with nearly half of this dynamic ice loss being attributed to frontal ablation (Enderlin et al., 2014; Mouginot et al., 2019).

Response: The two sentences have been combined and rephrased for conciseness as follows:

Text from the revised version (L37-L40)

Mass loss from the Greenland Ice Sheet contributes significantly to global sea-level rise (Shepherd et al., 2018; Frederikse et al., 2020), with nearly half of this dynamic ice loss attributed to frontal ablation (Enderlin et al., 2014; Mougnot et al., 2019).

6. **L44 & L47** –Consider adding references for Fahrner et al., (2025) and KC et al., (2025) as these publications address frontal ablation over large spatio-temporal scales. Consider adding Fahrner et al., (2021) to the reference list as you do refer to it later on.

Response: Added citations Fahrner et al. (2021, 2025) and Kc et al. (2025).

7. **L66** –“... marine-terminating glaciers show highly non-linear behaviour “– While this is true for individual glaciers, many studies have shown that their behaviour on a regional scale is linear (e.g. Cowton et al., 2018; Fahrner et al., 2021). Consider rephrasing.

Response: Thank you for this comment. The sentence has been rephrased accordingly to reflect this distinction.

Text from the revised version (L66)

These approaches are especially valuable for monitoring Greenland’s marine-terminating glaciers, which can exhibit nonlinear behavior in response to climatic and oceanic forcing at the individual-glacier scale (Brough et al., 2023; Cowton et al., 2018; Fahrner et al., 2021).

8. **Figure 1** –Please use the full names of the glaciers (e.g. Petermann Glacier, Jakobshavn Isbræ), check the spelling (e.g. Isstrøm) and be consistent (Gletscher vs Glacier). Panels should be labelled from A) overview map to G) Kangilinguata Sermia. Also "Sermia" should be spelled with a capital "S". This glacier is also not marine-terminating as is clearly visible in the image which contradicts previous statements that the dataset is looking at marine-terminating glaciers.

Response: As noted previously, all glacier names will be corrected and standardized according to map from NunaGIS and Bjørk et al. (2015). In addition, since the revised dataset will include both marine- and non-marine-terminating glaciers, the related descriptions in the text will be updated to avoid implying an exclusive focus on marine-terminating glaciers.

9. **L108** –GEEDiT is the acronym for the “Google Earth Engine Digitisation Tool” not the Glacier Extraction and Evaluation Dataset Tool as stated (Lea, 2018).

Response: The full name has been corrected.

10. **L113–114** –How does the use of these auxiliary images impact delineation uncertainties - e.g. how can you be certain that the calving front has not changed within ± 15 days?

Response: Thank you for raising this important point. The ± 15 -day temporal window was applied only when the target Landsat-7 scene was affected by Scan Line Corrector (SLC) gaps or when crucial frontal areas were partially obscured by cloud cover. In such cases, auxiliary scenes within ± 15 days were used only for visual cross-checking, not for direct delineation. The final glacier front was always

- digitized from the primary scene to maintain temporal consistency. This clarification has been added to the revised methods section, together with a note acknowledging that this approach may introduce minor additional uncertainty for glaciers experiencing rapid frontal change.
11. **L120** –Be consistent with the naming of the satellites e.g. most satellites are introduced by the written-out name followed by the abbreviation in parenthesis except for ERS1/2.
Response: We have standardized the naming of all satellite missions throughout the manuscript. Full names are now provided at first mention, followed by their abbreviations in parentheses.
12. **Table 2** –Landsat-5 does not have a panchromatic band so it cannot be resampled to a 15 m resolution.
Response: We appreciate the reviewer’s correction and apologies for the oversight. The information for Landsat 5 has been corrected.
13. **L133–137** –“...quality control and metadata creation” replace “and” with “,”; Repetition from line 125-126. Remove
Response: The redundant sentence has been removed, and the phrasing has been revised to replace “and” with a comma as suggested.
14. **L137** –“ MODIS OD09GQ?” - Is that specific MODIS imagery and therefore different from the other MODIS imagery you are using? Be consistent with the naming of data - if this is the MODIS data you use throughout, stick with the actual name. Also, auxiliary dataset for what? MODIS data?
Response: The MODIS product used is specifically MOD09GQ, and we have now standardized this naming throughout the manuscript for consistency. Glacier frontal changes are generally gradual and follow seasonal patterns, so consecutive delineations typically capture the overall migration trend. However, when two temporally distant delineations appeared nearly coincident, MOD09GQ imagery was used to assess whether intermediate advance–retreat cycles or local nonlinear variations occurred, ensuring that all key transitions were captured. In a few cases—where high-resolution imagery was unavailable, but the terminus exhibited predictable stable and near-linear behavior—MOD09GQ scenes were also used for delineation to maintain seasonal continuity.
15. **L138**–Is the NDWI only calculated for MODIS?
Response: Yes, the NDWI was calculated only for MODIS imagery. Since the MOD09GQ product provides the red and near-infrared bands at 250 m resolution, we adopted a red–NIR normalized difference index (Red–NIR NDWI) to enhance the contrast between ice and open water.
16. **L141–142**–I struggle to see how an image with a resolution of 250m provides a reliable source for detecting a calving front even as a reference. Maybe provide more detail here.
Response: As noted above, MODIS imagery was used only in a limited number of cases, and not as the primary source for delineating a calving front. These situations occurred at glaciers where consecutive high-resolution scenes (Landsat/Sentinel) showed a stable, near-linear pattern of terminus change over an extended period. In such situations, MODIS served only to support interpolation between well-constrained high-resolution positions and to maintain seasonal continuity, rather

- than to identify the front independently. All instances in which MODIS contributed to temporal interpolation have been explicitly documented in the metadata and can be filtered by users as needed.
17. **L142-143**—“ENVISAT ASAR (Level-1B) and ERS-1/2 SAR (Level-1.5) scenes were downloaded from the ESA archives” can be removed as you already state prior that the data was downloaded from ESA.
Response: It has been removed.
18. **L143-144:** Be more specific regarding the calibration and terrain correction, e.g., what DEM was used for the terrain correction?
Response: Thank you for the suggestion. We have added details specifying that radiometric calibration and terrain correction were performed using the Sentinel Application Platform (SNAP) toolbox, with topographic correction based on the ArcticDEM. This information has been included in the revised Methods section to ensure transparency and reproducibility.
19. **L145**—Same as above, what DEM was used and what filter applied.
Response: We thank the reviewer for the comment. The revised manuscript now specifies that ArcticDEM was used and that a Refined Lee filter was applied. These details have been added to the method section for clarity.
20. **L146**—Why was the data not reprojected to EPSG:3413 Polar Stereographic? This is the common format for datasets that cover Greenland.
Response: The dataset was not deliberately projected to a specific coordinate system; rather, EPSG:4326 (WGS84 geographic coordinates) was adopted as a common reference to unify all imagery and derived products. This ensures consistency across data sources with different native projections and facilitates straightforward use in most GIS and web-based visualization platforms.
21. **L148-151**—These sentences contradict each other as you first say that delineations were done on native-resolution imagery and then say that you used pan-sharpened images. For example, you say in Line 113-114 that this was only done for Landsat-7, but in Line 150-151 now it reads like it was done for other satellites too?
Response: Thank you for the comment. The methods section originally described multiple data sources separately, which may have caused confusion. To clarify, all optical imagery was delineated at its native spatial resolution, with GEEDiT providing visually enhanced true-colour composites for clearer interpretation. In a few specific cases, auxiliary images from nearby dates were used for reference: for example, optical images affected by cloud cover, and Sentinel-1 images where the ice–water boundary was indistinct. These auxiliary scenes were used only for visual cross-checking, not for direct delineation. The text has been revised accordingly.
22. **L152-153**—How did you manually adjust the delineations? Did you use a different image or the same image? Be more specific. Please explain what the plausibility check entails as this is the only mention of it.
Response: When clear inconsistencies were observed between the current front and preceding and/or following fronts, suggesting that the delineation might not be reliable, it was rechecked and manually refined using the original image. The plausibility check refers to a post-processing step applied after extracting calving front position variation along the glacier centerline. Time series of frontal positions were examined for expected seasonal or gradual variations. If an outlier or abrupt

change inconsistent with neighboring observations was identified, the corresponding delineation was rechecked against the original imagery to confirm or correct potential errors. This procedure ensured temporal coherence and reduced spurious fluctuations in the final dataset.

23. Figure 2 –The figure needs to be redone as it is currently confusing. Under preprocessing you mention everything that has been done to the satellite imagery, but you don't specify what was done to which imagery. It is somewhat clear that the imagery on the left corresponds with the boxes, but I think it could be made clearer.

I don't think the "Sets parameter." is necessary unless you specify which parameters. I am still confused by the editing and modification of the delineations.

The QA process is not clearly explained at all and needs additional details. Also, again check the glacier names for consistency, spelling etc.

Response: Figure 2 will be fully redesigned to improve clarity and logical consistency. The revised version will distinguish preprocessing steps for different imagery types (Landsat, Sentinel, MODIS, SAR), specify the applied parameters and operations, and adds further detail to the delineation and QA stages (see revised manuscript for more information).

24. L162 –You mention uncertainties in delineation of 25m, however Brough et al., (2019) and (Fahrner et al., (2021) contradict this statement as they performed error checks for Landsat 4 (30 m) and Landsat 8 (15 m) and came to the conclusion that the error is much less than the pixel resolution.

Response: The 25 m uncertainty mentioned refers to the positional deviation relative to the original satellite imagery (precision), rather than to repeated manual digitization on a single image. As noted by another reviewer, this represents a conservative estimate, since additional uncertainties can arise from cloud cover, side-wall shadows, or fragmented fronts—conditions under which errors can reach several hundred meters (e.g., TermPicks, median error = 107 m, range = 58.6–7,350 m). The studies by Brough et al. (2019) and Fahrner et al. (2021) indeed evaluated the internal consistency of manual delineation (i.e., repeated tracing of the same image by the same operator), **which measures precision rather than accuracy.** This sentence has been revised for clarity to distinguish between these error sources.

25. L165–166 –Planet imagery is only available from 2014, so there is no data for over half of the time period that your dataset covers. I fail to see how this can be a quality check, especially since it is also only a visual validation. I suggest the authors should find a way to quantify the uncertainties.

Response: As noted, PlanetScope imagery (available only after 2014) was used solely for qualitative visual assessment to illustrate the geometric reliability of the manual delineations, rather than as a dataset-wide validation tool. In the revised manuscript, PlanetScope scenes are additionally used for multiple repeated measurements of fronts, where AMD values are computed directly to provide a quantitative estimate of manual delineation precision and, where available, are quantitatively compared against contemporaneous entries from other datasets. To provide a more comprehensive and quantitative evaluation, the revised version also incorporates an expanded cross-dataset validation against TermPicks, AutoTerm, CALFIN, and Black & Joughin (2023) using Average Minimum Distance (AMD)

- metrics. Temporal interpolation between adjacent records is applied to address the scarcity of exact same-day overlaps, allowing near-coincident comparisons across the entire dataset. This approach provides a more representative and quantitative measure of overall dataset reliability.
26. **Line 175** –“Its fine spatial detail and temporal coverage make it a valuable benchmark for evaluating glacier front delineations.” That might be correct, but it is only available from 2014 onwards so that it is not useful for longer time series such as your dataset.
Response: We agree that PlanetScope imagery is only available from 2014 onward. Similar to the previous point, in the revised manuscript, its role is clarified as a quantitative assessment to confirm delineation precision. And the primary long-term validation is based on cross-comparison with multi-temporal products (TermPicks, AutoTerm, CALFIN, and Black & Joughin, 2023) using AMD metrics.
27. **Line 177** – Glacier names are incomplete and spelled wrong.
Response: We thank the reviewer for pointing this out. All glacier names will be checked and corrected according to their official spellings based on NunaGIS and Bjørk et al. (2015).
28. **L183** –“...capturing key features with sub-pixel accuracy...” if I understand this correctly you say that your delineations capture sub-pixel features. If so, I think that might just be coincidence as I don’t think you’d be able to accurately delineate features within a pixel. Otherwise, please rephrase.
Response: The phrase “*sub-pixel precision*” referred to the delineations made relative to the imagery on which they were drawn, as verified against higher-resolution PlanetScope imagery. We agree that this wording was unclear and could be misinterpreted, so it has been revised to clarify that the delineations accurately capture the ice–ocean boundary at the resolution of the original source imagery.
- Text from the revised version (L183)*
In all six cases, the manually delineated fronts showed strong geometric agreement with PlanetScope imagery, accurately capturing the visible ice front–ocean boundary at the higher resolution imagery.
29. **L188–189**–I don’t think the fact that this glacier is included in this dataset but not in others warrants the claim that this dataset has a greater spatial coverage. Especially since some of the other datasets do not contain land-terminating glaciers such as Kangilinguata Sermia. Consider rephrasing.
Response: We agree that the inclusion of this single glacier does not in itself justify a claim of broader spatial coverage. As noted above, this statement will be revised to clarify that the dataset complements existing products by incorporating additional glacier types—such as land- and lake-terminating glaciers—that are underrepresented by previous datasets.
30. **L207**– Rephrase. I think you're trying to say that your dataset outperforms automated approaches but not all manually delineated datasets.
Response: Thank you for the suggestion. Once the dataset update is completed, we will perform a comprehensive cross-site comparison, and the corresponding statement will be adjusted accordingly.
31. **Line 227:** Make sure the citation is correct – You are citing Goliber et al., (2022)

which is the publication, but you should be citing the dataset Goliber and Black, 2021)

Response: Thank you for your careful attention to detail. The citation has been corrected to reference the appropriate dataset — Goliber and Black (2021) — instead of the related publication (Goliber et al., 2022).

32. **L226–228** –I struggle to see how a close alignment to existing datasets is showing that your dataset is novel or more accurate. Maybe rephrase?

Response: The statement on higher accuracy referred specifically to comparisons with automated products, where our manually delineated fronts showed smaller deviations in complex terminus regions. The close agreement with other manually delineated datasets (e.g., TermPicks) instead demonstrates consistency and reliability, confirming that our dataset is accurate and compatible with established products rather than implying methodological novelty. This sentence has been revised.

33. **L236** –It is unclear here if TermPicks of MEaSUREs is used as reference line. I assume that the AMD was calculated for both and only the maximum is mentioned? Please be more specific.

Response: Yes, the AMD values shown in Figure 4 and related figures were calculated between this dataset and the most reliable available manual reference—TermPicks when available, and MEaSUREs otherwise. The reported values represent the mean differences across the entire calving front, which are slightly larger than the corresponding median values. This clarification has been added to the revised manuscript.

34. **L279–280** –Replace “high positional consistency with” with “corresponds well with”

Response: Thank you for the suggestion. It has been replaced.

35. **L280–281** –“... depending on glacier geometry, image quality, and surface contrast.” I would leave this out as you do not provide any measure on how any of these affect your accuracy.

Response: It has been removed.

36. **Line 281**: This is the first time semi-automated methods are mentioned. Unless it is introduced prior with a specific reference, I would leave it out.

Response: It has been removed.

37. **L286–287** –From the figures in the manuscript, I struggle to see the increased spatial and temporal coverage compared to TermPicks. If anything, the dataset presented here seems to have less temporal coverage than existing datasets (e.g. Figure 7).

Response: The original spatial-coverage statistics were based on the total number of delineated glaciers, which exceed the glacier number of existing products. As noted above, we are refining the dataset’s classification and expanding it to explicitly include more marine-, land-, lake-terminating, and floating-tongue glaciers. Once the updated dataset is completed, the total and marine-terminating glacier count will further increase. The corresponding figure will be revised and annotated to reflect the expanded coverage.

TermPicks and our datasets differ in scope and objectives, and 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. 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)).

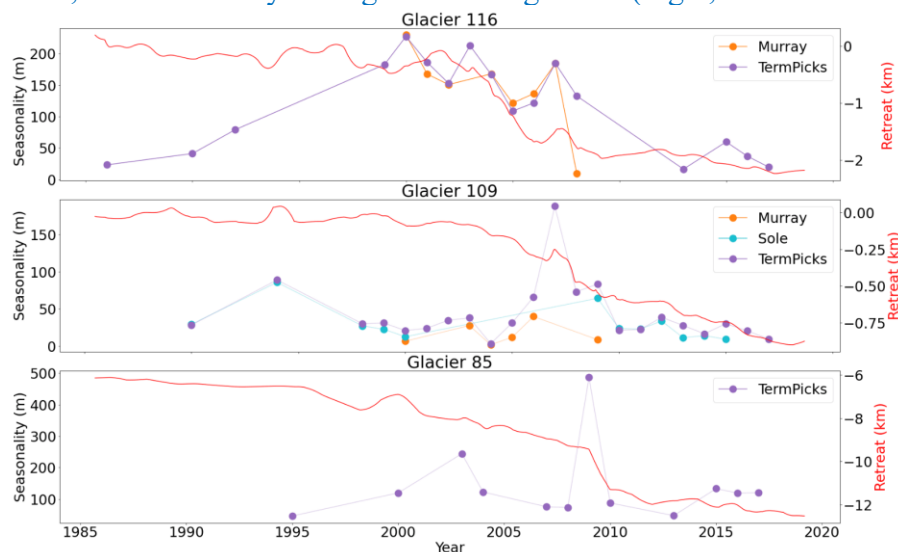


Fig.2 Seasonal sampling examples from the TermPicks dataset for Graae Gletscher (116), Heinkel Gletscher (109), and Humboldt Gletscher (85) (Goliber et al., 2022).

38. **L287–288** –Given that the dataset contains a multitude of glaciers that are land- and lake-terminating and has no consistent temporal resolution, I would remove this sentence. Currently the dataset contains 227 marine-terminating glaciers, and the temporal resolution varies from annually to monthly.

Response: Thank you for this comment. As discussed above, the initial version of the dataset was indeed limited to glaciers within the Landsat-based velocity coverage and thus included a mix of marine-, land-, and lake-terminating types with variable temporal density. Following your suggestion, we are refining the dataset’s scope and classification to explicitly distinguish glacier types and to standardize temporal resolution.

In the revised version, all glaciers will be clearly labelled by type (marine, land, lake, or floating-tongue) in the metadata, and the temporal sampling will be harmonized to a consistent seasonal interval where image availability allows. The sentence in question has been removed or rephrased accordingly to avoid overstatement of temporal uniformity.

39. **L289–290** –It states here that there is “high-precision agreement with PlanetScope imagery” however earlier it is stated that this was only accessed visually. There is

no proof in the manuscript that would support the claim of “positional discrepancies are consistently below 10m”. I suggest the authors rephrase or provide a detailed description of how this was determined.

Response: Thank you for pointing this out. In the revised manuscript, we have clarified the role of PlanetScope imagery and rephrased the corresponding statements. PlanetScope scenes are now used for multiple repeated measurements of calving fronts, where AMD values are computed directly to quantify the precision of the repeated manual delineations and, where possible, to allow quantitative comparisons with contemporaneous entries from other datasets. In addition, quantitative accuracy assessment is now addressed through an expanded AMD-based cross-dataset validation with TermPicks, AutoTerm, CALFIN, and Black & Joughin (2023), as detailed in the revised validation section.

40. **L301** –The name of the individual files “Calving_front_positions_for_Greenland_outlet_glaciers_2002_2021 -New_NeighborGID115” is unnecessarily long. I would recommend to either just use the Glacier ID or use the glacier name. In the current format, the key information (GlacierID) is almost hidden and I have to use half my screen in QGIS to see it.

Response: Thank you for this practical suggestion. The updated dataset no longer relies on glacier IDs. Instead, each glacier is identified by its official name according to map from NunaGIS and Bjørk et al. (2015), as the supplemented dataset also includes several non-marine-terminating glaciers not covered by existing ID.

41. **L308** –The metadata should contain at least the official glacier name as specified in TermPicks or Bjørk et al., (2015).

Response: The official glacier names will be added to the metadata according to map from NunaGIS and Bjørk et al. (2015).

42. **L313** –The “season” field can be left out as it is quite clear from the “Date” field what season the delineation was taken in.

Response: It will be removed from the updated dataset.

43. **L318** –Check the glacier names for spelling and consistency.

Response: They will be checked and corrected according to map from NunaGIS and Bjørk et al. (2015).

44. **L344** –Write these values as values $\pm$ STD km.

Response: They will be expressed as mean $\pm$ SD km.

45. **L350–351** –If you only model one glacier it's not really the dataset that you are using but a subset. I would leave this sentence out as it doesn't really add anything.

Response: It has been removed.

46. **L376** –It should be mentioned here that the validation against PlanetScope imagery was visual only, and that the AMD was only determined for TermPicks and/or (?) MEaSUREs. Currently it reads like the validation of the dataset was extensive, which unfortunately, it was not.

Response: In the revised manuscript, this section has been rewritten to clarifying their respective purposes and scope. As noted in earlier responses, we are now conducting repeated measurements to quantify the precision of the manual delineations based on PlanetScope imagery. And an expanded cross-dataset validation that includes TermPicks, AutoTerm, CALFIN, and Black & Joughin

(2023) to provide a more systematic and transparent assessment of accuracy and representativeness.

We thank the reviewer again for the thorough and constructive feedback, and we hope that our responses and revisions fully address all concerns.

Reference

- 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.
- Dømgaard, M., Kjeldsen, K., How, P., and Björk, A.: Altimetry-based ice-marginal lake water level changes in Greenland, *Communications Earth & Environment*, 5, 365, 10.1038/s43247-024-01522-4, 2024.
- 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.
- Kc, A., Enderlin, E. M., Fahrner, D., Moon, T., and Carroll, D.: Seasonality in terminus ablation rates for the glaciers in Greenland (Kalaallit Nunaat), *The Cryosphere*, 19, 3089-3106, 10.5194/tc-19-3089-2025, 2025.
- 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.