Author response to "Editor decision: Publish subject to revisions"

At the recommendation of the editor, we included two new results that focus on the demonstration of the usefulness of TermPicks that was missing in individual datasets. Section 2.6 in the methods describes how we estimated seasonality and sinuosity from the TermPicks dataset. Estimating seasonality showcases that with TermPicks, seasonality can be determined for a far greater number of glaciers over a longer time span than any author alone. This is shown in Figure 7 and described in Results (Line 205) and Discussion (Line 281). We also showcase the need for the fullwidth terminus trace instead of a centerline-only trace through the calculation of sinuosity for glaciers shown in Figure 8 and described in Results (Line 213) and Discussion (Line 285). As we included two new figures, one of which showing seasonality, we did not include the updated figure 11 (Originally Figure 8) that we included in our response to reviewer # 3. Our new figure 7 shows how the inclusion of additional data estimates seasonality in a more sufficient way than our previous figure for many glaciers. Inclusion of these results led to some restructuring of the methods, results, and discussion section.

The Track-changes document is not in the proper Copernicus format as my compiler time-out (Overleaf) when attempting to use a LaTeX Diff file using a Copernicus document class. All relevant changes should be there, except for changes made to the Authors list (adding town, Country information).

TermPicks Referee #1

In this manuscript, the authors have described a dataset of manually digitized terminus positions for outlet glaciers of the Greenland ice sheet compiled from previously-published datasets, in order to provide a consistently-formatted training dataset for future machine learning applications. This is an excellent and timely undertaking that highlights the power of collaborative efforts.

On the whole, the manuscript does a good job describing the issues involved in combining "input" datasets from multiple authors, as well as describing the "ouptut" dataset, and even manages to show an example application of combining data sources. Accordingly, I only have a few minor comments/suggestions to make on the manuscript. The bulk of my comments/suggestions have to do with the description of the metadata - I think a Table with a few different example entries would help clarify this for a reader.

We appreciate the constructive and positive feedback on the manuscript. We address the comments bellow, but also included a new figure (10) of the metadata of three glaciers to better clarify the structure. As we addressed comments, the original line numbers of the text may have changed in the final manuscript. The changed text has been noted in the responses to individual comments. Our responses are in blue below each comment.

Comments to Address:

- l. 10: is this the mean (± standard deviation)?
    - Yes. This was changed to “The TermPicks data set includes 39,060 individual terminus traces for 278 glaciers with a mean of 136±190 and median of 93 of traces per glacier” to be more clear. The SD is higher than the mean due to the high variation of picks between certain glaciers.
- l. 52: check that months are removed from the reference dates
    - These have been checked and have removed from the text.
- l. 104: is the Howat reference here for the MODIS image?
    - The incorrect MEaSUREs image was being cited here. It has been changed to “MEaSUREs Greenland Ice Mapping Project (GIMP) 2000 Image Mosaic (Howat et al, 2014; Howat,2018).”
- l. 130 (Date): I found this description slightly confusing - are there 5 columns (one column for the date string, four columns for the year, month, day, and

decimal date)? From the dataset, I see that it is indeed five individual columns, but the header makes it seem like there's only one column here (Date).

    - Text changed to “Date Columns: The Date column represents the acquisition time for the image used to pick the terminus for that trace. There are 4 additional columns for year, month, day and decimal date” for clarity.

- l. 135 (Satellite): How is this formatted/written?
    - Added text “The names used are in listed in Table 2”. Table 2 lists the satellite names.
- l. 144 (Scene ID): here again, it would be helpful to have more information about this. The Landsat Product ID/other satellite IDs are relatively straightforward, but what about the aerial images?
    - If an author provided satellite ID information, then we do not change it - if someone is using TermPicks for machine learning, then they may need access to the original data. This assumes it is easier for them to request it with the original name. We added text “It includes information on the date and location for the original image. This may be listed as a file name the original author used and may store locally (Figure 10; Glacier 291) or a scene ID from a different satellite (e.g. Sentinel-1 product folder name)” for clarity. Figure 10 Glacier 291 shows an example of an original image name.
- l. 155 (Quality Flag): What does this entry look like for a given image? From the dataset, I see that it's comma-separated 2-digit strings (00, 01, 02, 03, 04, 05) - I'm not sure I would have gotten that from the description here.
    - Added text “If there are multiple flags, they are separated by commas (Figure 10; Glacier 278)” for clarity. Figure 10 Glacier 278 shows an example of multiple flags.
- l. 170: where do the glacier centerlines come from?
    - Text added “Centerlines are manually mapped from the MEaSUREs Greenland Ice Mapping Project (GIMP) 2000 Image Mosaic (Howat et al., 2014; Howat, 2018).”
- l. 226: how many of these picks needed manual checking?
    - Only 220 traces were checked manually for this section. Text changed to “Traces with >500 m error between traces were manually checked for errors (220 traces).”

- l. 228: wouldn't it make more sense to compare the image (assuming it exists) against the different picks, rather than using the completeness of the metadata?
    - The method we used to compare traces between large errors in multiple authors assumes the large error is due to mislabeling the date (i.e. the trace did not appear to be from the same front on the same date as there is a large step change in the traces). The author that included the original image likely kept detailed record of what image was used and therefore is less likely to have incorrectly listed the date. As this was a very small subset of the dataset (~0.4%) we chose not to manually check each trace.
- Figure 5: I really like this figure.
    - Thank you!
- The GEEDiT walkthrough is great - have you thought about putting it on github pages (https://pages.github.com/) so that it's more widely visible/available?
    - GEEDiT TermPicks has been put into a repository (https://github.com/jmlea16/GEEDiT-TermPicks) documenting the walkthrough and program.

TermPicks Referee #2

The manuscript from Goliber et al. collates terminus shapefile from a variety of different published studies into one dataset, complete with metadata, with the ultimate aim that the dataset could be used as training data for machine learning.

I think this is both an excellent manuscript and dataset and I enjoyed having a look through the dataset and the associated Google Earth file. I certainly recommend the publication of this manuscript in The Cryosphere. I do have a few very minor comments which the authors may wish to consider.

Thank you for the positive feedback and comments on the manuscript, and we are glad you enjoyed looking through the data. As we addressed comments, the original line numbers of the text may have changed in the final manuscript. The changed text has been noted in the responses to individual comments. Our responses are in blue below each comment.

Line 91: Why exclude glaciers with less than two authors digitizing them? What is the rationale for this?

- My text here is unclear and overly complicated. We decided to exclude glaciers with only a single trace and therefore no timeseries information. These were generally glaciers that were very small. The text has been changed to "We excluded terminus picks where only one pick was available for the glacier over all authors as well as land-terminating glaciers" for clarity.

Section 3.2: Is there a bias here, in that most of the repeated terminus picks I presume are from the later periods i.e. 2000-2020. Here the imagery is of much superior quality, which would result in a lower error. In particular most of the Landsat-1 scenes have a pretty poor geolocation accuracy and often require a manual correction, could this result in a much larger error?

- Yes, this may be the case and we do find slight difference between errors pre-2000s to post 2000s. However, much of the largest errors (>5k) are in the 2000-10 due to differences in tracing of fractured ice tongues. The figure below is for dates with more than one trace from at least 2 different authors with a Hausdorff distance of <500 m. It shows there is a slight increase in error (<200m) in the 2000s but there are also more duplicated traces during that time overall. These are all the dates pre-removal noted in line 232.

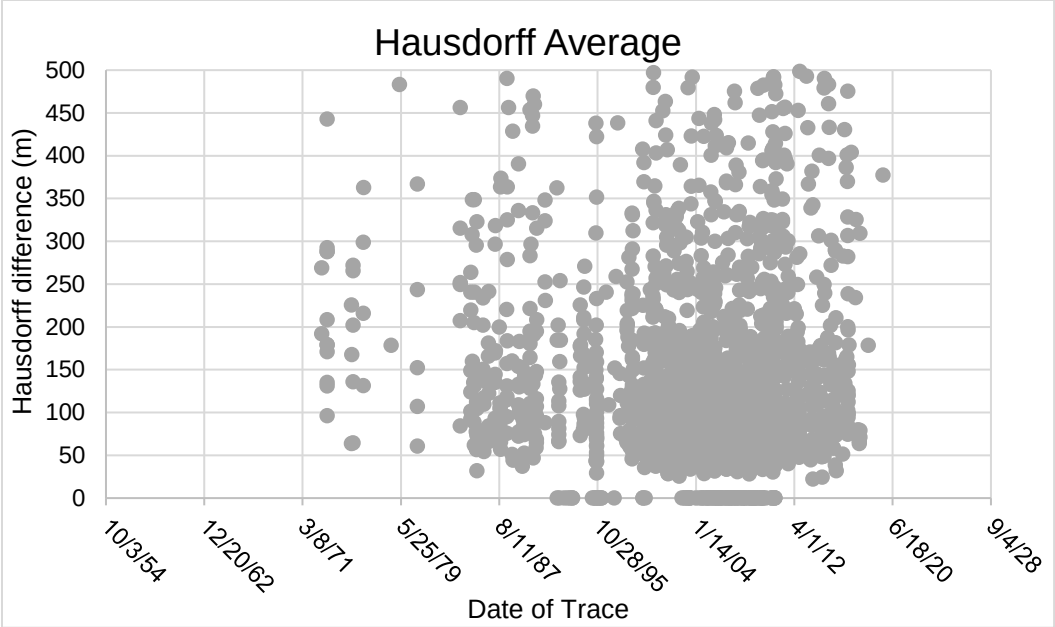


Figure 9: There seems to be a large difference between the authors in this figure in the calculated retreat, but I can not distinguish any difference on the figure due to the thickness of the shapefile. Could the thickness of the shapefiles be reduced to help with this?

- Figure has been updated with new colors, reduced line thickness and opacity to help distinguish the difference.

TermPicks Referee #3

Summary: The authors compiled all publicly-available Greenland marine-terminating outlet glacier positions from a wide variety of authors and performed a rigorous standardization procedure with the aim of creating a terminus trace database that could train machine learning algorithms. A description of qualitative and quantitative differences between the sources is provided, as well as a cursory review of the terminus position data coverage and estimated retreat rates relative to single datasets. The discussion focuses on recommendations for use of these data in machine learning algorithms as well as generation of additional manual terminus trace data using the updated GEEDiT tool (called GEEDiT-TermPicks).

The manuscript is easy to read and documents much-needed work. Although I hope the standardized datasets and the "ideal" approach and output format for the terminus data will advance our field, I am a bit disappointed that this manuscript did not describe any novel insights gained from the combined dataset. I assume that is the topic of another manuscript, but it would have been nice to have this manuscript go a bit beyond a dataset description.

We appreciate the constrictive feedback and positive comments on the manuscript. Based on Reviewer #3's comments, we expanded on the usefulness of the dataset for both scientific and machine learning purposes in the text, primarily by improving figure 8. While we appreciate the desire for additional analysis, the manuscript itself is meant to present a new dataset that will be widely used by the glaciology community to produce new science with estimates of errors and temporal and spatial biases present in terminus traces. Additionally, many results regarding retreat have been published by the original data providers. As we addressed comments, the original line/section numbers of the text may have changed in the final manuscript. The changed text has been noted in the responses to individual comments. Our responses are in blue below each comment.

Major Points:

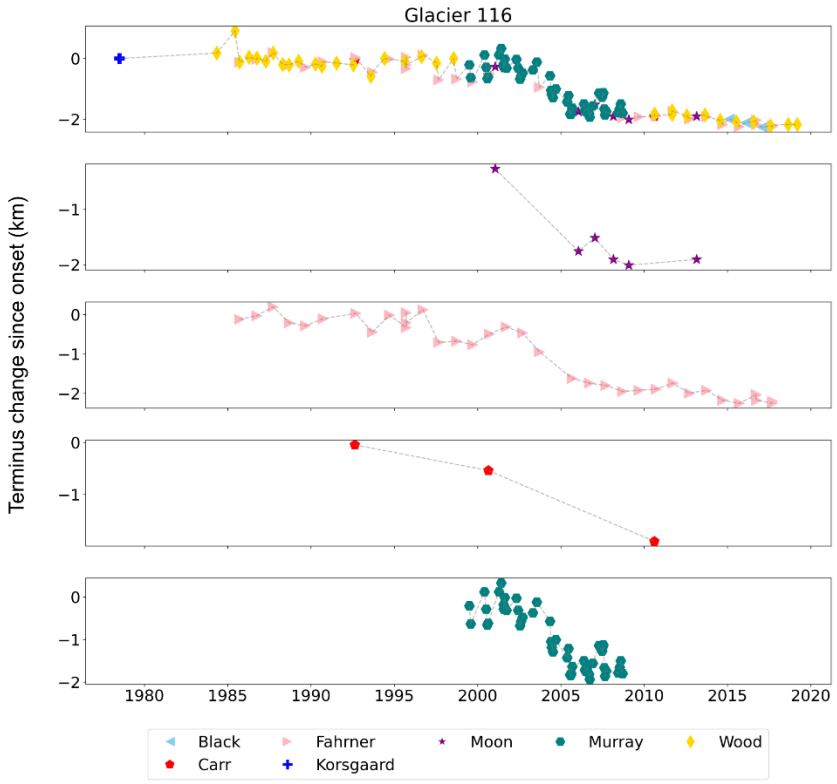
1. I'm not a huge fan of the title. A think there are lots of other applications for this dataset and I think it does the dataset a disservice for the title to suggest it can only be beneficial to machine learning applications. Also, there is no demonstration how the dataset improves machine learning applications (although the authors site machine learning manuscripts focused on glacier change). Instead, I recommend something broader, like "A standardized dataset and workflow for Greenland glacier terminus positions".

- Title changed to “TermPicks: A century of Greenland glacier terminus data for use in scientific and machine learning applications.” While we do not claim that we will improve machine learning itself, the addition of the new training data that includes image IDs will aid in improving the ability of machine learning to identify fronts in times of obstruction due to environmental factors and poor image quality (ice mélange, image saturation in early Landsat, etc.). This was an identified need to improve machine learning application by our co-authors who work on these issues. We agree with the reviewer that this data set will not only be useful for machine learning scientists. In section 2.4, we added the sentence “Including scene IDs is also useful in cases where scientists want to explore other features in the scene at the time of a terminus trace (e.g. iceberg distribution, sediment plume occurrence)” to make this more clear.

2. I appreciate that the results focus on errors and biases for individual traces, but I would also like more information on what the dataset can tell us about changes over time. This does not have to be a Greenland-wide description, but it is important to demonstrate how the combined dataset is much improved over individual datasets. There is one example figure (Figure 8) that is briefly mentioned in the discussion section as an example of the more “complete view of the change” for a glacier. It would be helpful if more examples were given, say as a series of subplots, and that some patterns in retreat rate, magnitude, or timing of changes in those metrics were presented for the broader dataset. Figure 6 gets close to doing this sort of broad overview to demonstrate merit, but doesn’t adequately emphasize the value added by combining the datasets. If these sorts of metrics were presented for some of the contributing datasets as well, I think that information would really emphasize the need for coordination of efforts so that records are detailed in time but also extensive in both space and time. Right now there isn’t anything that demonstrates the broad importance of the dataset you worked hard to create.

    - The authors plan on publishing subsequent papers on the application of the dataset, however the goal of the manuscript is to present a combined dataset with the addition of standardized metadata and image IDs for scientists to easily use these data. One of the largest indicators of the need for coordination is not only the usefulness, but the time it takes to create these datasets. In line 50, we estimate that it took approximately 48 hours per glacier to pick all available images in the Catania and others (2018) paper. Duplication of efforts precludes scientists from working on new questions and the goal of this paper is to reduce that.

- To showcase the datasets merit further, we included subplots of individual author data in addition to the overall TermPicks dataset in figure 8 and compare the magnitude and retreat rate for a subset of authors (Moon, Fahrner, Carr, Murray) in 2000-2010. The retreat magnitude and rates are comparable, the seasonality is only apparent when you include more data points. While the Fahrner data provides a single trace per year and the Carr and Moon data provide <1 trace per year on average to get the long-term magnitude of retreat, the lack of additional traces per year precludes the calculation of seasonality. While the record covers a shorter time, with an average of 6 traces per year for this glacier the Murray data provides enough traces per year to calculate a seasonal signal. The addition of the other authors (Korsgaard, Black, Wood) allows longer term retreat study and analysis of seasonality over the entire record.
- Updates Figure 8:



Author	Start	End	Retreat magnitude (km)	Retreat rate (m/yr)	Seasonality (m)
TermPicks	5/29/2000	9/21/2010	-2.01	-194.8	106
Moon	1/22/2001	1/28/2009	-1.74	-216.9	N/A
Carr	8/24/2000	8/13/2010	-1.365	-136.8	N/A
Fahrner	9/18/2000	9/11/2009	-1.425	-158.6	N/A
Murray	5/29/2000	9/15/2008	-1.92	-231.2	157

3. I'm not sure if this should be swapped in as a main figure or added as a supplemental figure, but I'd like to see heat maps or actual maps of the average temporal resolution and coverage for each glacier. You could potentially use different symbol sizes and colors on an actual map to display those data. Right now the focus is on the number of traces for each glacier, which is important for machine learning, but the temporal resolution and coverage is much more important for someone who would want to analyze these data.
    - Figures A9-11 in the Appendix demonstrate the number of traces per year for each glacier in our dataset. This shows how the temporal distribution of picks varies over each glacier. Additionally, we provide a Google Earth .kmz file in our data submission available on Zenodo that includes a Landsat coverage figure (examples shown in Figure 5) for each glacier so users can see the temporal coverage over the year for each glacier. While this only includes the Landsat data, as 70% of the dataset is Landsat, it provides a good overview of the temporal resolution and coverage for glaciers of interest.
4. In my opinion, the data formatting section should be below the metadata creation section. You mention scene IDs in the metadata creation but that comes after you already describe how you assigned IDs for datasets that did not contain that bit of metadata.
    - The name of the section was changed to "Landsat image scene identifiers" and moved below "Metadata Creation" section for clarity.

Minor Comments:

- Why is the ID flag 005 but all the other flags begin with X?
    - The flag of 05 referenced in section 2.5 Landsat image scene identifiers (formally "data formatting) refers to assigning Landsat IDs to only manually-delineated traces, therefore the prefix (X) of the quality flag will be 0. If it were referring to automatic traces, it would be 1.
- Section 3.3: There needs to be more quantitative substance here. You briefly state that you observe changes in retreat rates. What are the retreat rates? See my major comment about including more of a comparison with the contributing datasets to demonstrate difference.
    - The goal of this paper is to present a dataset that can be used widely by the scientific community. Many previous studies have already published retreat (Murray et al., 2015a; Cowton et al., 2018; Wood et al., 2021) and

retreat rates (Box et al., 2017; King et al., 2020) and controls on retreat (Murray et al., 2015b; Catania et al., 2018; Fried et al., 2018; Slater et al., 2019). The purpose of the retreat section is to provide a check that our dataset does not differ greatly from any of these previous studies. We plan to publish more detailed results with our terminus dataset in upcoming publications.

References:

Box, J. E., & Decker, D. T. (2011). Greenland marine-terminating glacier area changes: 2000–2010. *Annals of Glaciology*, *52*(59), 91-98.

Catania, G. A., Stearns, L. A., Sutherland, D. A., Fried, M. J., Bartholomaus, T. C., Morlighem, M., ... & Nash, J. (2018). Geometric controls on tidewater glacier retreat in central western Greenland. *Journal of Geophysical Research: Earth Surface*, *123*(8), 2024-2038.

Cowton, T. R., Sole, A. J., Nienow, P. W., Slater, D. A., & Christoffersen, P. (2018). Linear response of east Greenland's tidewater glaciers to ocean/atmosphere warming. *Proceedings of the National Academy of Sciences*, *115*(31), 7907-7912.

Fried, M. J., Catania, G. A., Stearns, L. A., Sutherland, D. A., Bartholomaus, T. C., Shroyer, E., & Nash, J. (2018). Reconciling drivers of seasonal terminus advance and retreat at 13 Central West Greenland tidewater glaciers. *Journal of Geophysical Research: Earth Surface*, *123*(7), 1590-1607.

King, M. D., Howat, I. M., Candela, S. G., Noh, M. J., Jeong, S., Noël, B. P., ... & Negrete, A. (2020). Dynamic ice loss from the Greenland Ice Sheet driven by sustained glacier retreat. *Communications Earth & Environment*, *1*(1), 1-7.

Murray, T., Scharrer, K., Selmes, N., Booth, A. D., James, T. D., Bevan, S. L., ... & McGovern, J. (2015a). Extensive retreat of Greenland tidewater glaciers, 2000–2010. *Arctic, antarctic, and alpine research*, *47*(3), 427-447.

Murray, T., Selmes, N., James, T. D., Edwards, S., Martin, I., O'Farrell, T., ... & Baugé, T. (2015b). Dynamics of glacier calving at the ungrounded margin of Helheim Glacier, southeast Greenland. *Journal of Geophysical Research: Earth Surface*, *120*(6), 964-982.

Slater, D. A., Straneo, F., Felikson, D., Little, C. M., Goelzer, H., Fettweis, X., & Holte, J. (2019). Estimating Greenland tidewater glacier retreat driven by submarine melting. *The Cryosphere*, *13*(9), 2489-2509.

Wood, M., Rignot, E., Fenty, I., An, L., Bjørk, A., van den Broeke, M., ... & Zhang, H. (2021). Ocean forcing drives glacier retreat in Greenland. *Science advances*, *7*(1), eaba7282.