

The Cryosphere Discuss., referee comment RC1
<https://doi.org/10.5194/tc-2021-311-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on tc-2021-311

Anonymous Referee #1

Referee comment on "TermPicks: a century of Greenland glacier terminus data for use in scientific and machine learning applications" by Sophie Goliber et al., The Cryosphere Discuss., <https://doi.org/10.5194/tc-2021-311-RC1>, 2021

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 "output" 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.

- l. 10: is this the mean (\pm standard deviation)?
- l. 52: check that months are removed from the reference dates
- l. 104: is the Howat reference here for the MODIS image?
- 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).
- l. 135 (Satellite): How is this formatted/written?
- 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?
- 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.
- l. 170: where do the glacier centerlines come from?
- l. 226: how many of these picks needed manual checking?
- 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?

- Figure 5: I really like this figure.
- 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?