Dear Referee #2,

First of all we want to thank you for your review and for the constructive feedback on our manuscript. Your general comments in particular are very important points that will lead us to a significantly improved manuscript. Our responses for each of the comments raised and how we addressed them are given below. We would like to apologize in advance for the sometimes lengthy responses. But we believe that especially the results from the comparison with other automatic delineation products make a clear argument on the impact of our study. Some of your points were also raised by the other Referees and we thus refer occasionally to those answers.

Again, many thanks for helping to improve our manuscript!

Best wishes,

Erik Loebel and all co-authors

General Comments

This paper uses a deep-learning-based method to produce 9243 calving front positions across 23 Greenland outlet glaciers from 2013 to 2021 and discusses the relationship between terminus variation and basal topography. Overall, I think this paper is well-written and the figures are well-presented. With that being said, I have reservations about its originality, impact, and the extent of its literature review. Based on these concerns, I would recommend rejecting the paper in its current form. Below, I provide a detailed evaluation and rationale for my recommendation.

We thank you for your comments and your suggestions for overcoming the deficiencies of the manuscript.

1. Originality:

In recent years, there increasing number of deep learning-based studies to automate terminus extraction. Compared with the previous study, especially with the author's previous paper Loebel et al. (2022), what is the improvement of this study regarding the methodology?

The study *Loebel et al., 2022* focussed on methodology, namely on the selection of input features and their specific contribution to ANN prediction performance for calving front extraction. The results highlight the benefit of multispectral input features as their integration leads to more accurate predictions compared with conventional single band inputs, especially for challenging ice melange and illumination conditions.

This present study builds on the methodological insights from *Loebel et al. 2022*. We apply multispectral Landsat data to process dense calving front time series for Greenland outlet glaciers. Hence, the ANN architecture is the same as in *Loebel et al., 2022* and is not the focus of this manuscript. Apart from the ANN architecture, the present study uses a different set of methods specifically for processing and analyzing time series.

Overall, four key developments extend upon the previous analysis by *Loebel et al., 2022.* First, we improved our processing by limiting input information to multispectral data and by enlarging our training set. Second, we advanced the accuracy assessment by including the ESA-CCI and CALFIN data additionally to our test set (which now also included images from Svalbard and Patagonia). Thirdly, we apply our processing system to automatically delineate 9243 calving front positions and provide these time series for the cryosphere community. Fourthly, we apply our time series analyzing the interaction of seasonal calving front variation and bedrock topography. Additionally, the revised version will also include a comparison with the other three “big data” repositories (CALFIN, TermPicks and AutoTerm).

We agree that there have been an increasing number of studies advancing deep learning methodology. However, only a limited number of studies (*Cheng et al. 2021, Baumhoer et al. 2023 and Zhang et al. 2023*) exist that apply these methods for time series generation (i.e. applying the method beyond the test dataset), which is ultimately what the cryosphere community requires.

1. Impact:

The main objective of automating the terminus extraction is to produce as many termini as possible. However, compared with the CALFIN (Cheng et al, 2021), which produces 22 678 calving front lines across 66 Greenlandic glaciers from 1972 to 2019, this study seems not improve the temporal resolution, temporal coverage, and spatial coverage. A comparison between the product from this study and CALFIN would be helpful. For instance, which glaciers CALFIN did not cover but this study covers.

We understand this concern. A comparison to other automatic delineation products will not only be highly beneficial for future data users but will also help to emphasize the strengths of our data set and the impact of this study. This is a very good suggestion which has also been raised by the editor and referee #3. The revised version of the manuscript will include a new section presenting and discussing the results of this comparison. In addition to the CALFIN dataset (*Cheng et al. 2021)*, we will also analyze the recently published AutoTerm dataset (*Zhang et al. 2023)* as well as the manually delineated TermPicks repository (*Goliber et al. 2022)*. The results of this comparison are already at hand and will be briefly introduced below.

Firstly, we want to emphasize that a direct comparison of calving front traces is often problematic since the definition of what a single calving front contains varies from study to study. For example, an entry in our data set for the Upernavik Isstrøm consists of four calving front features whereas in CALFIN these calving fronts are listed separately. For Jakobshavn Isbræ, CALFIN considers the north and south branch separately and in our data set it is one calving front. When adding up shape file features, the number of entries in our data product contains 15150 entries – considerably more than the number of 9243 calving fronts quoted in our manuscript. We will raise and discuss this point in the revised version..

Secondly, we want to emphasize that it is beneficial to have multiple data products on glacier front lines. 10 of our 23 glaciers are not included in the CALFIN product. They include the three large glaciers Humboldt Glacier, Nioghalvfjerdsbrae and Zachariæ Isstøm. This adds up to 8092 (from 9243) traces which are not included in CALFIN. Compared to AutoTerm our data product has 3261 unique traces. Compared to TermPicks, our data product has 8217 unique traces. Compared to the combination of CALFIN, AutoTerm and TermPicks, our data product has 2963 unique traces.

To enable a comparison that also considers the different processing methods, we have set a reference period and reference glaciers by considering their temporal (2013 to 2019) and spatial overlap (13 glaciers). Within this reference we looked at mapped fronts, sampling rate and unique entries. Results are given in Table R1. Although having the same Landsat data basis our data product achieves a higher sampling rate and more unique front extractions than CALFIN. This is due to differences in input feature selection and processing. AutoTerm has the most mapped and unique fronts as well as the highest sampling rate. This is due to its data basis which included Landsat, Sentinel-2 and Sentinel-1. When comparing the results for the TermPicks repository it is important to consider that its sampling rate varies significantly across glaciers. As our reference includes numerous glaciers, which have a relatively large number of entries in the TermPicks product, the sampling rate here is likely overestimated. Overall, 372 from 3005 of calving fronts extracted by our method within the reference were not extracted by CALFIN, AutoTerm or TermPicks although all use Landsat-8 imagery.

Table R1: *Tabular comparison of the CALFIN, AutoTerm, TermPicks product as well as the data product presented in this study. The reference period (2013 to 2019) and the reference glaciers (13 glaciers) are defined by the temporal and spatial overlap of the four data products.*

Dataset	Glaciers	Mapped fronts	Time span	Reference period and glaciers		
				Mapped fronts	Sampling rate (yr^{-1})	Unique entries
This study (Loebel et al., 2023)	23	9243	2013-2021	3005	33.02	372
CALFIN (Cheng et al., 2021)	66	22678	1972-2019	1322	14.53	15
AutoTerm (Zhang et al., 2023)	295	278239	1984-2021	7220	79.34	3724
TermPicks (Goliber et al., 2022)	278	39060	1948-2021	2287	25.13	505

We also looked at individual glaciers and compared the data. Figure R1 shows the time series of CALFIN, AutoTerm and TermPicks compared to our study for four examples. Results not only emphasize the different sampling rates, but also highlights the differences in quality control and the associated signal-to-noise ratio. For Kangiata Nunaata Sermia (Fig. R1 (b)) our data product is the only one which captures the seasonal ice tongue (which has been described for example in *Motyka et al., 2017* and *Moyer et al., 2017*).

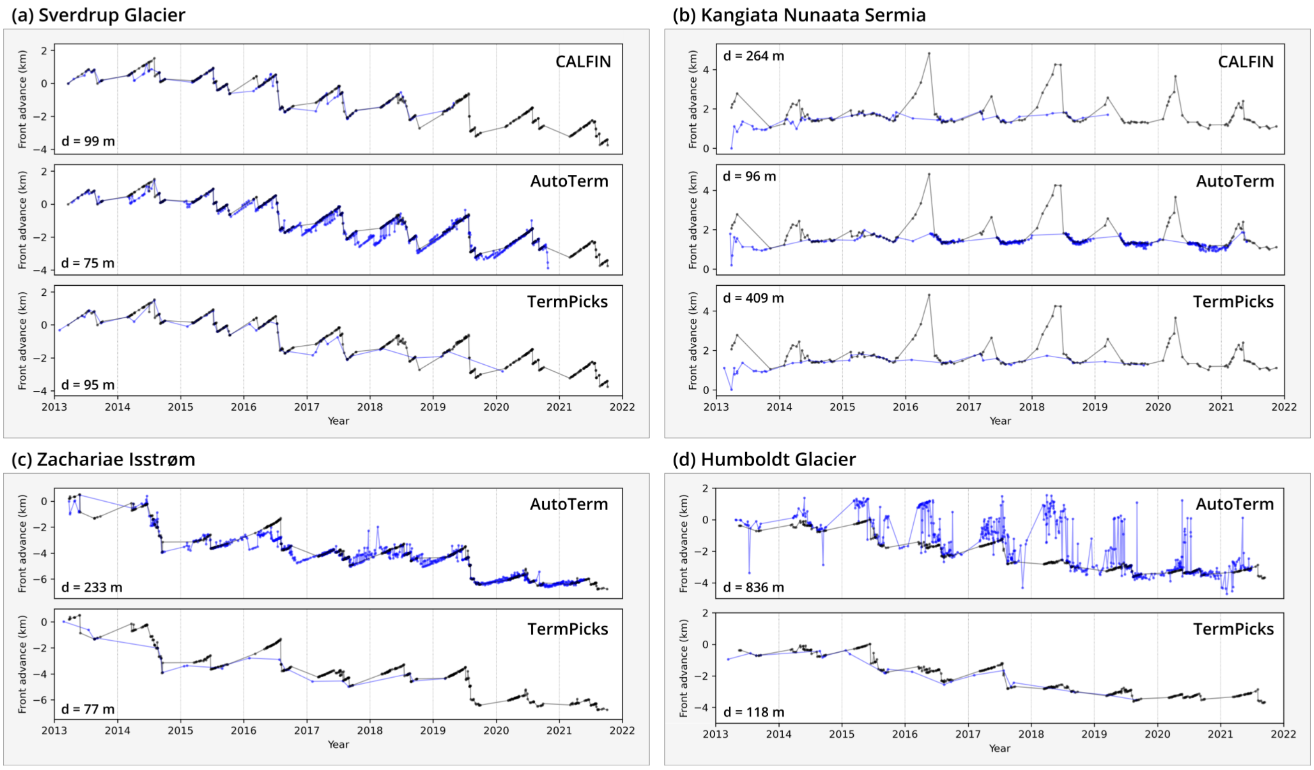

Figure R1: *Comparison of the CALFIN, AutoTerm, TermPicks product (blue) as well as the data product presented in this study (black) for four example glaciers. Time series are derived along the central flow line of the glacier. Every comparison specifies the mean distance d of between calving front delineations at identical days.*

In Figure R1 we also indicate the mean distance, *d*, for same-day calving front acquisitions for each pair of time series. On the basis of these differences, some of which extend well beyond estimated delineation uncertainties, it is clear that there are significant and more importantly unexplained differences between different delineation methods. This is a clear argument in favor of having multiple glacier front data products.

Overall, we expect a high impact of this study for two reasons. Firstly, our method enables a considerably higher extraction rate compared to other methods that use the same data basis (in particular compared CALFIN). This results in a significant amount of calving front traces (13% within our reference) that could not be extracted by the other methods. Importantly, these 13% include extractions under challenging conditions such as the appearance and disintegration of the seasonal ice tongue at Kangiata Nunaata Sermia . Secondly, we see enormous value in having multiple calving front data products even if there is a significant overlap. Glacier area change is a designated Essential Climate Variable (ECV, see https://public.wmo.int) and constitutes a basis for glaciological studies and a new generation of ice dynamic models. All current data products are different. They differ in coverage but also differ for duplicate extractions for identical glacier front traces. A better understanding of these differences is crucial and requires further investigation.

We now see that a comparison to previous products is essential for supporting our case, and we thank the referees for raising this important point. The results, which were only briefly presented here, will be discussed in detail in the relevant section of the revised manuscript.

1. Literature Review:

The discussion about the terminus variation and basal topography is interesting, but studies have been investigating this for many years (Joughin et al., 2008, 2014; Kehrl et al., 2017; Bunce et al., 2018, Catania et al., 2018), suggesting that retrograde bed slopes can cause glacier dynamic instabilities (Meier and Post, 1987) and substantial retreat. However, these papers are missing in the manuscript. Therefore, I suggest the author include a literature review about how the glacier geometry influences the terminus variation, also remove the expression throughout this manuscript that this study is the first application analyzing the interaction between calving front variation and bedrock topography.

We agree with this comment and similar comments of the other referees. With our discussion we wanted to showcase the potential of our data product for future glaciological analyzes. We agree that such a discussion cannot take place without considering previous work.

The discussion (Section 4) will be substantially revised.

1. At the start of the discussion, we will introduce seasonal terminus variability and how it is driven.
2. After that, we will discuss geometric controls of terminus change and how different bedrock topography modulates glacier retreat and advance.
3. Then, our results will be linked to the results of existing studies. For example, *Catania et al. (2018)* show an ongoing retreat of Ingia Isbræ from 2002 until 2016 and suggest further retreat of more than 1 km inland. Our observations confirm this, and extend this analysis by showing that Ingia Isbræ (Figure 8(a)) has stabilized in 2018 due to the prograde slope.
4. We will also emphasize that our time series not only allow to analyze glacier retreat and advance, but also to better differentiate between different calving patterns (e.g. tabular, non-tabular and mixed). Our analysis for Daugaard Jensen (Figure 8 (c)) is a perfect example.
5. Throughout the discussion will refer to relevant studies, including all studies mentioned by you and the other referees.

With the phrase "first application" we wanted to refer to a first application of our data product rather than to imply that we are the first to carry out such an analysis. We apologize for this misunderstanding and we will clear this up by rephrasing these expressions.

Considering the points mentioned above, I believe that rejecting the paper would be appropriate. However, I recommend that the authors be given an opportunity to revise and resubmit their work, addressing the concerns mentioned above. By providing constructive feedback and clear expectations for revision, the authors might have the chance to strengthen their manuscript and overcome the present limitations.

Specific Comments:

Line 51: Why the usage of multispectral sensor information can increase the temporal solution, compared with using the single band? I was asking because using multi-band information could not yield more images.

Using multispectral information does not yield more images, but it does increase the information we have in each image. Different surface types have different reflective curves along the electromagnetic spectrum. This facilitates a better separability between surface types (Kääb et al., 2014). This results in a more robust and accurate image segmentation using ANN which leads to a higher success rate of our method and an increased temporal resolution compared to using single band inputs (Loebel et al., 2022).

The use of multispectral bands is one reason that for the above mentioned reference period our method yields a significantly higher sampling rate compared to CALFIN.

Line 53: I believe the CALFIN can also resolve sub-seasonal terminus variations as it uses all the available Landsat images.

With this sentence we wanted to highlight the ten glaciers which were not included in CALFIN. But we acknowledge that these formulations (also at L166-168 and L173) are problematic, the more so after the recent release of the AutoTerm product. This concern has also been raised by the other referees. The formulations will be reframed

Line 64: What modification did the author apply here? Does that modification improve the results? It would be better to have a more detailed description of this.

As a main modification, the present study only uses multispectral information and no textural and topographic features. This reduces the input layer from 17 to 9 layers. Furthermore we increased our TUD reference data set from 728 to 898 calving fronts. These new calving front traces focus specifically on cloudy, low illumination and scene border conditions, enhancing the method in this regard.

We appreciate this comment, and we will expand the revised manuscript with a more detailed description.

Section 2.2: It would be better if the author can change the name of this section (validation) to avoid confusion. Following the deep learning convention, there are three sets: the training set, the validation set, and the test set. This section is actually about the test set but is called validation.

Also, the descriptions of the validation set should be included in the manuscript.

We will follow this recommendation and rename section 2.2 to *accuracy assessment*. Furthermore we will follow the suggested convention and throughout the manuscript use the naming: training set, validation set and test set.

A short description of the validation set (in the article *internal validation set*) is already in appendix B and will be moved to section 2 in the main manuscript.

Figure 8: It would be more helpful to change the second column to the time series of the bed elevation at the terminus. Time series might better reflect the verbal descriptions in the discussion section.

After looking at the time series of bed elevation at the terminus position, we prefer the current display of profile distance and bedrock depth. Mainly because this lateral view of the glacier and its bedrock topography is very easy and fast to comprehend. In addition, this provides further context regarding the surrounding bedrock topography (for example that Kangerdlugssuaq Glacier is about to retreat into a steep retrograde topography).

Nonetheless, we agree that it is very difficult to assign the time information referred to in the text to a particular bedrock position using only the color coded lines. To resolve this problem we modified the third column of the figure which now includes the bedrock slope at the specific calving front positions in respect to the date. This makes it much easier to identify the bedrock slope at the calving front for a given date. The descriptions in the discussion are now easier to follow.

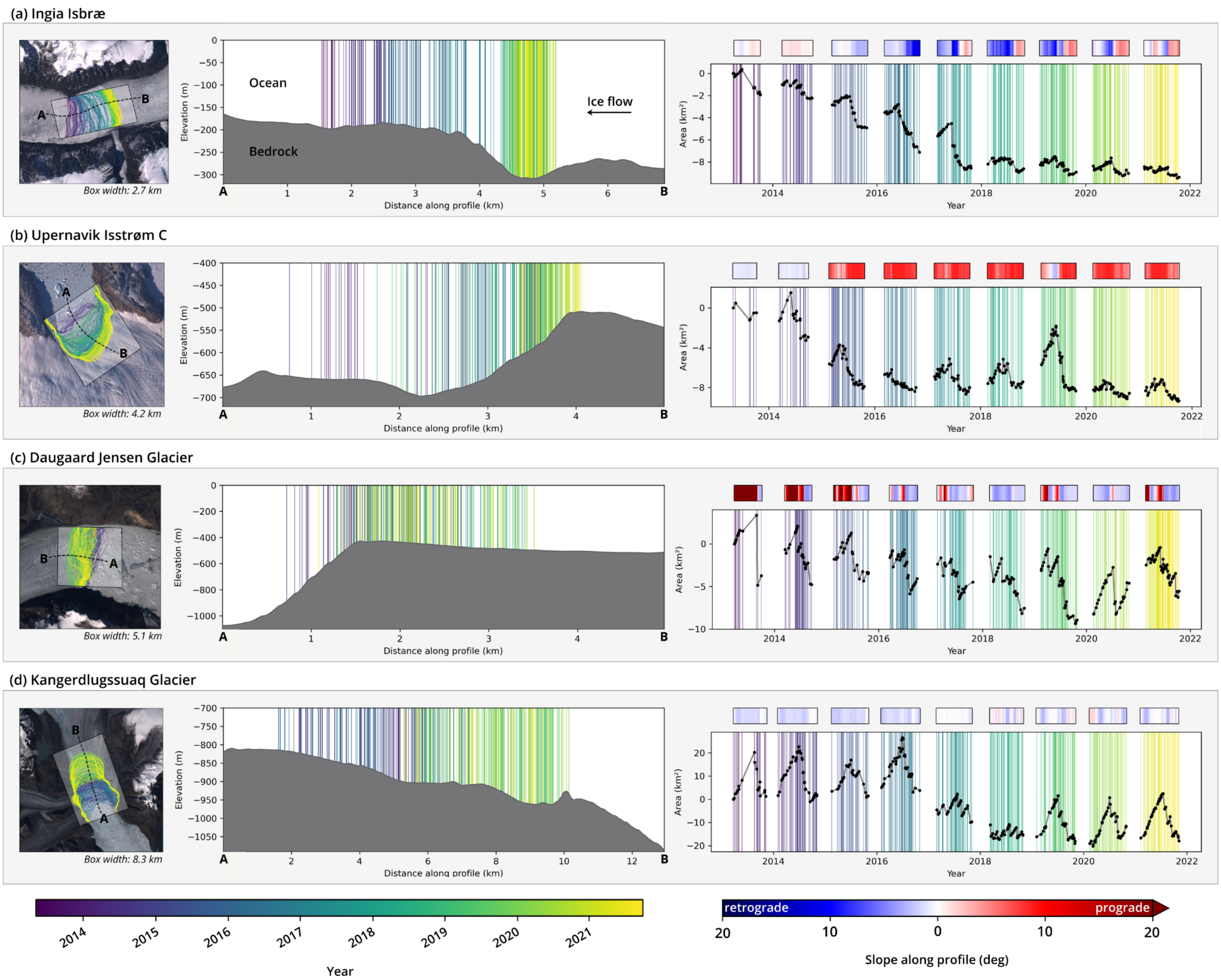

__Figure R2:__ Modified Figure 8 in the manuscript. This figure now includes the bedrock slope at the calving front for a given date. This is shown as a color coded bar in the third column above the terminus area change time series.

Reference

Loebel, E., Scheinert, M., Horwath, M., Heidler, K., Christmann, J., Phan, L. D., ... & Zhu, X. X. (2022). Extracting Glacier Calving Fronts by Deep Learning: The Benefit of Multispectral, Topographic, and Textural Input Features. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1-12.

Cheng, D., Hayes, W., Larour, E., Mohajerani, Y., Wood, M., Velicogna, I., & Rignot, E. (2021). Calving Front Machine (CALFIN): glacial termini dataset and automated deep learning extraction method for Greenland, 1972–2019. The Cryosphere, 15(3), 1663-1675.

Joughin, I., Howat, I., Alley, R. B., Ekstrom, G., Fahnestock, M., Moon, T., Nettles, M., Truffer, M., and Tsai, V. C.: Ice front variation and tidewater behavior on Helheim and Kangerdlugssuaq Glaciers, Greenland, J. Geophys. Res.-Earth, 113, F01004, https://doi.org/10.1029/2007JF000837, 2008.

Kehrl, L. M., Joughin, I., Shean, D. E., Floricioiu, D., and Krieger, L.: Seasonal and interannual variabilities in terminus position, glacier velocity, and surface elevation at Helheim and Kangerlussuaq Glaciers from 2008 to 2016, J. Geophys. Res.-Earth, 122, 1635–1652, https://doi.org/10.1002/2016JF004133, 2017.

Catania, G. A., Stearns, L. A., Sutherland, D. A., Fried, M. J., Bartholomaus, T. C., Morlighem, M., Shroyer, E., and Nash, J.: Geometric Controls on Tidewater Glacier Retreat in Central Western Greenland, J. Geophys. Res.-Earth, 123, 2024–2038, https://doi.org/10.1029/2017JF004499, 2018.

Bunce, C., Carr, J. R., Nienow, P. W., Ross, N., and Killick, R.: Ice front change of marine-terminating outlet glaciers in northwest and southeast Greenland during the 21st century, J. Glaciol., 64, 523–535, https://doi.org/10.1017/jog.2018.44, 2018.

Meier, M. F., & Post, A. (1987). Fast tidewater glaciers. *Journal of Geophysical Research: Solid Earth*, *92*(B9), 9051-9058.

References

Zhang, E., Catania, G., & Trugman, D. T. (2023). AutoTerm: an automated pipeline for glacier terminus extraction using machine learning and a “big data” repository of Greenland glacier termini. *The Cryosphere*, 17(8), 3485-3503.

Goliber, S., Black, T., Catania, G., Lea, J. M., Olsen, H., Cheng, D., Bevan, S., Bjørk, 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 (2022). TermPicks: a century of Greenland glacier terminus data for use in scientific and machine learning applications. *The Cryosphere*, *16*(8), 3215-3233.

Motyka, R. J., Cassotto, R., Truffer, M., Kjeldsen, K. K., Van As, D., Korsgaard, N. J., ... & Rysgaard, S. (2017). Asynchronous behavior of outlet glaciers feeding Godthåbsfjord (Nuup Kangerlua) and the triggering of Narsap Sermia's retreat in SW Greenland. *Journal of Glaciology*, *63*(238), 288-308.

Moyer, A. N., Nienow, P. W., Gourmelen, N., Sole, A. J., & Slater, D. A. (2017). Estimating spring terminus submarine melt rates at a Greenlandic tidewater glacier using satellite imagery. *Frontiers in Earth Science*, *5*, 107.

Kääb, A., Bolch, T., Casey, K., Heid, T., Kargel, J.S., Leonard, G.J., Paul, F. and Raup, B.H. (2014). Glacier mapping and monitoring using multispectral data. *Global land ice measurements from space*, 75-112.