CALFIN Subseasonal Greenland Glacial Terminus Positions, Version 1

USER GUIDE

How to Cite These Data

As a condition of using these data, you must include a citation:

Cheng, D. L., W. Hayes., and E. Larour. 2021. *CALFIN Subseasonal Greenland Glacial Terminus Positions, Version 1*. [Indicate subset used]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/7FILV218JZA2. [Date Accessed].

FOR QUESTIONS ABOUT THESE DATA, CONTACT NSIDC@NSIDC.ORG

FOR CURRENT INFORMATION, VISIT https://nsidc.org/data/NSIDC-0764

National Snow and Ice Data Center

TABLE OF CONTENTS

1 DATA DESCRIPTION

This data set contains shapefiles of Greenland's glacial termini and basins for the years 1972 to 2019. These vector data were created from Landsat 1-8 satellite imagery using the Calving Front Machine (CALFIN) an automated processing workflow utilizing neural networks for extracting calving fronts from satellite images of marine-terminating glaciers.

1.1 Parameters

- Glacial terminus positions

Shapefiles come in two types: polylines and polygons. The polyline product consists of isolated, georeferenced calving fronts. The polygon product consists of an ocean mask clipped to domains of interest, the fjord boundaries, and the calving front(s), for each domain (Cheng et. al., 2021). Domain subsets were created to clip features to specific regions, but are not included in this data set.

Caution must be exercised when using this data product for evaluating glacier width and ice area lost. Direct measurement of area loss is cautioned against as irregular time gaps between observations may result in missed calving events.

Digitized lines are intended to cover the full width of the glacier terminus, but may end prematurely or continue along the fjord boundaries beyond the terminus-fjord junction.

Digitized polygons are intended to address the limitation of digitizing the lines that represent the glacial terminus or calving fronts by connecting them to fjord boundaries. However, this may result in errors inherited from the lines.

1.2 File Information

1.2.1 Format

Data are available as shapefiles.

1.2.2 File Contents

The data set consists of a polyline and a polygon shapefile for each of the glaciers studied. In addition, the polylines and polygons for all the glaciers have been compiled into two shapefiles that contain all of the glaciers studied in Greenland. The number of lines and polygons may differ for the same glacier. One reason for this is that a single glacier potentially has multiple branches which are provided as independent lines and then merged into a single polygon.

For a list of the glaciers and a table of temporal coverage, see Figure 1 of Cheng et al. (2021) and Figure S1 of Cheng et al. Supplement (2021).

Note: NSIDC provides an updated set of shapefiles in which the data for Akullersuup Sermia and Kangiata Nunaata Sermia were combined. Thus, the files provided by NSIDC may differ from the descriptions in related publications.

The following table describes the various data fields in the shapefiles.

Table 1. Data Field Descriptions and Examples[1]

Data Field	Description	Format (Values)
GlacierID	Numerical ID assigned to each glacier	# ([1, 246]), Numbers 1-238 apply to by this data set and data from NSIDC-0642. Numbers after 238 are unique to this data set.
Center_X	Mean X coordinate in EPSG:3413	# ([-463626, 682313])
Center_Y	Mean Y coordinate in EPSG:3413.	# ([-2821269, -906747])
Latitude	Latitude of center	# ([64.29, 81.24])
Longitude	Longitude of center	# ([-63.17, -28.21])
QualFlag	Quality flag to indicate digitization conditions	# (0 – Manually digitized, 3– Manually digitized, w/ L7 SCE, 10 - Automatically digitized, 13 – Automatically digitized, w/ L7 SCE. "SCE" represents "Scan Line Error".
Satellite[2]	Satellite/sensor of the digitized source image	LXSS where L represents Landsat, X, represents the sensor, and SS represents the two-digit number assigned to the satellite For Example: LE07 represents Landsat ETM+ satellite 07. Possible values for "X", the sensor are: "C"=OLI/TIRS combined, "O"=OLI-only, "T"=TIRS-only, "E"=ETM+, "T"="TM, "M"=MSS
Date	Date of the digitized source image	YYYY-MM-DD ([1972-09-06, 2019-06-25])
ImageID[2]	Source image file name.	LXSS_LLLL_PPPRRR_YYYYMMDD_yyyymmdd_CC_TX (LC08_L1TP_026006_20170702_20170715_01_T1, etc.) Where LXSS is the Satellite and Sensor, LLLL represents the processing correction level, PPP represents the WRS path, RRR represents the WRS row, YYYYMMDD represents the processing year, month day, CC represents the collection number, and TX represents the TX collection category ("RT = Real-Time, "T1" = Tier 1, and "T2" equals Tier 2)
GrnlndcN	Greenlandic glacier name	NAME (New_Greenl names from Bjørk et al. 2015 database of Greenland glacier names)
OfficialN	Officially recognized glacier name	NAME (Official_n names from Bjørk et al. 2015 database of Greenland glacier names)

Data Field	Description	Format (Values)
AltName	Alternative, Foreign, Old Greenlandic, or other glacier names	NAME (Foreign_na, Old_Greenl, Alternative names (Bjørk et al. 2015), or other names)
RefName	Reference glacier name, non-authoritative names used in CALFIN to denote grouped/unnamed glaciers	NAME (New_Greenl, Official_n, Foreign_na, Old_Greenl, Alternative names (Bjørk et al. 2015), or other names)
Author	Digitization author's name	LastName_FirstInitial (Cheng_D)

[1]The data fields for shapefiles in this data set share a feature schema derived from NSIDC-0642, including information about the date, IDs, name(s), and quality flag. See, MEaSUREs Annual Greenland Outlet Glacier Terminus Positions from SAR Mosaics, Version 1. (NSIDC-0642).

[2]For details, please refer to the USGS Landsat Missions page regarding naming conventions for Landsat Collections Level-1 scenes.

1.2.3 File Naming Convention

The following table provides examples along with a description of the variables used for file naming.

Sample file names:

- Polyline file: `termini_1972-2019_Kjer-Gletsjer_lines_v1.0.shp`
- Polygon file: `termini_1972-2019_Kjer-Gletsjer_polygons_v1.0.shp`

Table 2. File Name Variable Descriptions

Variable	Description
termini	Project name
1972-2019	Temporal coverage of data set
Kjer-Gletsjer	Glacier or geographic names based upon Bjork et al. (2015)
lines or polygons	Type of shapefile
v1.0	Data set version

Variable	Description
.ext	Shapefile extensions: .cpg – specified the codepage for identifying the characterset used .dbf – dBASE table that stores the attribute information of features. .prj – file that stores the coordinate system information .shp – main file that stores the feature geometry .shx – index file that stores the index of the feature geometry

1.3 Spatial Information

1.3.1 Coverage

Greenland:

Latitude: 60°N to 80N
Longitude: 15°W to 75°W

For a list and map of the glaciers studied, see Figure 1 of Cheng et al. (2021).

1.3.2 Resolution

Approximately 30 m with an accuracy of less than 90 m.

1.3.3 Geolocation

The following table provides geolocation information for this data set.

Table 3. WGS 84 / NSIDC Sea Ice Polar Stereographic North

Geographic coordinate system	WGS 84
Projected coordinate system	NSIDC Sea Ice Polar Stereographic North
Longitude of true origin	45° W
Latitude of true origin	70° N
Scale factor at longitude of true origin	1
Datum	WGS 1984
Ellipsoid/spheroid	WGS 84
Units	meter
False easting	0
False northing	0
EPSG code	3413

PROJ4 string	proj4.defs("EPSG:3413","+proj=stere +lat_0=90 +lat_ts=70 +lon_0=-45 +k=1 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no_defs");
Reference	https://epsg.io/3413

1.4 Temporal Information

1.4.1 Coverage

Start Date: 06 September 1972

End Date: 25 June 2019

A detailed list of temporal availability for this data set is provided in Cheng et al. Supplement (2021).

1.4.2 Resolution

Sub-seasonal

2 DATA ACQUISITION AND PROCESSING

2.1 Background

The Calving Front Machine (CALFIN) provides an automated approach to extracting calving fronts from satellite images of marine-terminating glaciers. Understanding tidewater glacier evolution, and the evolution of ice sheets in general and specifically with respect to Greenland, reduces gaps and uncertainty when modeling past and projected climate change. Current, time-consuming manual methods to delineate calving fronts underutilize available satellite imagery. CALFIN leverages machine learning and deep neural networks to provide an automated approach to the challenges of delineating calving fronts from the ever-increasing volume of raster data. The CALFIN is based upon a modified version of the DeepLabV3+ Xception neural network. For a detailed explanation of CALFIN, see Cheng et al. (2021).

2.2 Acquisition

The Shapefiles were produced from satellite images acquired by the following satellites and sensors: Landsat 1-5 MSS, Landsat 4-5 TM, Landsat 7 ETM+, and Landsat 8 OLI. Additionally, a few images from Landsat Level-1 products (L1GS/L1GT) were used to fill in Landsat 1 and 2 time series gaps between 1972 to 1985.

The Landsat Product Generation System attempts to process all Landsat scenes to level L1TP: Precision and Terrain Correction. However, scene and/or sensor issues, or insufficient reference data, can cause L1TP processing to fail.

Systematic Terrain Correction (L1GT) products are created when the systematic product lacks sufficient ground point control for L1TP, but has consistent and sufficient locational accuracy to apply a terrain model.

Systematic Correction (L1GS) products are created when the locational accuracy is not sufficient to apply terrain correction. For additional details on Landsat and Landsat products, please see the USGS Landsat Missions page.

TerraSAR-X and Sentinel-1A/B synthetic-aperture radar (SAR) images were added to the training and validation data to ensure the applicability of CALFIN across different sensors and domains. As such, the trained and validated area of interest for this method includes Antarctic SAR data as well as Greenlandic Landsat optical data.

2.3 Processing

CALFIN processes raw image data into calving front shapefiles in three stages: preprocessing, neural network processing, and postprocessing (see Figure 1). The following sections summarize these stages. For more detail, see Cheng et al. (2021).

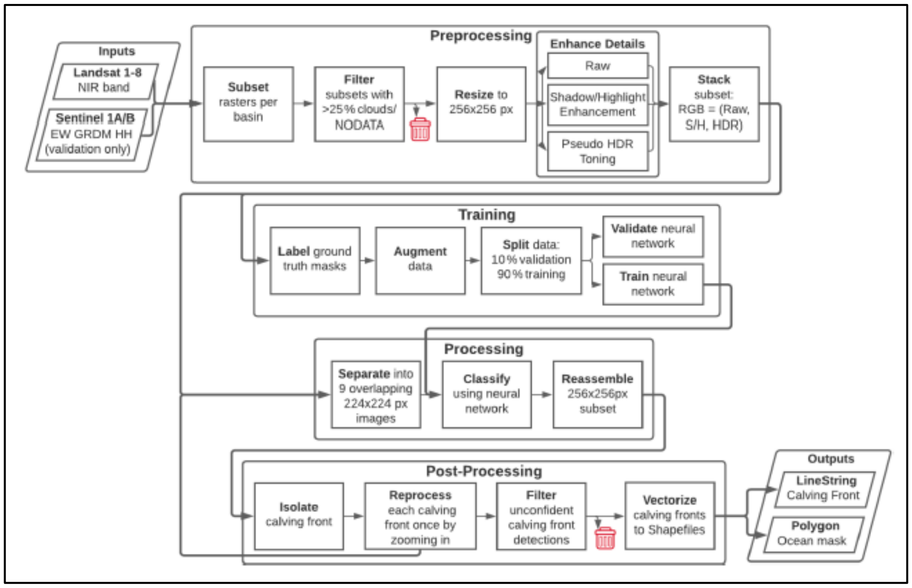

Figure 1. Methodology flowchart. The CALFIN workflow, which processes single band raster imagery into calving front and ocean mask shapefiles. Note that Sentinel-1A/B imagery is only used for validation, as it is not corrected and thus not qualified for geolocation/extraction. Image courtesy of Cheng et. al., 2021.

2.3.1 Preprocessing

Raster images were selected from areas centered around one the following nine glacial basins: Kong Oscar, Hayes, Rink Isbrae, Upernavik, Jakobshavn, Kangiata Nunaata Sermia, Helheim, Kangerlussuaq, and Petermann. Next, the L1TP rasters from Landsat 1–8 with < 20% cloud coverage were collected and supplemented with manually georeferenced L1GS/L1GT products (see 2.2 Acquisition) to fill in Landsat 1–2 time series gaps. This resulted in a total of 4956 Landsat rasters.

Next, the rasters were: clipped using basin domain shapefiles which enclosed the glacier terminus; and filtered to remove any remaining subsets that still contained ≥ 30% "NODATA" pixels or ≥ 20% cloud pixels in the Landsat QA band. The remaining image subsets were resized to 256 x 256 pixels and enhanced using Adobe Photoshop Pseudo-HDR Toning and Shadows/Highlights. Finally, the HDR- and Shadows/Highlights-enhanced subsets were stacked into a single RGB image. Refer to Figure 2 for a visual representation of the preprocessing steps.

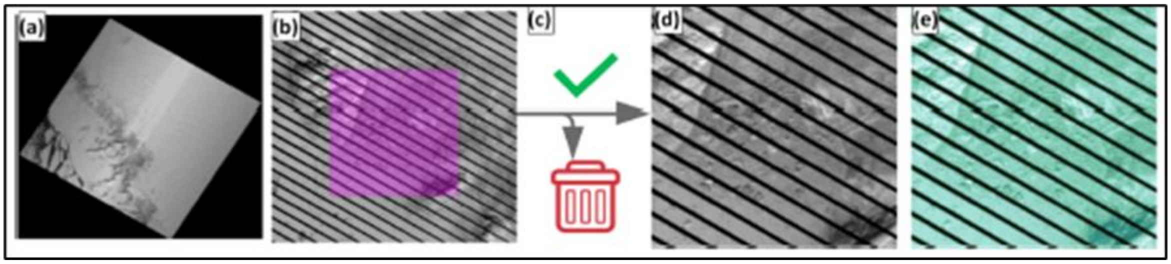

Figure 2. Preprocessing Pipeline: (a) First, input the raw Landsat GeoTIFF rasters with <20% clouds. (b) Next, subset using QGIS/GDAL and the domain shapefile to clip each raster. (c) Then, filter the clouded/NODATA subsets. (d) Now, resize the subsets to 256x256 px. (e) Final, enhance and stack with the raw data subset. Image courtesy of Cheng et. al., (2021).

2.3.2 Neural Network Processing

The CALFIN neural network is trained using manually delineated calving front masks. Once trained, neural network processing extracts basic elements, like edges and corners, and assembles them into more abstract features, such as glacier and land textures. Then, a decoder up-samples these features to predict the final outputs: a probability mask that shows each pixel's likelihood of lying on the coastline or calving front, plus a secondary ice-ocean probability mask.

CALFIN's neural network architecture, training set, and techniques used to improve accuracy and performance are detailed in "Section 3.2 **|** Neural Network Processing" of Cheng et. al., 2021.

2.3.3 Post-Processing

The two-channel pixel mask output is post-processed to extract the shapefile products. A broad overview of the post-processing pipeline is show in Figure 3 In more detail, first, a polyline is fit to the coastline/calving front pixel mask by: converting each pixel in the mask to nodes in a graph; connecting the nearest neighboring nodes; and finding the single longest path in the graph's minimum spanning tree, see Figure 4 for a visual representation of these steps.

Next, the calving front is isolated from the coastline polyline using static masks of the average fjord boundaries, created manually for each basin using the image subsets and bed topography from BedMachine V3 as a reference (see Morlighem et al., 2017). Calving fronts are located by calculating the distance from each point in the coastline to the nearest fjord boundary pixel and then selecting the contiguous pixels which are the farthest from the fjord boundaries. See Figure 4 step e for the results of these steps.

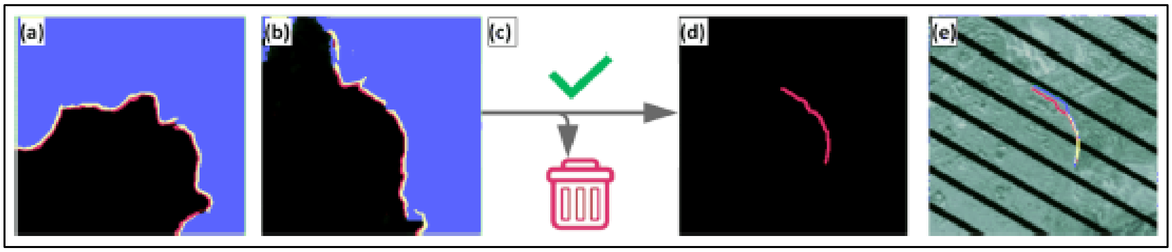

Figure 3. Post-processing pipeline: (a) First, get the processed image from CALFIN (b) Then, isolate and re-process each front. (c) Next, filter unconfident predictions. (d) Now, fit line and mask static coastline (see also Figure 3). (e) Lastly, export and validate Shapefile. Image courtesy of Cheng et al. (2021).

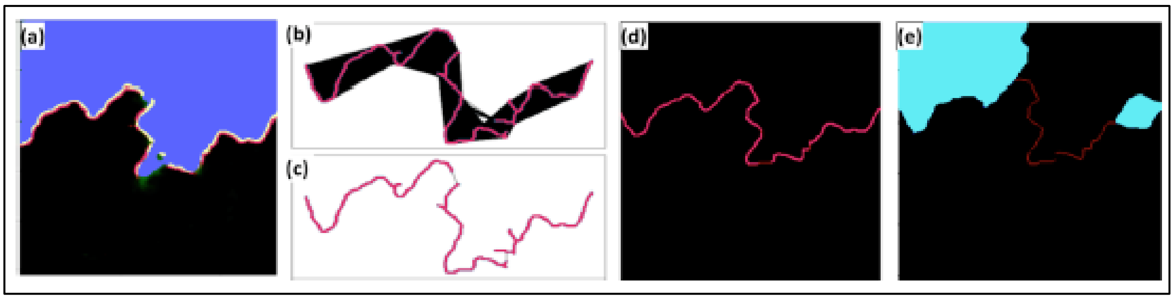

Figure 4. Mask to polyline algorithm. (a) First, extract the coastline mask (red/yellow) from the CALFIN output. (b) Then create a graph, connecting each pixel (red) to 15% of its nearest neighbors with an edge (black). (c) Next, create a Minimum Spanning Tree (MST) from the graph. (d) Now, extract the longest path from the MST. (e) Finally, mask the static coastline using the fjord boundaries (cyan) to extract the calving front. Image courtesy of Cheng et al. (2021).

Once isolated, each calving front's bounding box is used to extract a higher-resolution subset from the original and reprocessed images. Fronts are then filtered using the ice-ocean probability mask to remove uncertain detections.

Finally, the polylines are smoothed to eliminate noise artifacts inherited from previous steps, and the smoothed polylines, fjord boundary mask, and probability masks are combined to create a polygonal ocean mask. The polylines and their corresponding polygons are then exported as georeferenced shapefiles. Refer to "3.3 | Post-processing" in Cheng et al. (2021) for complete details.

2.4 Quality, Errors, and Limitations

CALFIN produces results which deviate by on average 86.76 +/- 1.43 m from the measured front (Cheng et al., 2021).

The primary quality assessment method, error estimation, utilizes the mean-median distances between closest pixels in predicted and manually delineated calving fronts. Also referred to in literature as the “area over front,” this method may be seen as a generalization of the method of transects along arbitrarily oriented fronts (Mohajerani et al., 2019; Baumhoer et al., 2019).

Classification accuracy, a second quality assessment method, compares the degree of overlap between predicted and manually delineated masks of the calving front using the “intersection over union” approach from Baumhoer et al. (2019). It is calculated by dividing the number of pixels in the intersection of two masks by the number of pixels in the union of the two masks. For a visual representation of error methods, see Figure 5.

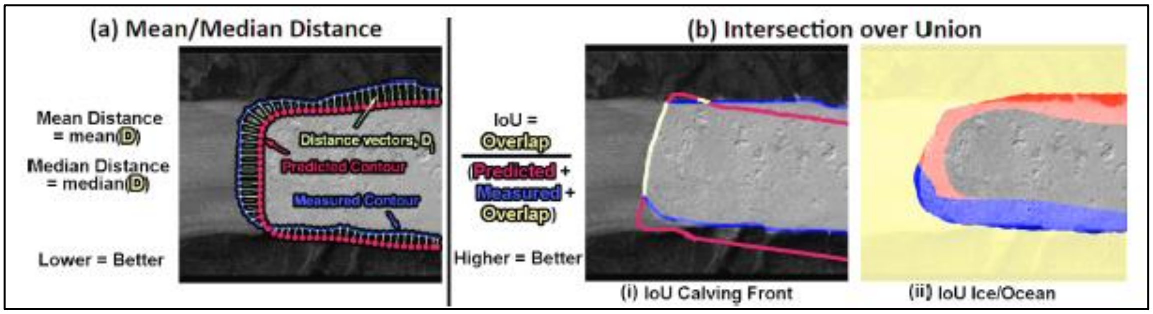

Figure 5. Error Measures: (a) A visual outline of mean-median distance error estimation and (b) classification accuracy using intersection over union (IoU) for the (i) primary calving front, and (ii) the secondary ice/ocean mask, respectively. Image courtesy of Cheng et al. (2021).

CALFIN results are evaluated against several validation sets taken from existing studies. They are: 10 Landsat Helheim subsets used in Mohajerani et al. (2019), the six TerraSAR X Jakobshavn subsets used in Zhang et al. (2019), and 62 Sentinel-1 Antarctic basins taken from the 11 validation scenes used in Baumhoer et al. (2019).

In tests, CALFIN automatically filters out images in which it does not detect calving fronts, confirmed by 13 images in the CALFIN validation set which did not contain calving fronts discernible to the human eye.

CALFIN does not output any false positives on the CALFIN validation set. While this ensures accurate fronts are output rather than incorrect fronts, this filtering behavior removes potentially large errors and must be accounted for when comparing errors across other sets. For a detailed discussion of error analysis and quality assessment see Cheng et al. (2021).

3 SOFTWARE AND TOOLS

Users can access shapefiles with GIS software such as QGIS or ArcGIS.

4 RELATED DATA SETS

Alaska Tidewater Glacier Terminus Positions

MEaSUREs Annual Greenland Outlet Glacier Terminus Positions from SAR Mosaics

MEaSUREs Greenland Annual Ice Sheet Velocity Mosaics from SAR and Landsat

5 CONTACTS AND ACKNOWLEDGMENTS

Daniel L. Cheng and Wayne Hayes
University of California Irvine
Department of Computer Science

Eric Larour
Jet Propulsion Laboratory
Pasadena, CA

6 REFERENCES

Baumhoer, C. A., Dietz, A. J., Kneisel, C., and Kuenzer, C. (2019). Automated Extraction of Antarctic Glacier and Ice Shelf Fronts from Sentinel-1 Imagery Using Deep Learning, *Remote Sensing*, 11, 2529.

Bjørk, A. A., Kruse, L. M., and Michaelsen, P. B. (2015). Brief communication: Getting Greenland's glaciers right – a new data set of all official Greenlandic glacier names, *The Cryosphere*, 9, 2215–2218, https://doi.org/10.5194/tc-9-2215-2015.

Chen, L., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, *European Conference on Computer Vision*, pp. 801–818.

Cheng, D., Hayes, W., Larour, E., Mohajerani, Y., Wood, M., Velicogna, I., and Rignot, E. (2021) Calving Front Machine (CALFIN): glacial termini dataset and automated deep learning extraction method for Greenland, 1972–2019, *The Cryosphere*, 15, 1663–1675. https://doi.org/10.5194/tc-15-1663-2021

Cheng, D., Hayes, W., Larour, E., Mohajerani, Y., Wood, M., Velicogna, I., and Rignot, E. (2021). Supplement of Calving Front Machine (CALFIN): Glacial Termini Dataset and Automated Deep Learning Extraction Method for Greenland, 1972–2019, *The Cryosphere*. https://doi.org/10.5194/tc-15-1663-2021-supplement

Mohajerani, Y., Wood, M., Velicogna, I., and Rignot, E. (2019). Detection of Glacier Calving Margins with Convolutional Neural Networks: A Case Study, *Remote Sensing*, 11.

Zhang, E., Liu, L., and Huang, L. (2019). Automatically delineating the calving front of Jakobshavn Isbræ from multi-temporal TerraSAR-X images: a deep learning approach, *The Cryosphere* 1–20.

7 DOCUMENT INFORMATION

7.1 Publication Date

17 September 2021