Summary: This paper entitled "Brief communication: Rapid machine learning-based extraction and measurement of ice wedge polygons in airborne lidar data" presents a workflow for rapid delineation and microtopographic characterization of ice wedge polygons within high-resolution digital elevation models. The authors have been addressed all my concerns in their revised version. The paper is well conducted and presented. In summary, I think this paper is publishable in The Cryosphere. Please consider improving two places before final publishing.

Minor comments:
1. Title: Is it necessary to include "brief communication" in the title? The authors creatively proposed a new way to rapidly delineate and microtopographic characterize ice wedge polygons rather than summarizing previous work. Please remove "brief communication" if possible.

We originally submitted the manuscript as a "brief communication" as it was somewhat shorter than most regular manuscripts. We are comfortable with either changing the status of the manuscript or keeping it a brief communication, depending on the opinion of the editor.

2. Abstract: Please consider including some validation result with numbers in the abstract.

We now include a statement in the abstract that manual validations indicate 70-96% accuracy by area at the kilometer scale.

Brief communication: Rapid machine learning-based extraction and measurement of ice wedge polygons in high-resolution digital elevation models~~airborne lidar data~~

Charles J. Abolt[1,2], Michael H. Young[2], Adam L~~A~~. Atchley[3], Cathy J. Wilson[3]

[1]Department of Geological Sciences, The University of Texas at Austin, Austin, TX, USA
[2]Bureau of Economic Geology, The University of Texas at Austin, Austin, TX USA
[3]Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM, USA

Correspondence to: Charles J. Abolt (chuck.abolt@beg.utexas.edu)

Abstract. We present a workflow for rapid delineation and microtopographic characterization of ice wedge polygons within high-resolution digital elevation models. At the core of the workflow is a convolutional neural network used to detect pixels representing polygon boundaries. ~~The workflow, which is extensible to other forms of remotely sensed imagery, incorporates a convolutional neural network to detect pixels representing troughs.~~ A watershed transformation is ~~then~~ subsequently used to segment imagery into discrete polygons. Fast training times (<5 minutes) permit an iterative approach to improving~~e~~ s~~s~~kill ~~improvement~~ as the routine is applied across broad landscapes. ~~Regions of non-polygonal terrain are partitioned out using a simple post-processing procedure.~~ Results from ~~training and validation~~study sites near~~at~~ Barrow and Prudhoe Bay, Alaska demonstrate robust performance in diverse tundra ~~landscapes~~settings, with manual validations demonstrating 70-96% accuracy by area at the kilometer scale. The methodology permits ~~fast~~precise, spatially extensive measurements of polygonal microtopography and trough network geometry.

1 Introduction and Background

This research ~~paper~~ addresses the problem of delineating and measuring ice wedge polygons within high-resolution digital elevation models (DEMs). ~~The objective of this research is to develop and report on a workflow for rapid delineation and microtopographic analysis of ice wedge polygons in airborne lidar data.~~ Ice wedge polygons are the surface expression of ice wedges, a form of ground ice nearly ubiquitous to coastal tundra environments in North America and Eurasia (Leffingwell, 1915; Lachenbruch, 1962). High resolution inventories of ice wedge polygon microtopography are of hydrologic and ecologic interest, because decimeter~~e~~-scale variability in polygonal relief can drive pronounced changes to soil drainage (Liljedahl et al., 2016), and surface emissions of CO_2 and CH_4 (Lara et al., 2015; Wainwright et al., 2015). At typical sizes, several thousand ice wedge polygons~~Several thousand ice wedge polygons of typical size~~ may occupy a single square kilometer of terrain, motivating our development of an automated approach to ~~map polygonal boundaries~~mapping. The key innovation in our method is the use of a convolutional neural network (CNN), a variety of machine learning algorithm, to identify ~~troughs (*i.e.*,~~

~~polygonal boundaries) within lidar-derived imagery~~pixels representing polygon boundaries. ~~The same techniques could be applied to any form of remotely sensed data with sufficient spatial resolution; however, using a digital elevation model (DEM) as input, our approach~~Integrated within a set of common image processing operations, this approach permits the extraction of microtopographic attributes from entire populations of ice wedge polygons at the kilometer scale or greater.

Previous geospatial surveys of polygonal microtopography have often aimed to map the occurrence of two geomorphic endmembers: basin-shaped low-centered polygons (LCPs), which are characterized by rims of soil at the perimeters, and hummock-shaped high-centered polygons (HCPs), which often are associated with permafrost degradation. ~~Several recent a~~Analyses of historic aerial photography have demonstrated a pan-Arctic acceleration since 1989 in rates of LCP conversion into HCPs~~High-resolution inventories of polygon geomorphology are of hydrologic and ecologic interest because decimeter-scale variability in polygon microtopography can drive pronounced changes to soil drainage (Liljedahl et al., 2016) and surface emissions of CO$_2$ and CH$_4$ (Lara et al., 2015; Wainwright et al., 2015). Several previous geospatial analyses of polygon geomorphology have been motivated by a recent acceleration in the rates at which low-centered polygons (LCPs) are converted into high-centered polygons (HCPs)~~, a process ~~driven by permafrost degradation~~ which improves soil ~~aeration~~ drainage and stimulates enhanced emissions of CO$_2$ (Jorgenson et al., 2006; Raynolds et al., 2014; Jorgenson et al., 2015; Liljedahl et al., 2016). ~~In these prior studies, the most common form of imagery used to demonstrate this transformation has been historic aerial photography. Thus far, analysis of available datasets has demonstrated a consistent uptick, beginning around 1989, in landscape-scale HCP formation throughout the high Arctic; however~~Nonetheless, precise rates of geomorphic change have been difficult to quantify, as ~~the surveys primarily have~~these surveys typically have relied on proxy indicators, such as the presence of ponded water in deepening HCP troughs ~~(Steedman et al., 2017)~~. In a related effort to characterize contemporary polygon microtopography, a landcover map of LCP and HCP occurrence across the Arctic coastal plain of northern Alaska was recently developed using multispectral imagery from the Landsat 8 satellite at 30 m resolution (Lara et al., 2018). This dataset offers a static estimate of variation in polygonal form over unprecedented spatial scales; however, geomorphology was inferred from the characteristics of pixels larger than typical polygons, preventing inspection of individual features.

Higher-resolution approaches to ~~explicitly~~ segment imagery into discrete ice wedge polygons have often ~~typically~~ been motivated by efforts to ~~analyze~~analyse trough ~~trough~~ network geometry. On both Earth and Mars, for example, paleo-environmental conditions in remnant polygonal landscapes have been inferred by comparing parameters such as ~~trough~~ boundary spacing and orientation with systems in modern periglacial terrain (*e.g.*, Pina et al., 2008; Levy et al., 2009; Ulrich et al., 2011~~; Ewertowski et al., 2017)~~). An early semi-automated approach to delineating Martian polygons from satellite imagery was developed by Pina et al. (2006), who ~~first used~~employed morphological image processing operations to emphasize polygonal boundaries~~troughs~~, then applied a watershed transformation (discussed in Section 4.1.3) to identify discrete polygons. This workflow was later applied to lidar-derived ~~imagery~~ DEMs from a landscape outside Barrow, Alaska~~K~~ by Wainwright et al. (2015), but in their experience and our own, robust results at spatial scales approaching a square kilometer or greater were elusive.

Our application of CNNs to the task of identifying polygonal troughs was inspired by the remarkable solutions that CNNs recently have permitted to previously intractable image processing problems~~.~~. Aided by advances in the performance of graphics processing units (GPUs) over the last decade, CNNs have demonstrated unprecedented skill at tasks analogous to ice wedge polygon delineation, such as cell membrane identification in biomedical images (Ciresan et al., 2012) or road extraction from satellite imagery (Kestur et al., 2018; Xu et al., 2018). Motivated by this potential, an exploratory study was recently conducted by Zhang et al. (2018), who demonstrated that a sophisticated neural network, the Mask R-CNN of He et al. (2017), is capable of end-to-end extraction of ice wedge polygons from satellite-based optical imagery, capturing ~79% of ice wedge polygons across a >134 km^2 field site and classifying each as HCP or LCP. The authors concluded that the method has potential for pan-Arctic mapping of polygonal landscapes. Here we explore an alternative approach, using a less complex CNN paired with a set of post-processing operations, to extract ice wedge polygons from high-resolution DEMs derived from airborne Lidar surveys. An advantage to this method is that training the CNN is rapid (~5 minutes or less on a personal laptop), permitting an iterative workflow~~,~~ in which supplementary data can easily be incorporated to boost skill in targeted areas. We demonstrate the suitability of this approach to extract ice wedge polygons with very high accuracy (up to 96% at the kilometer scale), applying it to ten field sites of 1 km^2 outside Barrow and Prudhoe Bay, Alaska. Because our method operates on high resolution elevation data, it enables direct measurement of polygonal microtopography, and we anticipate that~~,~~ in the future, the method ~~it~~ will permit precise monitoring of surface deformation in landscapes covered by repeat airborne surveys.~~As the performance of graphics processing units (GPUs) has steadily improved over the last decade, CNNs have demonstrated unprecedented skill at tasks analogous to trough delineation, such as cell membrane identification in biomedical images (Ciresan et al., 2012) or road extraction from satellite imagery (Kestur et al., 2018). In training a CNN to analyze lidar imagery from polygonal terrain, our goal was to improve the speed and accuracy of automated polygon delineation in modern tundra environments. Because our method extracts polygons from a high resolution DEM, in enables thorough, landscape-scale inventories of polygonal microtopography, and we anticipate that it will permit precise monitoring of contemporary surface deformation in landscapes covered by repeat airborne surveys.~~

2 Study areas and data acquisition

To demonstrate the flexibility of our approach, we applied it simultaneously ~~at study sites~~at two clusters of study sites near Barrow and Prudhoe Bay, Alaska, ~~two setting~~settings with highly divergent ice wedge polygon geomorphology, ~300 km distant from one another (Fig S1).

2.1 Barrow

The first cluster of study sites (Figs. S2-S3) ~~The first study site~~ is located within 10 km of the Beaufort Sea coast ~~(Fig S1A)~~ in the Barrow Environmental Observatory, operated by the National Environmental Observatory Network (NEON). Mean elevation is less than 5 m above sea level, and vegetation consists of uniformly low-growing grasses and sedges.

Mesoscale topography is mostly flat but marked by depressions up to 2 m deep associated with draws and drained lake beds ~~(Fig S1B)~~. In the landcover map of Lara et al. (2018), the area is characterized by extensive coverage by both LCPs and HCPs, with occasional lakes and patches of non-polygonal meadow. Microtopography at the sites reflects nearly ubiquitous ice wedge development, which becomes occluded in some of the depressions. Ice wedge polygons are of complex geometry and highly variable area, ranging from ~10~~-~~ m^2 to >2000 m^2. An airborne lidar survey was flown in August 2012 as part of the U.S. Department of Energy's Next Generation Ecosystems Experiment-Arctic program (https://ngee-arctic.ornl.gov/). The resulting point cloud was processed into a 2~~50~~5 cm horizontal resolution DEM ~~digital elevation model (DEM)~~ with an estimated vertical accuracy of 0.145 m (Wilson et al., 2012). In the present study, to compare algorithm performance on data of variable spatial resolution, the 25 cm DEM was resampled at 50 cm and 100 cm resolution. Two sites of 1 km^2, here referred to as Barrow-1 and Barrow-2, were extracted from the DEMs and processed using our workflow.

2.2 Prudhoe Bay

The second cluster of sites (Figs. S4-S11) ~~The second site~~ is approximately 300 km east of the first and farther inland, located ~40 km south of Prudhoe Bay, AK (Fig S1~~A~~). As at Barrow, vegetation consists almost exclusively of low and even-growing grasses and sedges. Mesoscale topography is ~~extremely~~generally flat, with a slight (<4%) dip toward the northwest ~~(Fig S1C)~~. In the landcover map of Lara et al. (2018), the area is primarily characterized by HCPs, with smaller clusters of LCPs, patches of non-polygonal meadow, and occasional lakes. Ice wedge ~~troughs are apparent throughout the training site, demarcating polygons~~ polygons are generally of more consistent area than those of Barrow, ~400-800 m^2. Airborne lidar data was acquired in August 2012 by the Bureau of Economic Geology at the University of Texas at Austin (Paine et al., 2015) and subsequently processed into 25 cm, 50 cm, and 100 cm resolution DEMs~~a 50 cm resolution DEM (Paine et al., 2015)~~. Vertical accuracy was estimated at 0.10 m. As the Prudhoe Bay survey area is substantially larger than the survey area at Barrow, eight sites of 1 km^2, here referred to as Prudhoe-1 through Prudhoe-8, were extracted from the DEMs and processed using our workflow.

3 Methods

3.1 Polygon delineation algorithm

A chart summarizing our iterative workflow is presented in Fig. 1, and several intermediate stages in the polygon delineation algorithm are illustrated in Fig. 2.~~A high-level summary of our workflow is presented in Fig 1, using a sample of the data from Barrow.~~ In the first (pre-processing) stage, regional trends were removed from a DEM (Fig. S~~2~~12), generating an image of polygonal microtopography (Fig. 2~~1~~A). Next, the microtopographic information was processed by a CNN, which was trained to use the 27×27 neighborhood surrounding each pixel to assign a label of "~~trough~~boundary" or "not ~~trough~~boundary" (Fig. 2~~1~~B). A distance transformation was then applied, (*i.e.*, each non-boundary pixel was assigned a negative intensity proportional to its Euclidean distance from the closest boundary), ~~Each non-trough pixel was then assigned~~

~~a negative intensity proportional to its Euclidean distance from the closest trough, ~~generating a grayscale image analogous to a DEM of isolated basins, in which the polygonal ~~troughs ~~boundaries appear as ridges (Fig. 2~~1~~C). Subsequently, a watershed transform was applied to segment the image into discrete ice wedge polygons (Fig. 2~~1~~D). These steps, and a post-processing algorithm used to remove non-polygonal terrain from the final image, are described in detail below.

3.1.1 Pre-processing

In the pre-processing stage, regional topographic trends were estimated by processing the DEM with a 2D filter, which assigned to each pixel the mean elevation within a 20 m ~~(40 pixel) ~~radius. This radius was chosen such that the area over which elevation was averaged would be larger than a typical ice wedge polygon. Polygon-scale ~~M~~microtopography was then estimated by subtracting the~~is~~ regional topography from the DEM (Fig S~~2~~12). In preparation for passing the data to the CNN, microtopography was ~~then ~~subsequently converted to 8-bit gray~~-~~scale imagery. The minimum intensity (0) was assigned to depressions of 0.7 m or greater, and the maximum intensity (255) was assigned to ridges of 0.7 m or greater. ~~, as t~~These bounds captured >99% of pixel values at each study site. Finally, one thumbnail-sized image was created for each pixel in the microtopography raster, capturing the immediate neighbourhood surrounding it. These thumbnail images were the direct input to the CNN. The CNN required the width in pixels of each thumbnail to be an odd multiple of 9; therefore, at 50 cm resolution the thumbnails were assigned a width of 27 pixels (13.5 m), at 25 cm resolution a width of 45 pixels (11.25 m), and at 100 cm resolution a width of 27 pixels (27 m). The width of these thumbnails was chosen such that each image would contain sufficient spatial context for a human observer to distinguish easily between polygonal boundaries, which typically were demarcated by inter-polygonal troughs, and other microtopographic depressions such as LCP centers.

3.1.2 Convolutional neural network

The function of the CNN in our workflow was to identify pixels likely to represent ~~troughs~~boundaries. Conceptually, a CNN is a classification tool that ~~which ~~accepts images of a fixed size (in our case, the thumbnails described in the previous section~~27×27 grayscale arrays~~) as input and generates categorical labels as output. The CNN determines decision criteria through training with a set of manually~~-~~labeled images. The architecture of a CNN consists of a user-defined sequence of components, or layers, which take inspiration from the neural connections of the visual cortex. We developed our CNN in MATLAB (R2017b) using the Image Processing, Parallel Computing, and Neural Network toolboxes. We ~~Our CNN was~~ purposefully constructed the CNN with an architecture of minimal complexity, to maximize the efficiency of training and application. Here we briefly ~~summarize ~~describe the function of each layer in our CNN; for more detailed description, the reader is directed to Ciresan et al. (2012).

Summarized in Table S1, t~~T~~he most important components of our CNN were a single convolutional layer, a max-pooling layer, and two fully connected layers~~ (Table S1)~~. In the convolutional layer, a set of 2D filters was applied to the input image, generating intermediate images in which features including concavities, convexities, or linear edges were detected. The max-pooling layer downsized the height and width of these~~se~~ intermediate images by a factor of three, ~~from 27×27 to 9×9~~

~~arrays~~ by selecting the highest intensity pixel in a moving 3×3 window with a stride of 3 pixels. Each pixel in the downsized intermediate images was then passed as an input signal to the fully connected layers, which functioned identically to standard neural networks. Two additional components of our CNN were Rectified Linear Unit (ReLU) layers, which enhance non-linearity by reassigning a value of zero to any negative signals output by a preceding layer, and a softmax layer, which converted the output from the final fully connected layer into a probability for each categorical label (*i.e.*, ~~trough~~ boundary or not ~~trough~~ boundary).

During training, the weights of the 2D filters in the convolutional layer and the activation functions of the neurons in the fully connected layers were optimized to correctly predict the labels in a training deck of images. Our workflow was designed to generate the training deck primarily by processing ~~square~~ 100 ×100 m tiles of manually-labeled imagery~~ with 100 m (200 pixel) edges.~~ In each of these tiles, ~~trough~~ boundary pixels were delineated by hand in a standard raster graphics editor, a process that required ~1 hour per tile at 50 cm resolution (Fig. S13~~3~~A). Our algorithm imported these tiles, identified the geographic coordinates of each pixel identified as a ~~trough~~boundary, then ~~extracted the 27×27 neighborhood surrounding each trough pixel~~created a thumbnail image centered on that pixel from the 8-bit microtopographic imagery. This procedure generated several thousand thumbnail~~-sized~~ ~~-~~images centered on ~~a trough~~boundaries from each manually delineated tile. Subsequently, an equal number of pixels not labeled as ~~troughs~~ boundaries were selected at random, and the thumbnail ~~image~~ extraction procedure was repeated~~.~~, ~~to~~ generating~~e~~ a set of non-boundary images for the training deck. Finally, for more targeted~~ing~~ training that did not require full delineation of a 100 ×100 m tile, individual instances of boundary or non-boundary features could also be added to supplement the training deck, based again on manual delineation (Fig. S13B). Just prior to training, 25% of the training deck assembled by these methods was set aside to be used for validation.

Once trained, the CNN was executed to assign a label of “boundary” or “not boundary” to the thumbnail image corresponding to each pixel of a study site. These labels were then reassembled into a binary image of polygon boundaries (Fig. 2B), which was further processed to extract discrete ice wedge polygons.~~Due to the compact architecture of our CNN, training and execution speeds were rapid. Processing information from ~30,000 images, classification accuracies >99% on the training deck and >98% on the validation deck were obtained in less than ten minutes on a single GPU. Application of the trained CNN to a square kilometer of gridded lidar data required ~90 seconds. This speed permitted us to iteratively train the CNN, identify regions of the survey area where it performed poorly, and supplement the training data with labeled imagery from those regions. Optionally, our workflow at this stage also permitted supplementation of the training deck with individual instances of problematic features, such as LCP centers or especially narrow troughs (Fig S3B). This permitted us to increase the skill of the CNN in a targeted manner without the time requirements of manually labeling a full 100 m tile of imagery.~~

3.1.3 Polygon extraction

After applying the CNN to classify all pixels at a site as ~~trough~~boundary or not ~~trough~~boundary, we extracted discrete ice wedge polygons by ~~through~~ applying ~~ication of~~ several standard image processing operations. The first step was elimination of “salt and pepper” noise in the binary image, which we accomplished by eliminating all contiguous sets of ~~trough~~boundary-identified pixels with an area < ~~-~~20 m^2. This threshold was selected based on the reasoning that most true boundary pixels should be part of a continuous network, covering an area arbitrarily larger than 20 m^2, while most false detections should occur in smaller clusters. Next, we applied a distance transform, assigning~~ed~~ to every non-~~trough~~ boundary pixel a negative intensity equal to its Euclidean distance from the nearest boundary~~trough~~. This created an intermediate image in which each ice wedge polygon appeared as a valley, surrounded on all or most sides by ridges representing the ~~trough~~ice wedge network (Fig. 2~~1~~C). At this stage, to prevent over-segmentation, valleys with maximum depths of ~~three pixel-widths or fewer~~1.5 meters or less were then identified and merged with the closest neighbor through morphological reconstruction (Soille, 1992). The effect of this procedure was to ensure that the algorithm would only delineate polygons whose centers contained at least one point greater than 1.5 m from the boundaries, as field observations indicate that ice wedge polygons tend to measure at least several meters across (Leffingwell, 1915; Lachenbruch, 1962)~~not delineate any polygon whose center did not include a point greater than 1.5 m from the troughs~~. Next, watershed segmentation was ~~then~~ applied to divide the valleys into discrete polygons (Fig. 2D). Our use of this operation was inspired by its incorporation in the polygon delineation method developed by Pina et al. (2006). Conceptually, this procedure was analogous to identifying the up-gradient region or area of attraction surrounding each local minimum.

3.1.4 Partitioning of non-polygonal ground ~~1 Post-processing~~

In the final stage of delineation, we partitioned out regions of a survey area that had been segmented using the techniques described above, but were unlikely to represent true ice wedge polygons. For example, polygons were eliminated from the draw in the southern half of Barrow-1 ~~the Barrow training site~~ (Fig. S~~3~~2A~~1B~~), where microtopography wa~~i~~s too occluded to permit accurate delineation. Toward this aim, our algorithm ~~tabulated within the boundaries of each delineated polygon~~analysed individually each boundary between two polygons (black line~~s~~ segments in Fig. 2~~1~~D), tabulating the number of pixels that had been identified positively by the CNN ~~as troughs~~ (white pixels in Fig. 2~~1~~B). It then dissolved all boundaries in which less than half the pixels had been classified positively, merging adjacent polygons. In practice, this procedure resulted in areas of non-polygonal terrain being demarcated by unusually large “polygons.”~~.~~ We removed these areas by partitioning out any polygon with an area greater than 10,000 m^2, a threshold selected to be arbitrarily larger than most real ice wedge polygons. ~~It then eliminated contiguous regions greater than 10,000 m^2 (representing the area of several large polygons) where less than half the boundary pixels of the contained polygons had been labeled as troughs.~~ This procedure had the strengths of being conceptually simple and providing a deterministic means of partitioning non-polygonal terrain from the rest of a survey area.

3.2 Microtopographic analysis

To demonstrate the ~~our workflow's~~ capabilities of our workflow for microtopographic analysis, we developed a simple method for measuring the relative elevation at the center of each delineated polygon~~. This served as~~, serving as a ~~rough~~ proxy for LCP or HCP form. In each polygon, we first applied a distance transform, calculating the distance from the closest boundary of all interior pixels. We then divided the area of the polygon in half at the median distance from boundaries, designating a ring of “outer” pixels and an equally sized core of center pixels. Microtopographic relief was then estimated as the difference in mean elevation between the center and outer pixels.~~calculated the Euclidean distance from the trough network of all interior pixels, then identified the pixel most distal from a trough. Elevation was extracted from this pixel and compared with mean elevation in the troughs surrounding the polygon. We thus assigned a number to each polygon which consistently was lower in LCPs and higher in HCPs.~~

3.3 Case study experimental design

The case study was first conducted using topographic data at 50 cm resolution, then repeated at 25 cm and 100 cm resolution. Training was focused primarily on sites Prudhoe-1 and Barrow-1. ~~We conducted our case study in two phases: calibration and validation. Calibration was performed at two 1 km^2 sites, sampled from the Barrow and Prudhoe Bay DEMs.~~ Leveraging the rapid training and application times of our CNN, we manually delineated one 100 ×100 m tile of imagery at a time from either site, trained the CNN, extracted results from both sites, then introduced additional training data from regions of poor performance to improve skill (Fig. 1). After four iterations of this approach, the CNN incorporated training data from three fully-delineated tiles at Barrow-1 and one at Prudhoe-1, representing 3% and 1% of the sites, respectively. From this point, we opted to “fine-tune” the CNN by supplementing the training deck directly with instances of problematic features, rather than using information from fully-delineated tiles. Several examples of boundary and non-boundary features were included from Barrow-1 and Prudhoe-1. Next, to test its extensibility, the CNN was applied across the remaining sites, and re-trained once more. In this final iteration, several instances of boundary and non-boundary features (but no fully delineated tiles) were incorporated into the training deck from sites Prudhoe-2, Prudhoe-3, and Prudhoe-4. No training data at all were incorporated from sites Prudhoe-5 through Prudhoe-8 or Barrow-2. (All training data used in the final iteration of our workflow can be viewed in the data and code repository accompanying this article.) Once this procedure was complete at 50 cm resolution, training decks at 25 cm and 100 cm resolution were prepared. To generate CNNs comparable to the network trained on 50 cm data, ~~T~~the 25 cm and 100 cm training decks were constructed using data sampled from identical geographic locations, but manual labeling was performed without reference to the labeled 50 cm resolution data.

After the CNN was trained and applied across all study sites, we quantified the performance of the polygon delineation algorithm through manual validation. At each site, we first calculated the total area and number of polygons extracted from the landscape. We then randomly sampled 500 of the computer-delineated polygons, and classified each as either whole, fragmentary, conglomerate, or false. Fragmentary polygons were defined as computer-delineated polygons that ~~which~~ included less than 90% of one real polygon; conglomerate polygons were defined as computer-delineated polygons comprising parts of two or more real polygons; and false polygons were defined as computer-delineated polygons occupying terrain in which no polygonal pattern was deemed visible to the human evaluator. The percentage of computer-delineated polygonal terrain corresponding to each class was then calculated by number of polygons and by area. This procedure was completed for all ten study sites at 50 cm data resolution, and for sites Prudhoe-1 and Barrow-1 at 25 cm and 100 cm resolution.~~took an iterative~~

~~approach to extracting polygons from both sites: after training and applying an initial model, new training data were introduced from regions of poor performance to improve skill. After several iterations, our final CNN was trained using manually labeled data from three 100 m tiles at the Barrow training site and one at the Prudhoe Bay site, representing 3% and 1% of the training plots, respectively. The non-trough training data were also supplemented with some additional examples of LCP centers from Barrow, to discourage inaccurate labeling as troughs (Fig S3B). After training the CNN, we tested its extensibility by applying it at two validation sites, also 1 km^2 in area, selected from the same landscapes. Once delineation was complete, to illustrate capabilities for microtopographic analysis, we calculated the relative elevations of polygon centers at the Prudhoe Bay training site, which was chosen for the tendency of its polygons to cluster into areas with high concentrations of either LCPs or HCPs.~~

4 Results and discussion

4.1 Training speed and accuracy

Due to the compact architecture of our CNN, training speeds at 50 cm resolution and 100 cm resolution were rapid. At 50 cm resolution, the training procedure operated on a deck of ~36,000 thumbnail images. Executed on a personal laptop with an Intel i7 CPU and a single GeForce MX150 GPU, accuracies >97% on the training deck and >95% on the validation deck of thumbnails were achieved in less than five minutes. At 100 cm resolution, the procedure operated on ~12,000 thumbnail images, achieving comparable levels of accuracy within 90 seconds. These speeds enable the iterative approach to training on which our workflow is based (Fig. 1), as the CNN can be re-trained quickly to incorporate new data when applied across increasingly large areas.

Using 25 cm resolution data, the training procedure operated on a set of ~115,000 thumbnail images. Accuracies >97% on the training deck and >95% on the validation deck were once more obtained, but training required just under one hour on the same computer. This substantial increase to training time is attributable to the facts that more thumbnail images were processed, and the number of pixels in each thumbnail was larger, making execution of the CNN more computationally expensive.

4.2 Delineation speed and validation

Operating at 50 cm resolution, delineation of ice wedge polygons within a 1 km^2 field site required ~2 minutes, including application of the CNN and subsequent post-processing. Results generally were very accurate; across study sites, ~1000-5000 ice wedge polygons were detected per square kilometer, of which 85-96% were estimated as “whole” during manual validation, representing 70-96% of the polygonal ground by area (Table 1). The most common type of error at all sites with <95% accuracy was incorrect aggregation of several real ice wedge polygons into a single feature. Unsurprisingly, performance was strongest ~~highest~~ at sites with clearly defined polygon boundaries and relatively simple polygonal geometry, such as Prudhoe-1, Prudhoe-7, and Prudhoe-8 (Figs. S4~~3~~, S10, S11). In contrast, performance was weakest at sites such as Barrow-2 or Prudhoe-6 (Figs. S2, S9), where considerable swaths of terrain are characterized by faint microtopography, as; ~~,~~

~~in which~~ ice wedge polygons appear to grade into non-polygonal terrain. In such locations, polygonal boundaries frequently went undetected, resulting in the delineation of unrealistically large conglomerate polygons. In general, the results of the delineation clearly illustrate that the polygonal network at Barrow possesses more complex geometry than Prudhoe Bay, with many instances where secondary or tertiary ice wedges appear to subdivide larger ice wedge polygons.

Although simple, our post-processing procedure for partitioning out non-polygonal ground from the results was generally accurate. Examples of features successfully removed from the 50 cm resolution DEMs included thaw lakes (Figs. S6-S10), drained thaw lake basins (Fig. S2), stream beds (Figs. S5, S9), non-polygonal marsh (Figs. S3, S7, S9), shallow ponds (Fig. S7), and the flood plain of a braided stream (Fig. S10). Because the partitioning procedure ~~worked by~~ removed~~ing~~ areas with a low density of boundaries identified by the CNN, we found that it could be ~~efficiently trained~~trained efficiently by supplementing the training deck with extra examples of "non-boundary" thumbnails extracted from these regions. Across the study sites, we encountered almost no cases where ~~in which~~ well-defined polygons were mistakenly partitioned out~~ of the results by the computer~~. Instances of non-polygonal ground mistakenly ~~being~~ classified as polygonal accounted for less than 1% of machine-delineated polygons by area.

When the delineation algorithm was repeated using data at 100 cm resolution, delineation speeds increased somewhat, but performance dropped significantly, with greater declines occurring in the challenging environment of Barrow (Table S2). We attribute declines in performance to an obscuration of fainter polygonal boundaries ~~that~~ at this resolution, and to ~~that~~ decreases in the amount of contextual information that can be derived from the neighbourhood of any given pixel, reducing the capacity for the algorithm ~~computer~~ to distinguish between polygonal troughs and other microtopographic depressions (Fig. S14). Somewhat surprisingly, performance also declined slightly~~ somewhat~~ using data at 25 cm resolution, with larger fractions of fragmentary and false polygons accounting for most of the increases in errors (Table S2). We attribute these mistakes to the larger number of distinguishable features ~~fact that, as more features become distinguishable~~ in the higher resolution data, leading the algorithm ~~computer appears~~ to more frequently mistake features not encountered in the training dataset as boundary pixels. As the imagery~~ is nonetheless crisper~~ nonetheless is of sufficient resolution for our purposes, we anticipate that, with augmentations to the training dataset, performance at 25 cm resolution could improve and even exceed performance at 50 cm resolution; ~~nonetheless~~however, performance speeds were generally inhibitory to our workflow, as delineation required ~50 minutes per square kilometer. We therefore conclude that 50 cm resolution data is optimal for our analysis.

4.3 Measurement of polygonal microtopography

The time required to execute our procedure for measuring polygonal microtopography varied from ~10-30 seconds per square kilometer at 50 cm resolution, depending on the number of polygons delineated. A comparison of calculated relief at polygon centers between Prudhoe-1 and Barrow-1 reveals that both sites are characterized by the prevalence of HCPs, which surround smaller clusters of LCPs (Fig. 3). Relief tends to be more extreme at Barrow, with the relative elevation of polygon centers commonly approaching 20-30 cm. Our automated calculations of relief align well with visual inspection of the DEMs, as rimmed LCPs are consistently assigned ~~a~~ negative center elevations. To our knowledge, our results represent the first direct measurement of polygonal relief at the kilometer-scale, demonstrating a spectrum of center elevations rather than a binary classification into LCP or HCP. We anticipate these measurements may be useful for further investigations into relationships between microtopography, soil moisture, and carbon fluxes (*e.g.*, Wainwright et al., 2015).

4.4 Comparison with Mask R-CNN and future applications

Several key differences are apparent between the workflow presented in this paper (hereafter termed the CNN-watershed approach) and a recent implementation of Mask R-CNN for mapping of ice wedge polygons (Zhang et al., 2018), revealing relative strengths to each approach. An advantage of the CNN-watershed approach, stemming from its sparse neural architecture, is that training times are extremely rapid, facilitating iterative improvements to skill (Fig. 1). In contrast, although inference times using the CNN-watershed approach are reasonable, extraction of ice wedge polygons over broad landscapes is several times faster using Mask R-CNN, with a reported time of ~21 minutes for inference over 134 km^2 of terrain (Zhang et al., 2018). Because the CNN-watershed approach operates on high-resolution DEMs, it enables direct quantification of polygonal relief, whereas Mask R-CNN instead produces a binary classification of each polygon as either LCP or HCP. The CNN-watershed approach is therefore useful for generating unique datasets summarizing polygonal geomorphology, demonstrating high-performance at spatial scales typical of airborne surveys using lidar or photogrammetry to produce high-resolution DEMs. In comparison, because Mask R-CNN has been trained to operate on satellite-derived optical imagery with global coverage, it is uniquely well-suited for application across very broad regions, with potential to generate pan-Arctic maps of land cover by polygon type. Because of differences in training ~~times,~~ and inference procedures~~,~~ and the spatial scales at which they ideally operate, the training data requirements and accuracy of the two approaches are difficult to compare directly; nonetheless, in several aspects, performance appears to be similar (Text S1).

Because the CNN-watershed approach generates direct measurements of polygonal microtopography, one application to which it is uniquely amenable is precise monitoring of microtopopographic deformation in areas covered by repeat airborne surveys. Through such analysis, we anticipate that it will permit polygon-level quantification of ground subsidence over timespans of years, potentially yielding new insights into the vulnerability of various landscape units to thermokarst. An additional research problem~~,~~ to which the CNN-watershed approach is well-suited~~,~~ is quanti~~fying~~ication of ~~fication of~~ polygonal network parameters, such as boundary spacing and orientation, to explore relationships to ~~with~~ environmental factors such as climate (*e.g.*, Pina et al., 2008; Ulrich et al., 2011). These boundaries (black line segments in Figs 2D, 3B, 3D) are naturally delineated through implementation of the watershed transformation. We acknowledge that, because the surface expression of ice wedges is sometimes subtle or non-existent, ground-based delineation methods are the highest-accuracy approach to mapping ice wedge networks (Lousada et al., 2018). Nonetheless, by segmenting machine-delineated networks into individual boundaries, our workflow permits the estimation of network statistics at spatial scales unattainable through on-site surveying.

~~Results from the Barrow training site demonstrate the capability of our workflow to extract polygons of highly~~

~~variable geometry (Fig 2A). The results also demonstrate success of our post-processing workflow in partitioning out a streambed traversing the southern half of the site and a lakebed in the northeast quadrant. The final delineation makes immediately apparent the complex distribution of polygon type at the kilometer scale. For example, just above the draw in the southwest quadrant is a cluster of comparatively small HCPs that stands in stark contrast to the larger LCPs and intermediate polygons dispersed elsewhere. The complexity of the delineated network underscores the value of an automated approach to polygon extraction, as the manual processing necessary to achieve similar results would require prohibitively large amounts of time. Moreover, the site contains several areas in which larger polygons typically are subdivided by secondary and tertiary ice wedges. Although polygonal boundaries are somewhat ambiguous in such situations, our algorithm provides a consistent, deterministic method of interpretation, which is set by the user's delineation of troughs in the training dataset.~~

~~Results from the Prudhoe Bay training site, using the same CNN, demonstrate the ease with which our workflow can be extended across diverse landscapes. We note that, while we had initially trained our CNN using three tiles of manually labeled data from Barrow, successful extension to Prudhoe Bay required incorporation of only one additional tile, from a region in the southeast quadrant characterized by LCPs with exceptionally narrow troughs. The final delineation (Fig 2B) emphasizes the far more consistent size and geometry of polygons relative to Barrow, which may be associated with the simpler mesoscale topography (Lachenbruch, 1962). The ice wedge network at Prudhoe Bay appears to be less riddled with secondary ice wedges than at Barrow, but some examples are evident in HCPs.~~

~~Results from the validation sites demonstrate similar skill in polygon delineation without incorporation of any new training data. At the Barrow validation site (Fig S4A), the algorithm partitioned out another stream bed and three regions where surface inundation prevented accurate delineation of polygonal microtopography. Elsewhere, a mixture of LCPs, HCPs, and intermediate polygons reminiscent of the training site was successfully delineated. At the Prudhoe Bay validation site (Fig S4B), the algorithm partitioned out a thaw lake and the flood plain of a braided stream, in which no ice wedge troughs are present. It also partitioned out a zone just north of the lake where ice wedges may be present in the subsurface, but microtopography is difficult to distinguish. Through the remainder of the area, the algorithm delineated a set of LCPs and HCPs of similar size and geometry to the training site, but with somewhat subtler troughs.~~

~~The capacity of our workflow for high-resolution microtopographic analysis is demonstrated in estimates of relief at the polygon centers at the Prudhoe Bay training site (Fig 3A). Through relatively simple calculations, we highlight the presence of a distinct cluster of LCPs in the southeast quadrant, which emerges from the mixture of HCPs and intermediate polygons present through the remainder of the site. Using this approach, our procedure provides a means to estimate polygonal geomorphology with unprecedented detail at the kilometer scale, which may improve predictions of soil wetness and polygon-scale fluxes of CO_2 and CH_4 (Lara et al., 2015; Wainwright et al., 2015). Simultaneously, the delineation performed at the same site provides a means to analyze trough geometric parameters, such as spacing and orientation. We acknowledge that, because the surface expression of ice wedges is sometimes subtle or non-existent, ground-based methods remain the highest-accuracy approach to mapping ice wedge networks (Lousada, 2018). Nonetheless, by segmenting machine-delineated polygonal boundaries into individual troughs (Fig 3B), very large datasets can rapidly be generated to estimate statistics at~~

~~spatial scales unattainable through on-site surveying.~~

5 Conclusions

A relatively simple CNN paired with a set of common image processing techniques is capable of extracting polygons of highly variable size and geometry from high-resolution DEMs of diverse tundra landscapes~~Convolutional neural networks enable rapid delineation of ice wedge networks at unprecedented spatial scales using airborne lidar datasets~~. Successful application of the CNN~~s~~ is facilitated by ~~employing relatively simple~~its sparse neural architecture, which permits rapid training, testing, and incorporation of new data to improve skill. The optimal spatial resolution for DEMs processed using the workflow is ~50 cm. ~~A single CNN is capable of extracting polygons of highly variable area and geometry from diverse tundra landscapes, using a training workflow that can be completed in a single afternoon.~~ Potential applications for the technology include: generation of high-resolution maps of land cover by polygon type, precise quantification of microtopographic deformation in areas covered by repeat airborne surveys, and rapid extraction of center elevations and ~~trough~~ boundary parameters including spacing and orientation. These capabilities can improve understanding of environmental influences on network geometry and facilitate assessments of contemporary landscape evolution in the Arctic.

Acknowledgements

We are grateful for the support provided for this research, which included: the Next Generation Ecosytem Experiments Arctic (NGEE-Arctic) project (DOE ERKP757) funded by the Office of Biological and Environmental Research in the U.S. Deparment of Energy Office of Science, and the NASA Earth and Space Science Fellowship program, for an award to CJA (80NSSC17K0376). We thank Dylan Harp (Los Alamos National Laboratory) for constructive conversations during manuscript preparation. We acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources that have contributed to the research results reported within this paper.

Data availability

Data and code are available through the corresponding author. Upon publication of the final paper, data and code will be uploaded to a Zenodo repository.

References

Ciresan, D., Giusti, A., Gambardella, L. M. and Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images, in: Advances in Neural Information Processing Systems 25, Pereira, F., Burges, C. J. C., Bottou, L. and Weinberger, Q (eds.), Curran Associates, Inc., 2843-2851, 2012.

He, K, Gkioxari, G., Dollar, P., Girshick, R.: Mask R-CNN, in: Proceedings of the 2017 IEEE International Conference on Computer Vision, IEEE, Piscataway, NJ, USA, 2017.

~~Ewertowski, M. W., Kijowski, A., Szuman, I., Tomczyk, A. M. and Kasprzak, L.: Low-altitude remote sening and GIS-based analysis of cropmarks: Classification of past thermal-contraction crack polygons in central western Poland, Geomorphology, 293, 418-432, 2017.~~

Jorgenson, M. T., Shur, Y. L. and Pullman, E. R.: Abrupt increase in permafrost degradation in Arctic Alaska, Geophysical Research Letters, 33(2), doi:10.1029/2005GL024960, 2006.

Jorgenson, M. T., Kanevskiy, M., Shur, Y., Moskalenko, N., Brown, D. R. N., Wickland, K., Striegl, R. and Koch, J.: Role of ground ice dynamics and ecological feedbacks in recent ice wedge degradation and stabilization, Journal of Geophysical Research: Earth Surface, 120(11), 2280–2297, doi:10.1002/2015JF003602, 2015.

Kestur, R., Farooq, S., Abdal, R., Mehraj, E., Narasipura, O. and Mudigere, M.: UFCN: A fully convolutional neural network for road extraction in RGB imagery acquired by remote sensing from an unmanned aerial vehicle, Journal of Applied Remote Sensing, 12, 016020, doi:10.1117/1.JRS,12.016020, 2018.

Lachenbruch, A. H.: Mechanics of thermal contraction cracks and ice-wedge polygons in permafrost, Special Paper, Geological Society of America, New York, 1962.

Lara, M. J., McGuire, A. D., Euskirchen, E. S., Tweedie, C. E., Hinkel, K. M. and Skurikhin, A. N.: Polygonal tundra geomorphological change in response to warming alters future CO_2 and CH_4 flux on the Barrow Peninsula, Global Change Biology, 21, 1634-1651, 2015.

Lara, M. J., Nitze, I., Grosse, G. and McGuire, A. D.: Tundra landform and vegetation trend maps for the Arctic Coastal Plain of northern Alaska, Scientific Data, 5, 180058, 2018.

Leffingwell, E. K.: Ground-ice wedges: The dominant form of ground-ice on the north coast of Alaska, J. Geol., 23, 635-654, 1915.

Levy, J. S., Marchant, D. R. and Head, J. W.: Thermal contraction crack polygons on Mars: A synthesis from HiRISE, Phoenix, and terrestrial analogue studies, Icarus, 226, 229-252, 2010.

Liljedahl, A. K., Boike, J., Daanen, R. P., Fedorov, A. N., Frost, G. V., Grosse, G., Hinzman, L. D., Iijma, Y., Jorgenson, J. C., Matveyeva, N., Necsoiu, M., Raynolds, M. K., Romanovsky, V. E., Schulla, J., Tape, K. D., Walker, D. A., Wilson, C. J., Yabuki, H., and Zona, D.: Pan-Arctic ice-wedge degradation in warming permafrost and its influence on tundra hydrology, Nat. Geosci., 9, 312–318, https://doi.org/10.1038/ngeo2674, 2016.

Lousada, M., Pina, M., Vieira, G., Bandeira, L. and Mora, C.: Evaluation of the use of very high resolution aerial imagery for accurate ice-wedge polygon mapping (Adventdalen, Svalbard), Science of the Total Environment, 615, 1574-1583, 2018.

Paine, J. G., Andrews, J. R., Saylam, K. and Tremblay, T. A.: Airborne LiDAR-based wetland and permafrost-feature mapping on an Arctic Coastal Plain, North Slope, Alaska, in: Remote Sensing of Wetlands: Applications and Advances, Tiner, R. W., Klemas, V. V., Lang, M. W. (eds.), CRC Press, Boca Raton, FL, USA, 413-434, 2015.

Pina, P., Saraiva, J., Bandeira, L. and Barata, T.: Identification of Martian polygonal patterns using the dynamics of watershed contours, in: Image Analysis on Recognition, Campilho, A., Kamel, M. (eds.), Springer-Verlag, Berlin-Heidelberg, 691-699, 2006.

Pina, P., Saraiva, J., Bandeira, L., and Antunes, J.: Polygonal terrains on Mars: A contribution to their geometric and topological characterization, Planetary and Space Science, 56, 1919-1924, 2008.

Raynolds, M. K., Walker, D. A., Ambrosius, K. J., Brown, J., Everett, K. R., Kanevskiy, M., Kofinas, G. P., Romanovsky, V. E., Shur, Y. and Webber, P. J.: Cumulative geoecological effects of 62 years of infrastructure and climate change in ice-rich permafrost landscapes, Prudhoe Bay Oilfield, Alaska, Global Change Biology, 20(4), 1211–1224, doi:10.1111/gcb.12500, 2014.

Soille, P.: Morphological Image Analysis, Springer-Verlag: Berlin Heidelberg, 2004.

~~Steedman, A. E., Lantz, T. C. and Kokelj, S. V.: Spatio-temporal variation in high-centre polygons and ice-wedge melt ponds, Tuktoyaktuk coastlands, Northwest Territories, Permafrost and Periglacial Processes, 28, 66-78, 2017.~~

Ulrich, M., Hauber, E., Herrzschuh, U., Hartel, S. and Schirrmeister, L.: Polygon pattern geomorphometry on Svalbard (Norway) and western Utopia Planitia (Mars) using high resolution stereo remote-sensing data, Geomorphology, 134, 197-216, 2011.

Wainwright, H. M., Dafflon, B., Smith, L. J., Hahn, M. S., Curtis, J. B., Wu, Y., Ulrich, C., Peterson, J. E., Torn, M. S. and Hubbard, S. S.: Identifying multiscale zonation and assessing the relative importance of polygon geomorphology on carbon fluxes in an Arctic tundra ecosystem, Journal of Geophysical Research: Biogeosciences, 120(4), 788–808, doi:10.1002/2014JG002799, 2015.

Wilson, C., Gangodagamage, C. and Rowland, J.: Digital elevation model, 0.5 m, Barrow Environmental Observatory, Alaska, 2012, Next Generation Ecosystem Experiments Arctic Data Collection, Oak Ridge National Laboratory, US Department of Energy, doi:10.5440/1109234, 2013.

Xu, Y. Y., Xie, Z., Feng, Y. X., Chen, Z. L.: Road extraction from high-resolution remote sensing imagery using deep learning, Remote Sensing, 10(9), 1461, 2018.

Zhang, W., Witharana, C., Liljedahl, A., and Kanevskiy, M.: Deep convolutional neural networks for automated characterization of Arctic ice-wedge polygons in very high spatial resolution aerial imagery, Remote Sensing, 10, 1487, 2018.

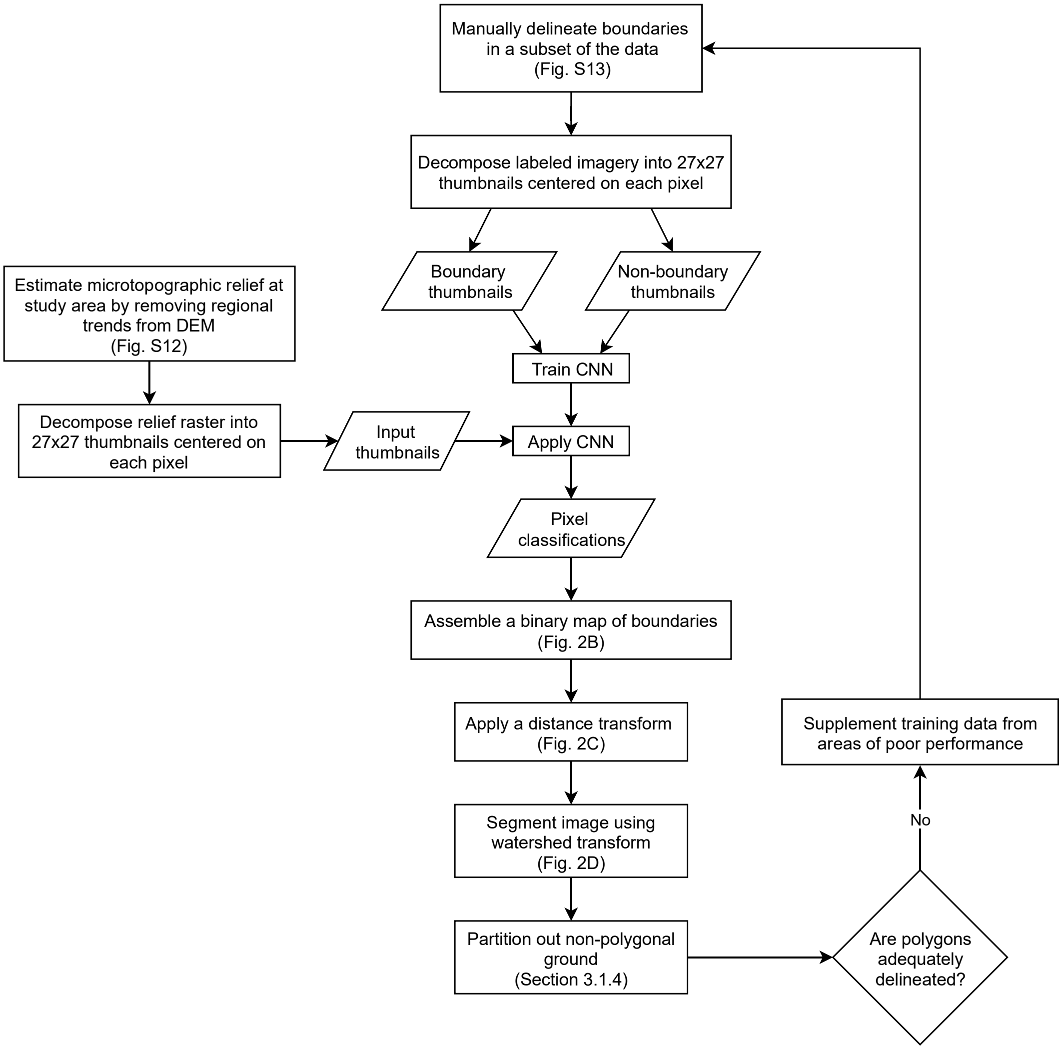

Figure 1. Schematic of our iterative workflow for polygon delineation.

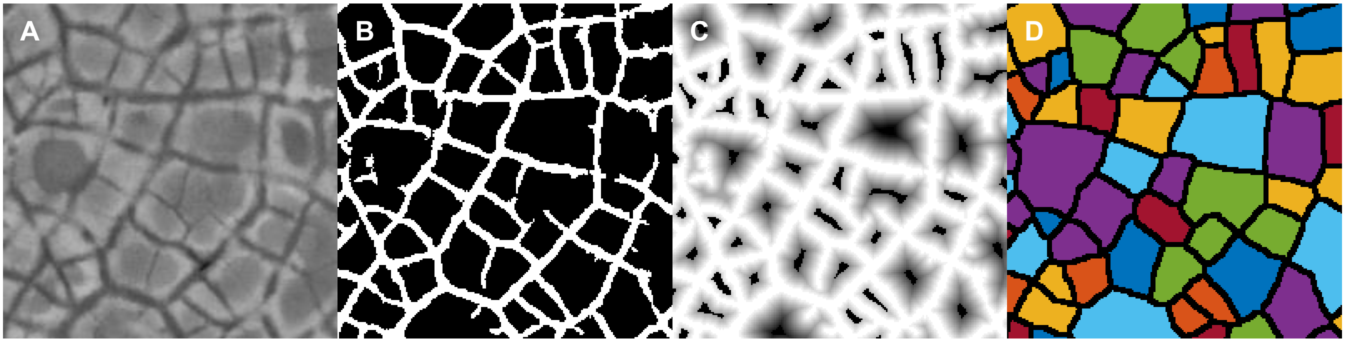

Figure ~~1~~2. ~~High-level schematic of our workflow~~Illustration of several intermediate stages in our workflow. The CNN processes information stored in ~~. The input to the CNN is~~ 8-bit grayscale imagery representing microtopography (A), estimated by removing regional trends from the lidar DEM (Fig S2). The CNN identifies pixels likely to represent troughs (B). Each non-trough pixel is assigned a negative intensity equal to its distance from a trough (C) and a watershed transformation is applied to segment the image into discrete polygons (D) (colors randomly applied to emphasize polygonal boundaries).

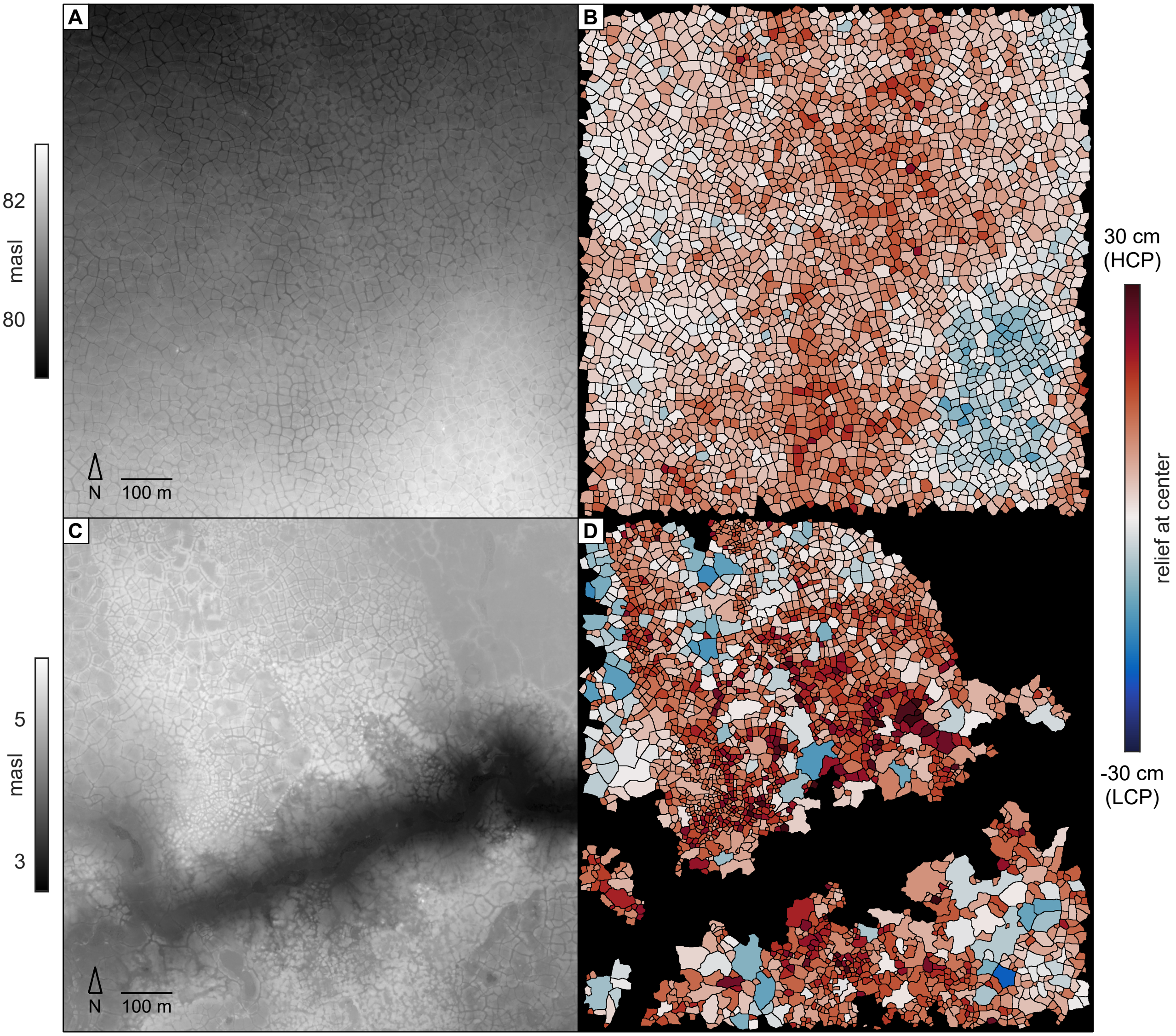

Figure 3~~2~~. DEMs and estimates of polygonal relief at sites Prudhoe-1 (A, B) and Barrow-1 (C, D). Polygons crossing the boundaries of study sites are removed in (B~~C~~, D).~~Delineation of ice wedge polygons using a single CNN at Barrow (A) and Prudhoe Bay (B) training sites. DEMs are shown on the left, and polygons are randomly colorized on the right.~~

Table 1. Results of manual validation at 50 cm data resolution (sites are 1 km^2).

			% of polygons by instance				*% of polygons by area*			
Site	Polygons identified	Polygonal area (%)	Whole	Fractional	Conglomerate	Non-polygonal	Whole	Fractional	Conglomerate	Non-polygonal
Barrow-1	2555	68.3	94.2	2.2	2.2	1.4	86.9	1.4	7.5	4.2
Barrow-2	2613	68.8	91.6	2.8	5.6	0.0	79.9	3.6	16.5	0.0
Prudhoe-1	3227	99.9	95.6	2.8	1.4	0.2	96.3	2.2	1.4	0.0
Prudhoe-2	4685	94.5	91.6	1.4	7.0	0.0	87.7	1.7	10.6	0.0
Prudhoe-3	1112	48.2	91.2	3.6	5.2	0.0	82.3	4.2	13.4	0.0
Prudhoe-4	1956	60.4	88.2	4.2	7.2	0.4	81.4	3.2	14.9	0.5
Prudhoe-5	2969	77.9	91.8	4.0	4.2	0	87.7	1.7	10.6	0
Prudhoe-6	1605	65.5	85.8	5.4	8.2	0.4	69.5	4.8	22.1	3.4
Prudhoe-7	1348	47.3	94.0	2.2	3.6	0.2	90.5	1.7	7.9	0.0
Prudhoe-8	3288	100	96.0	2.2	1.6	0.2	95.2	1.5	3.2	0.0

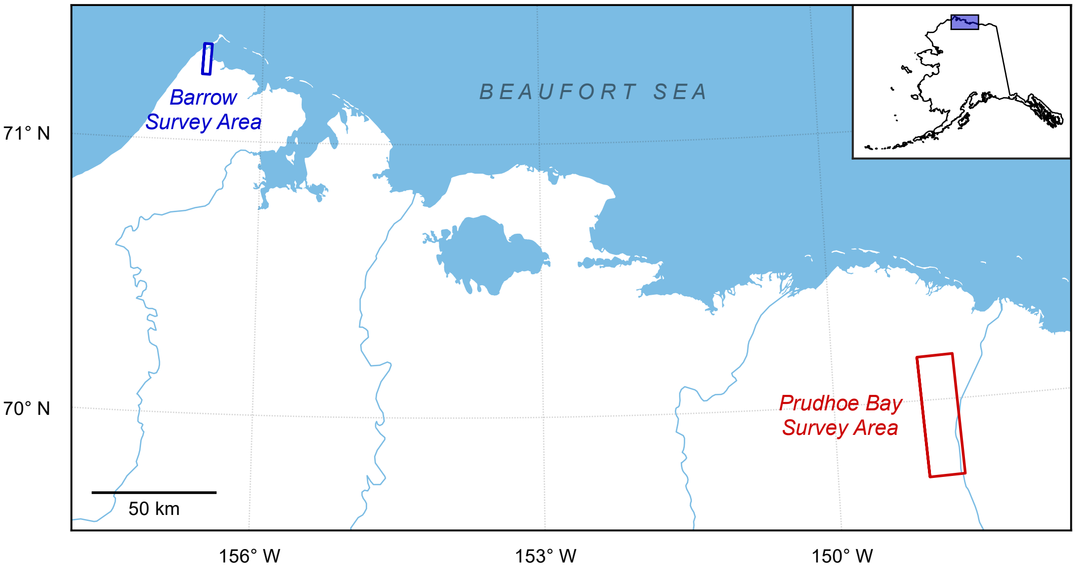

Figure S1. Bounding boxes of airborne lidar surveys~~ in Alaska Albers (NAD83) projection~~.

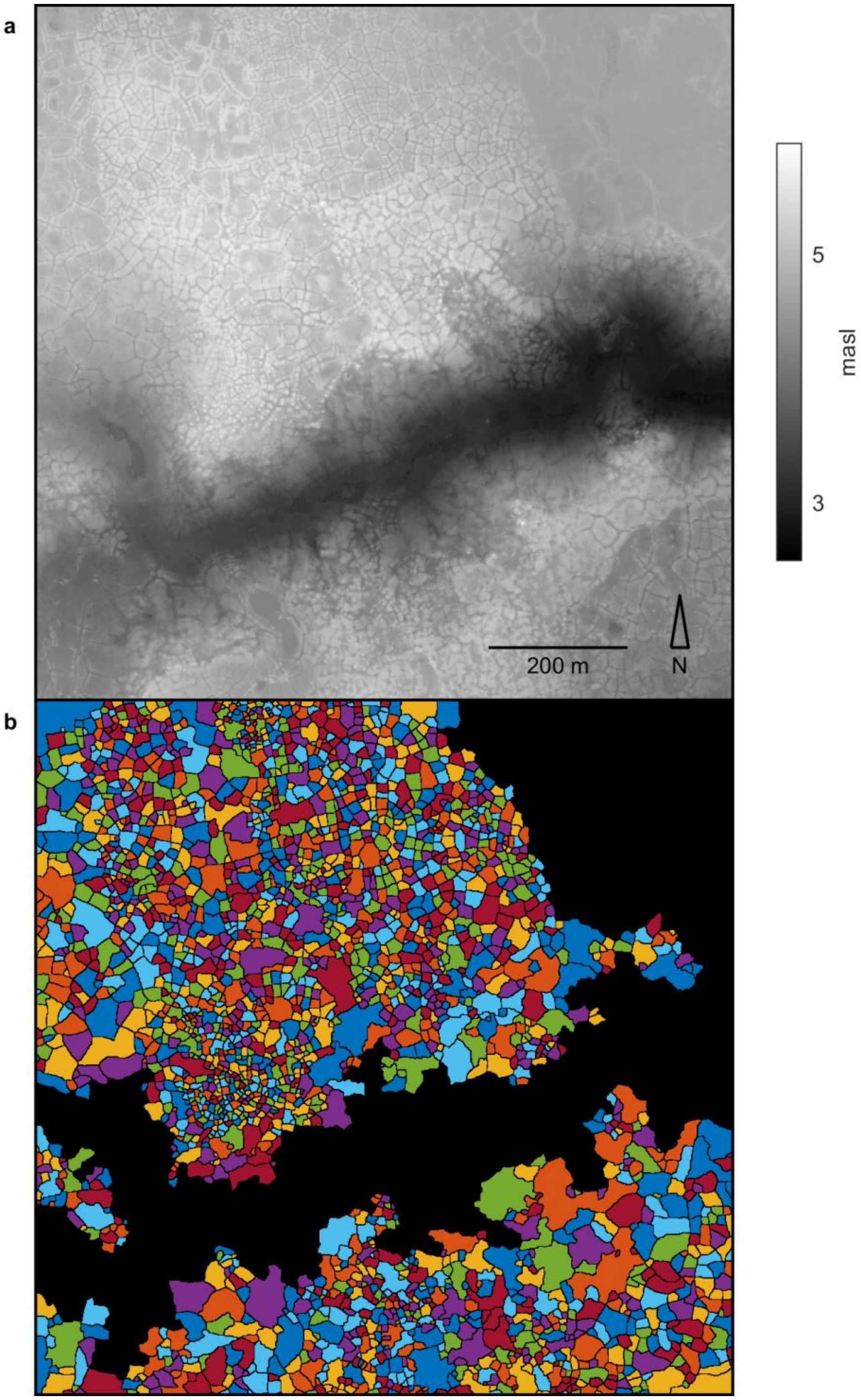

Figure S2. 50 cm DEM (**a**) and polygon delineation (**b**) at site Barrow-1.

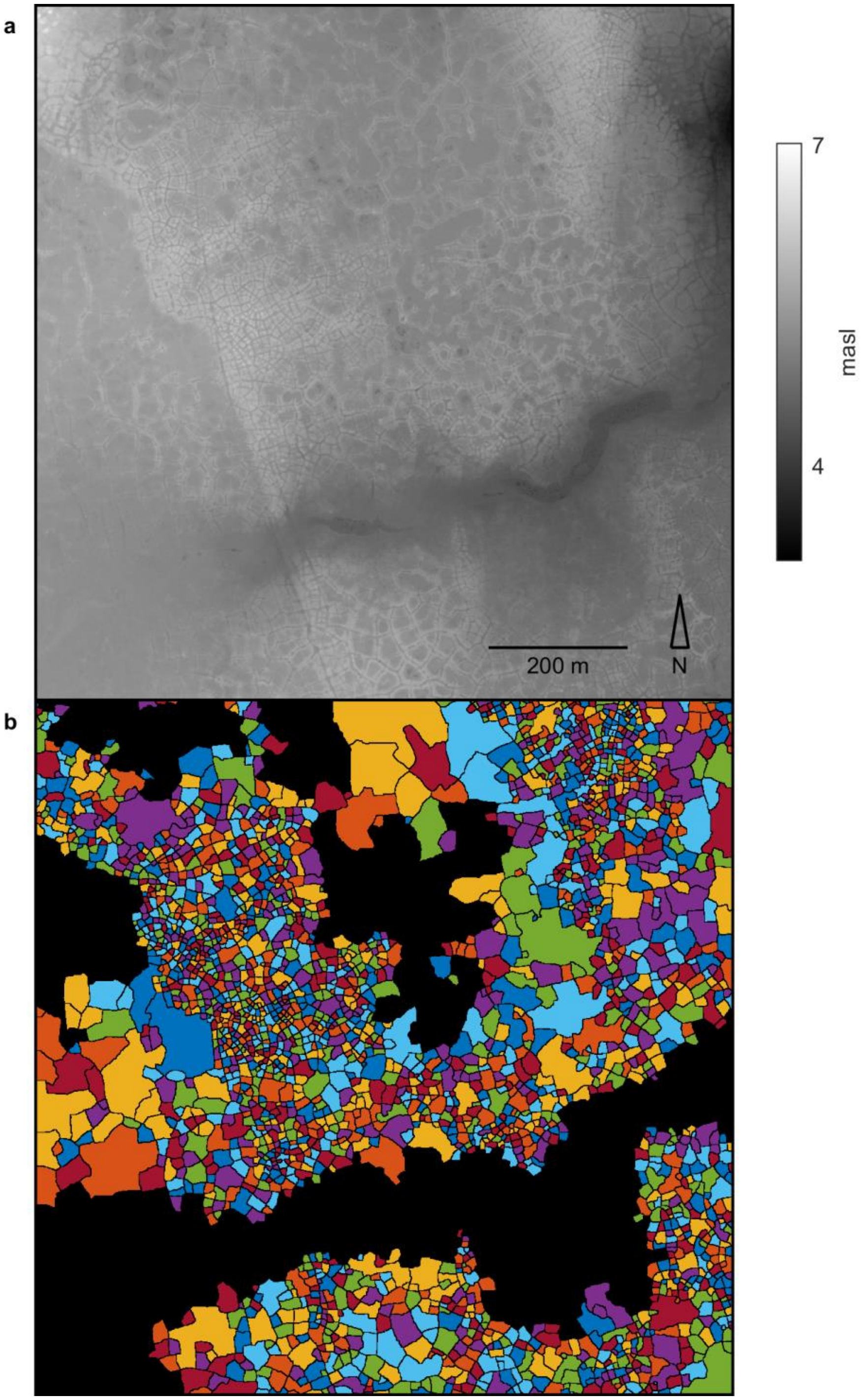

Figure S3. 50 cm DEM (**a**) and polygon delineation (**b**) at site Barrow-2.

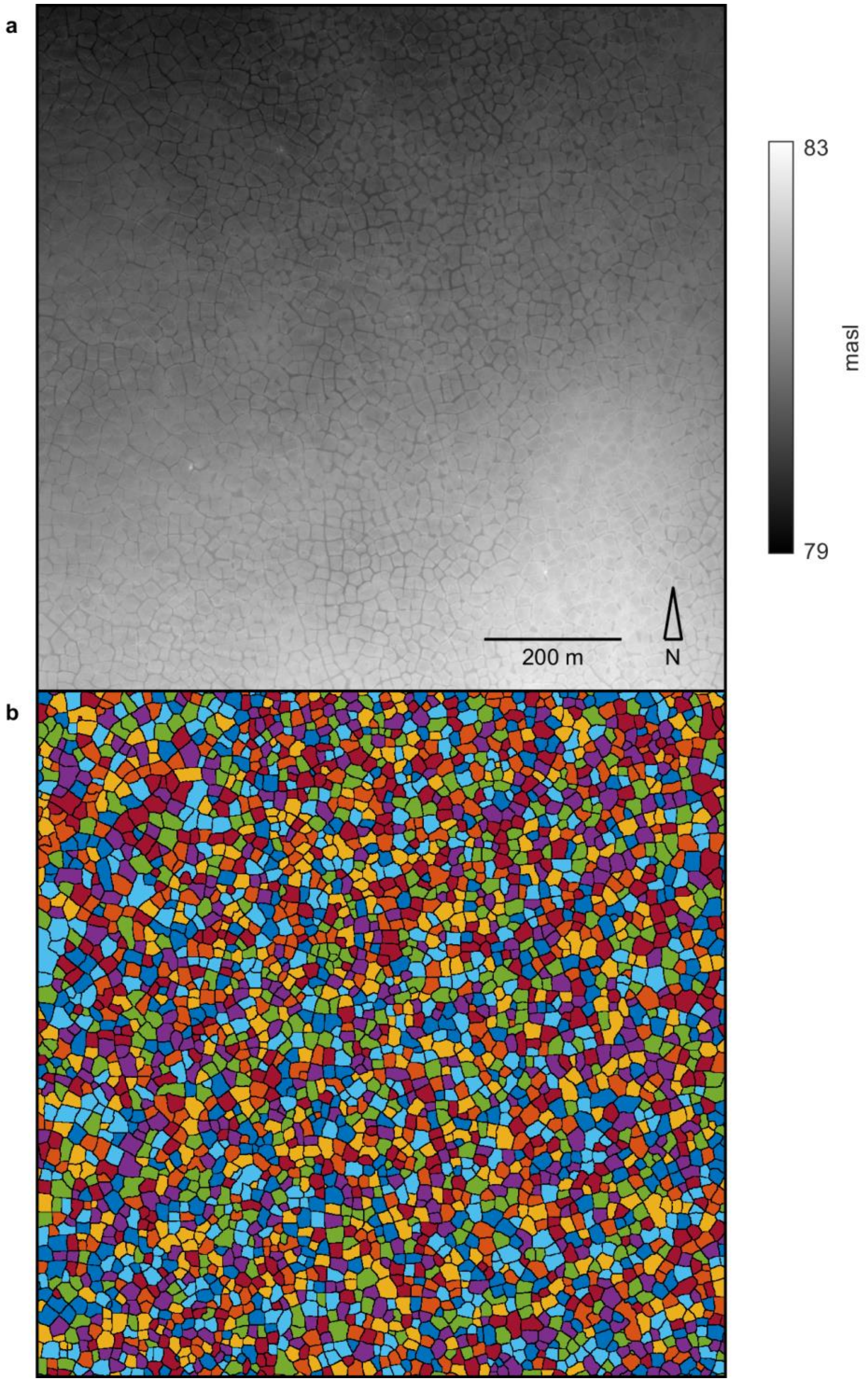

Figure S4. 50 cm DEM (**a**) and polygon delineation (**b**) at site Prudhoe-1.

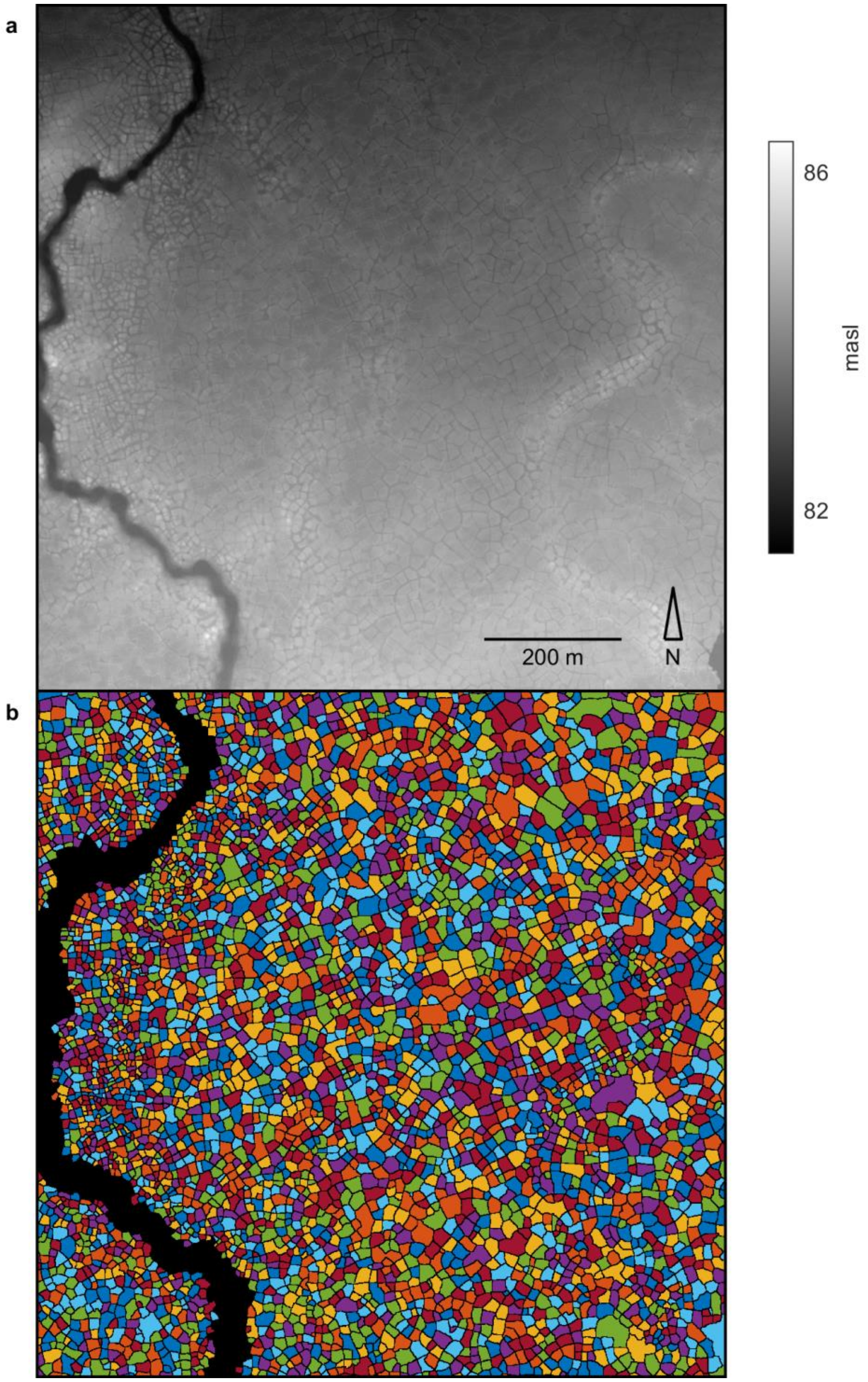

Figure S5. 50 cm DEM (**a**) and polygon delineation (**b**) at site Prudhoe-2.

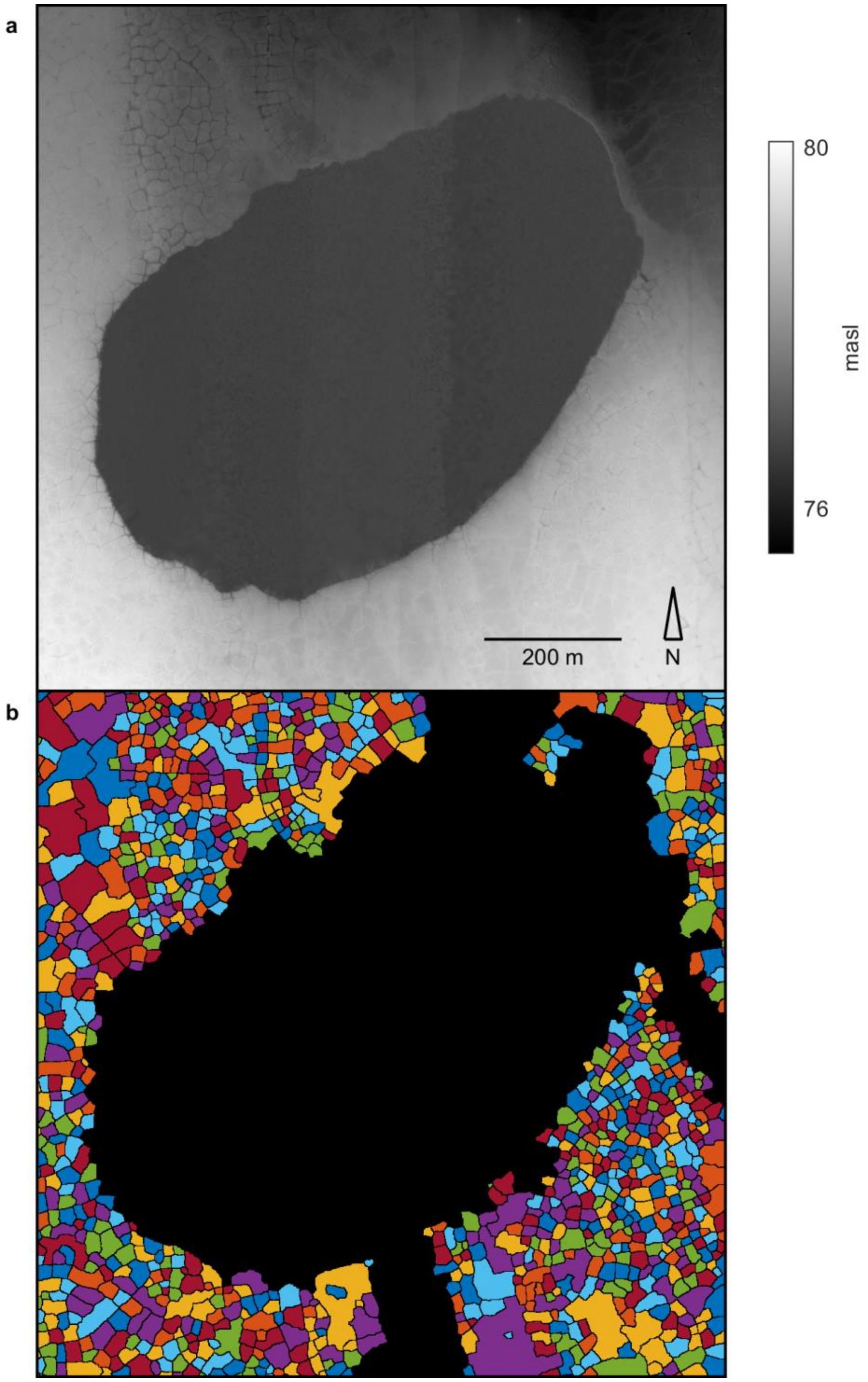

Figure S6. 50 cm DEM (**a**) and polygon delineation (**b**) at site Prudhoe-3.

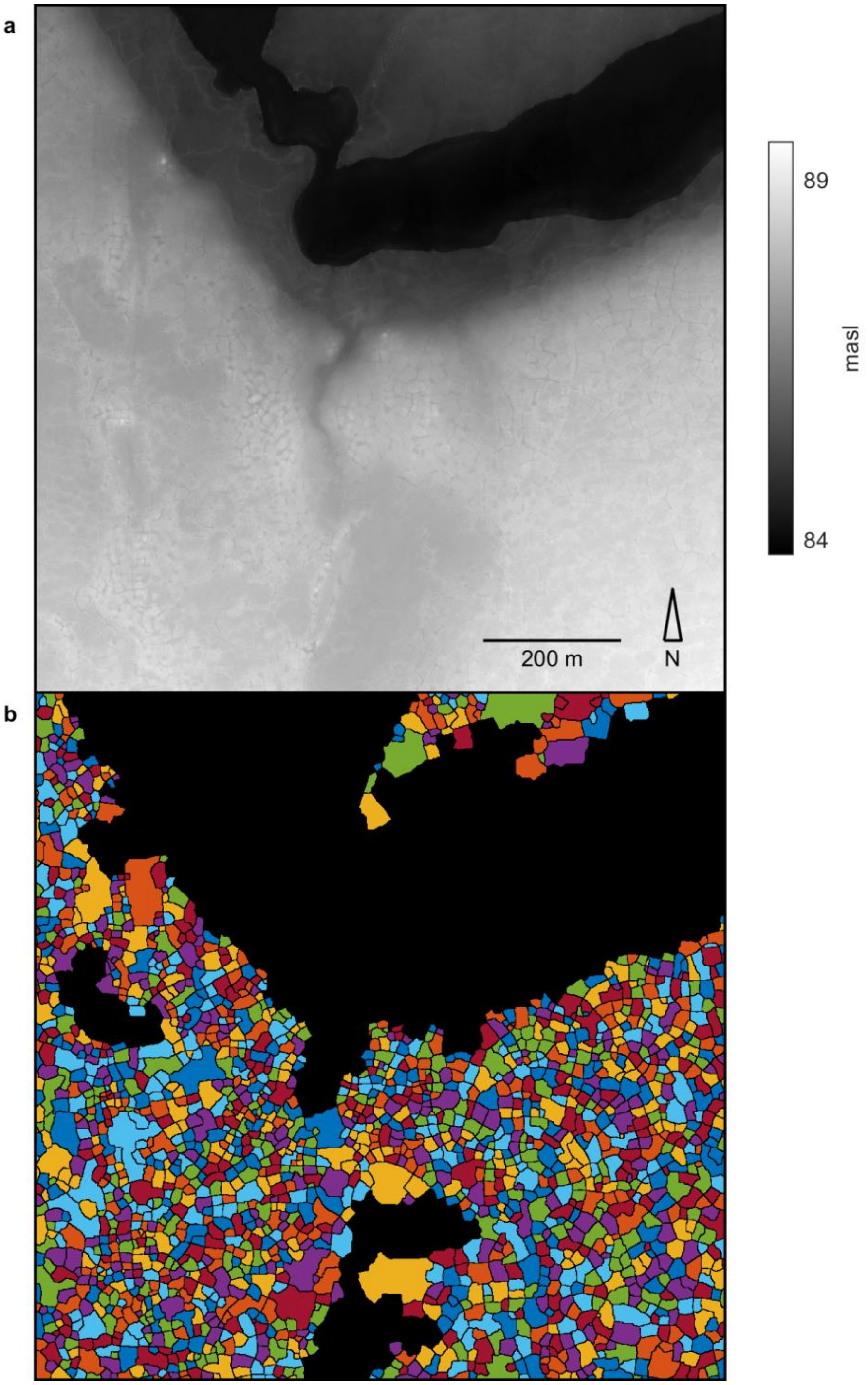

Figure S7. 50 cm DEM (**a**) and polygon delineation (**b**) at site Prudhoe-4.

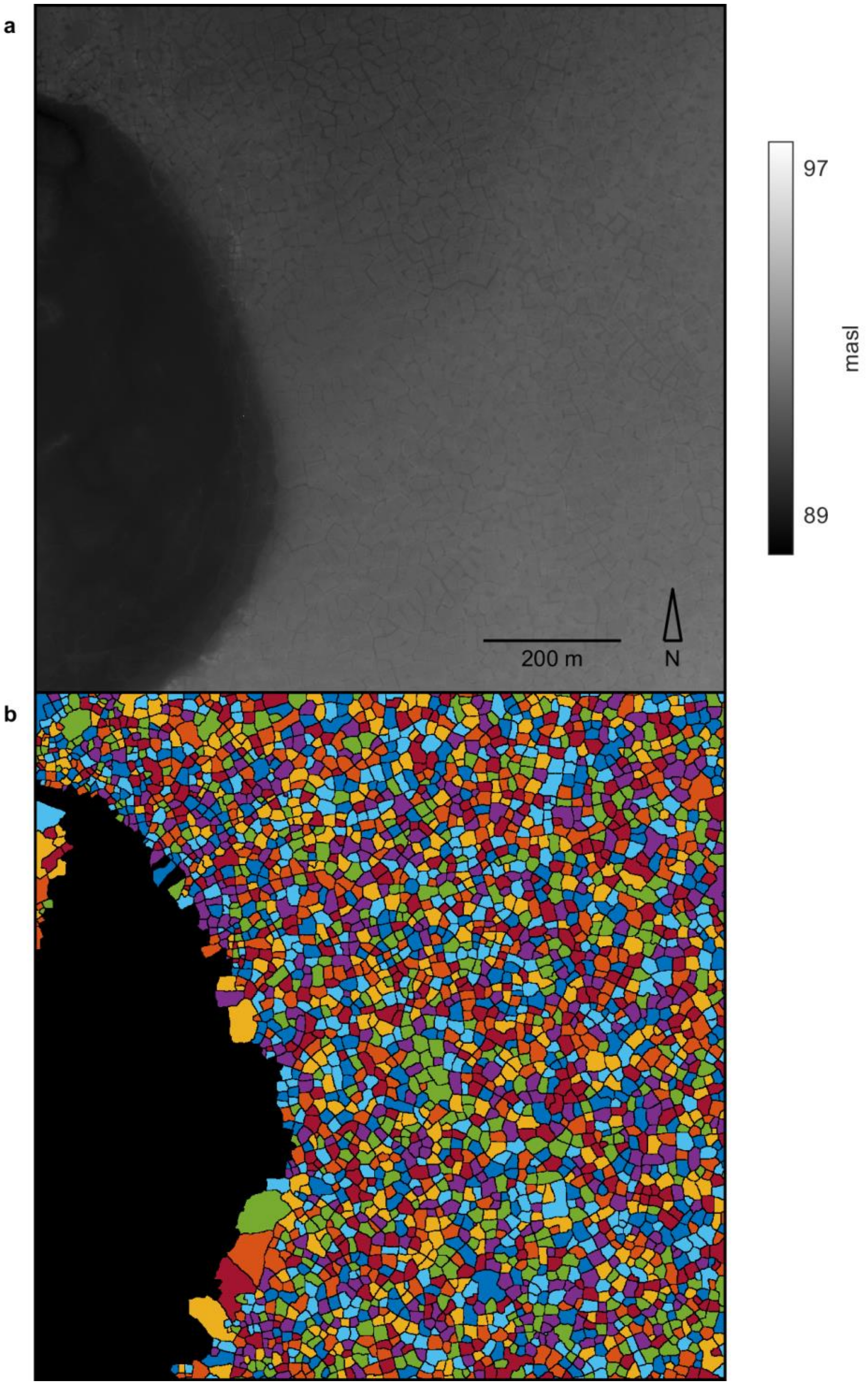

Figure S8. 50 cm DEM (**a**) and polygon delineation (**b**) at site Prudhoe-5.

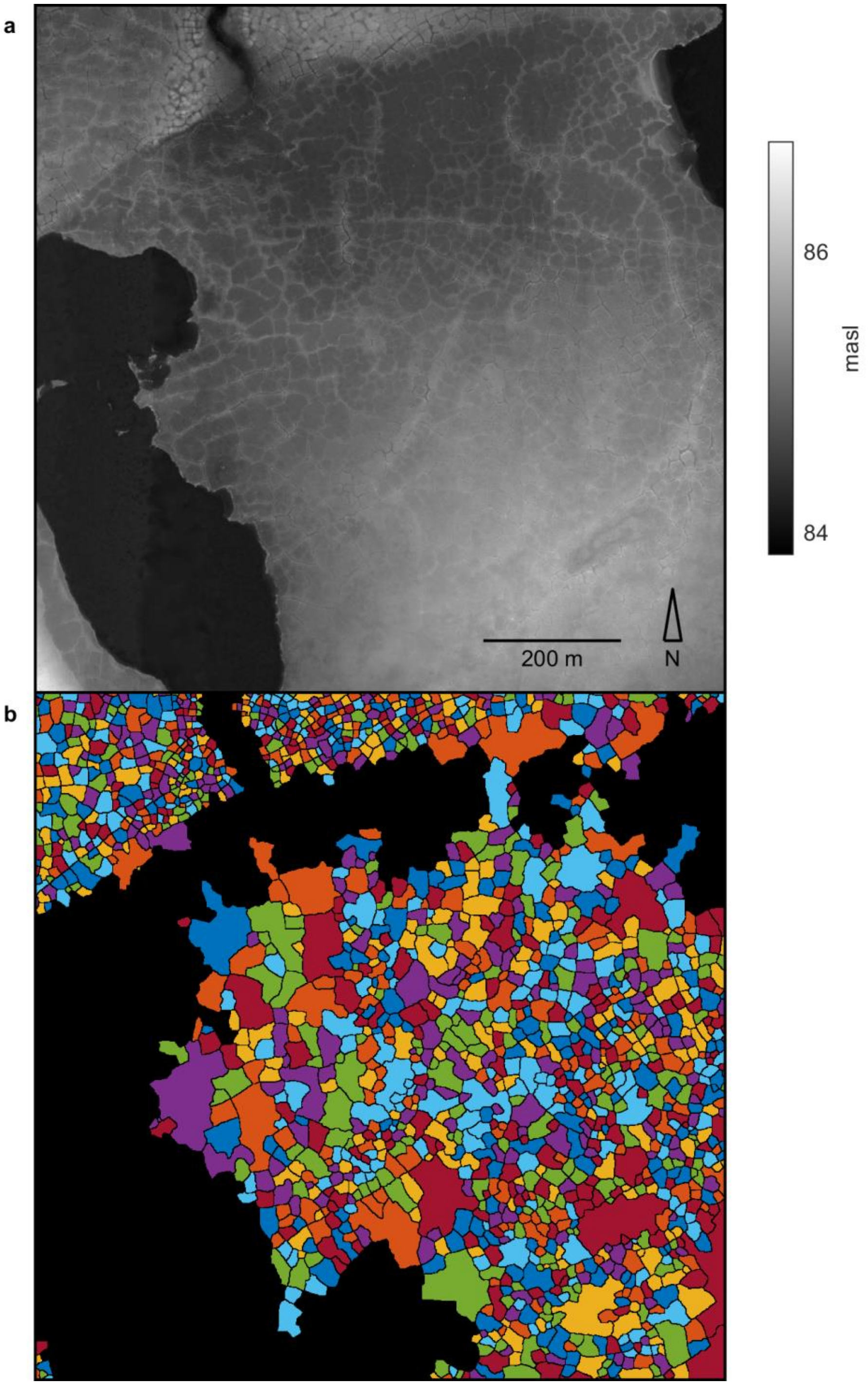

Figure S9. 50 cm DEM (**a**) and polygon delineation (**b**) at site Prudhoe-6.

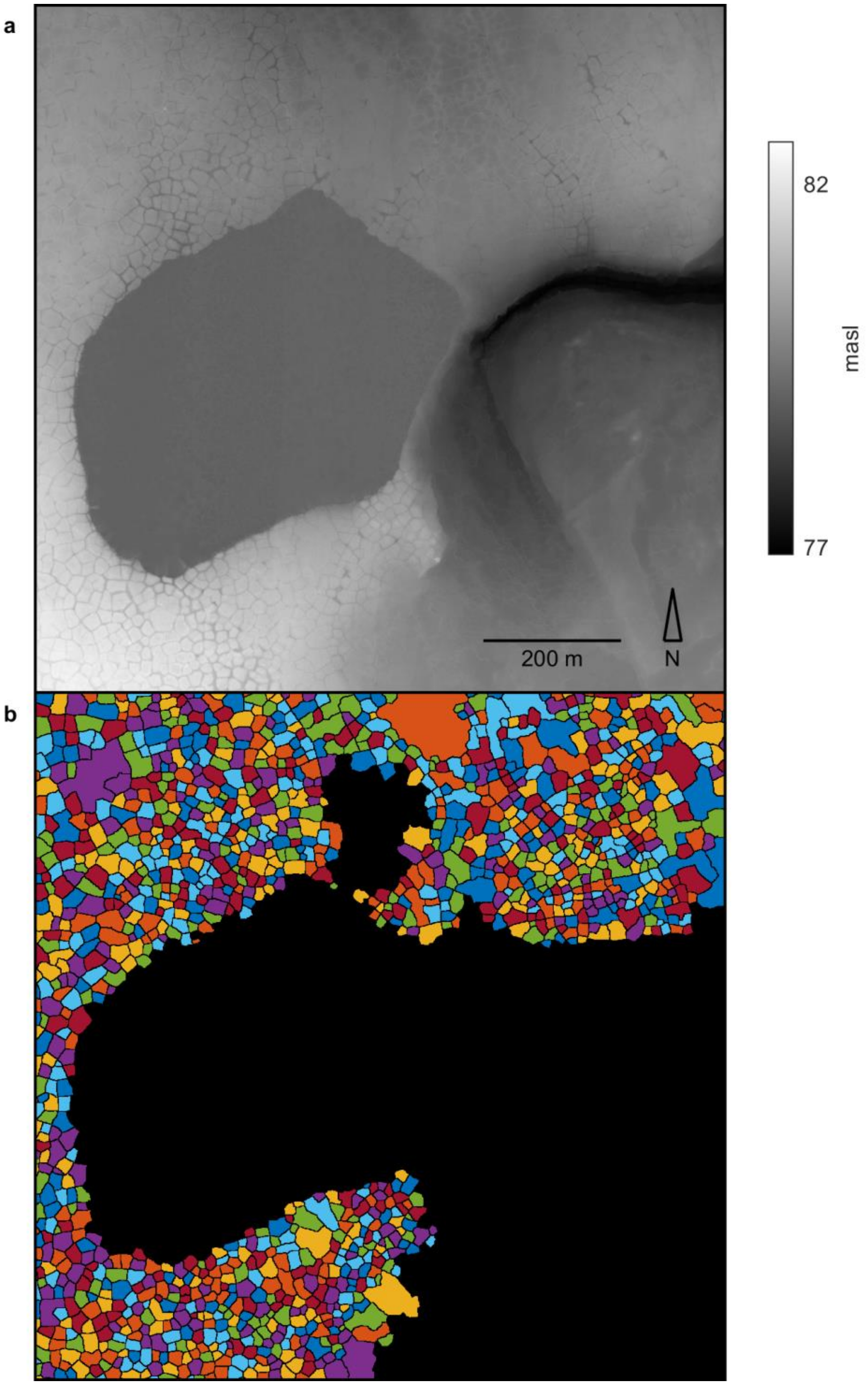

Figure S10. 50 cm DEM (**a**) and polygon delineation (**b**) at site Prudhoe-7.

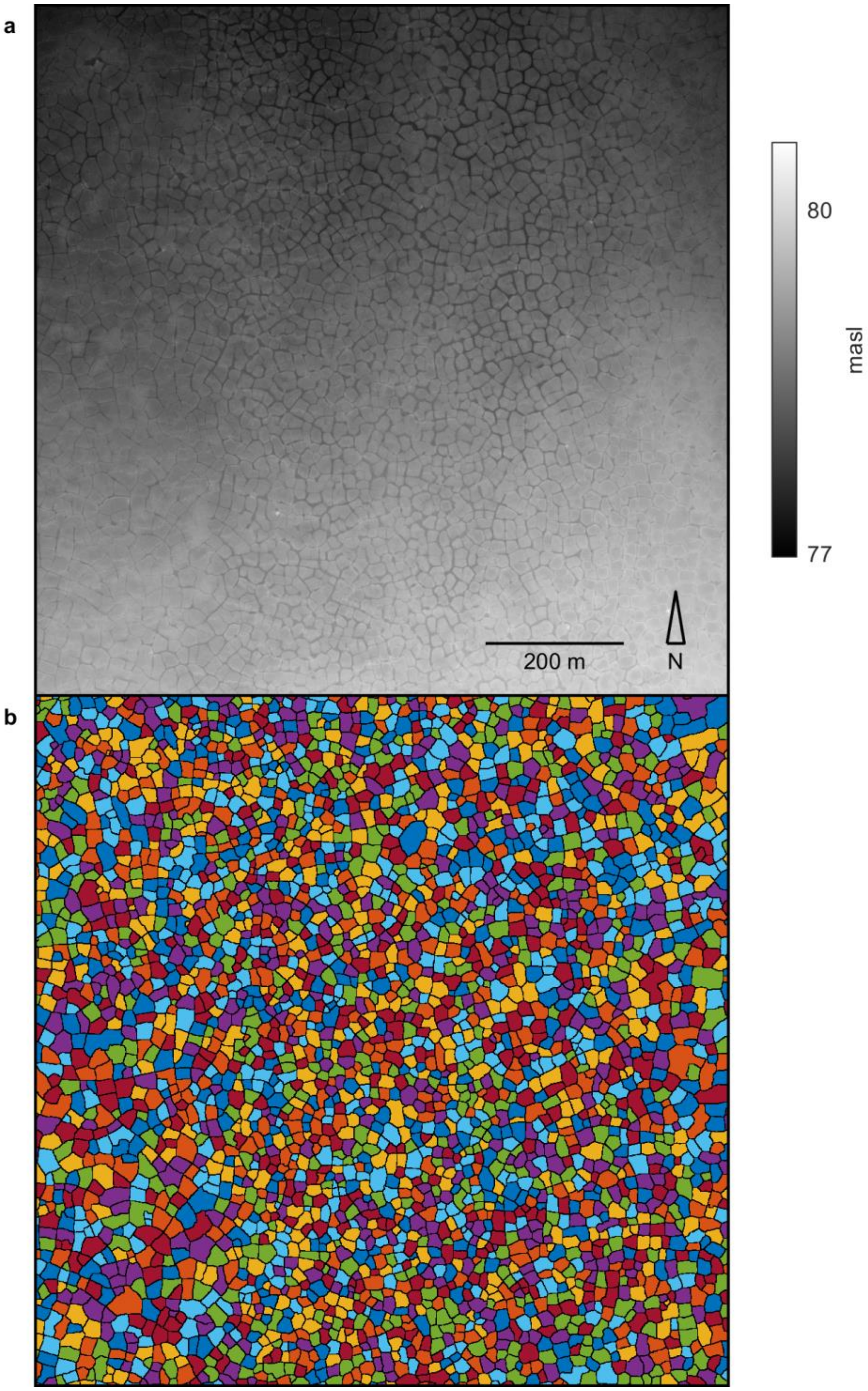

Figure S11. 50 cm DEM (**a**) and polygon delineation (**b**) at site Prudhoe-8.

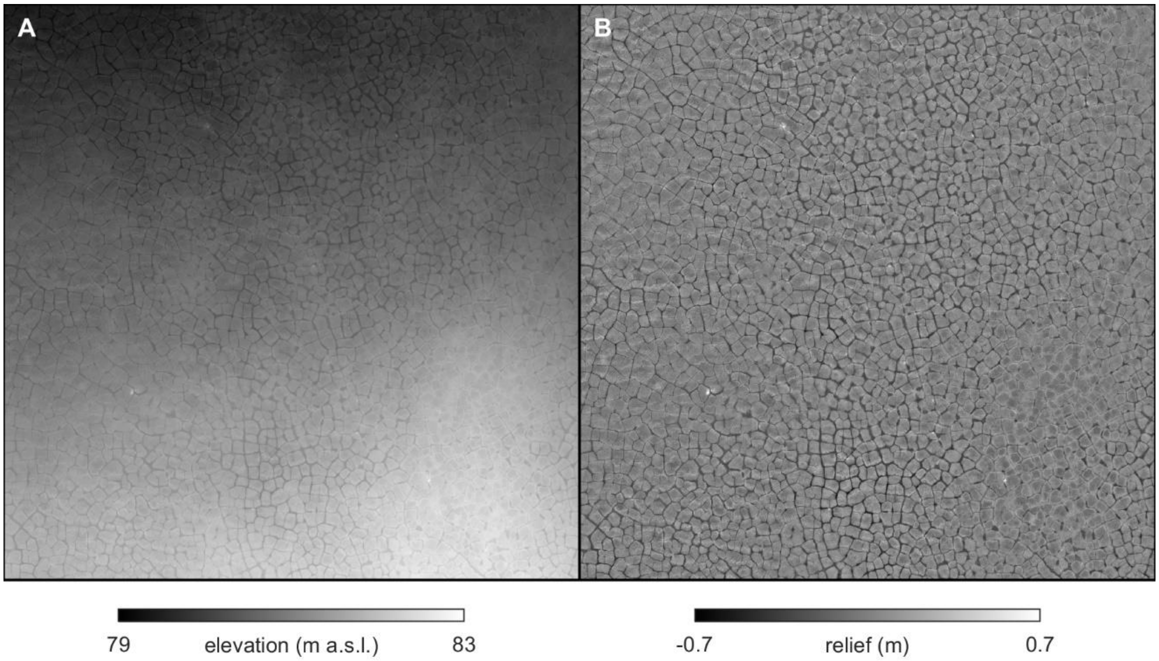

Figure S12. 50 cm DEM of the Prudhoe Bay training site before (**a**) and after (**b**) removing regional trends to isolate microtopography.

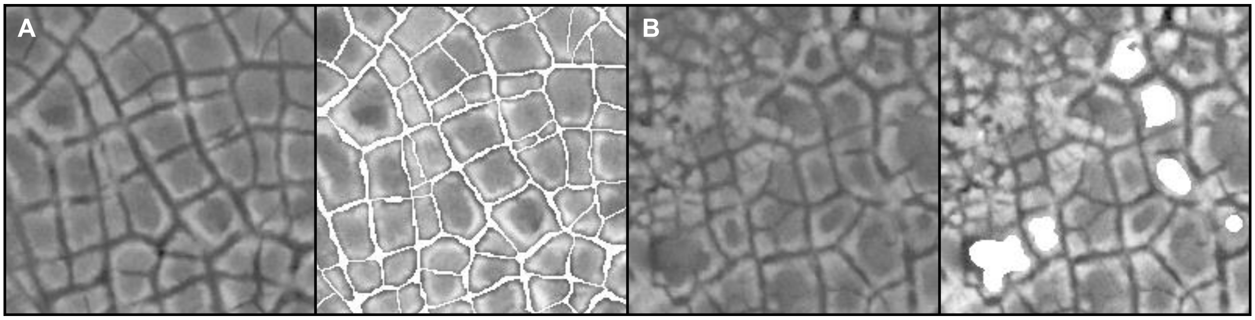

Figure S13. Samples of manually delineated data used to train the CNN, including a tile in which troughs are fully delineated (**a**) and a tile used to supplement the training deck with extra examples of non-trough pixels (**b**).

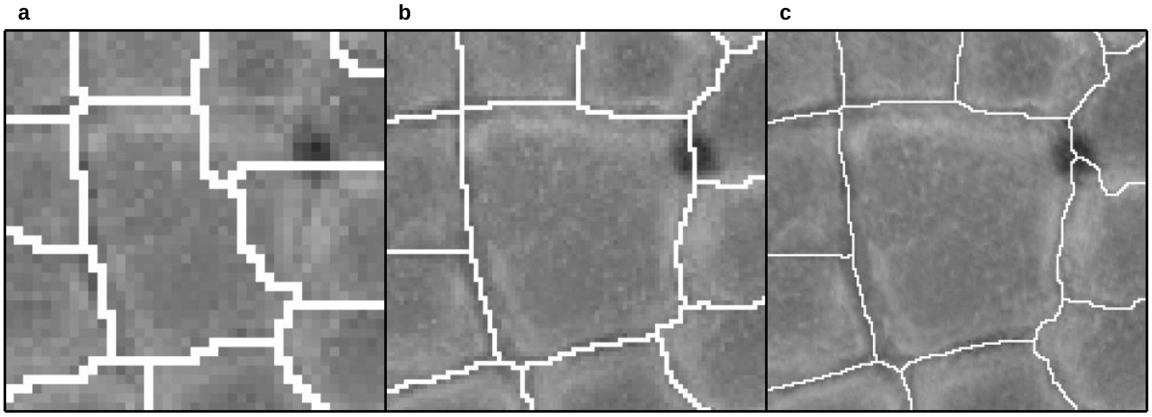

Figure S14. Delineation algorithm on the same ice wedge polygon at 100 cm (**a**), 50 cm (**b**), and 25 cm (**c**) resolution. Each image is 40 m across. Note that anomalously low pixels in the polygon center in (**a**) are mistaken as polygon boundaries, incorrectly fragmenting the polygon.

Text S1. Comparison of training requirements and accuracy between CNN-watershed and Mask R-CNN algorithms.

Due to differences in the training and inference procedures used by each algorithm, training data requirements and accuracy are difficult to compare directly. Nonetheless, in several aspects, performance appears to be similar. In the present study, the CNN-watershed approach is trained initially on data derived from four manually-labeled 100 × 100 m tiles, representing 0.04 km^2. This training data is supplemented with extra examples of boundary and non-boundary features, the convex hulls of which sum to ~0.07 km^2, and the trained model is extrapolated across 10 km^2. The training to application ratio is therefore ~0.011, or 1.1%. In comparison, Mask R-CNN was trained on data from 340 90 × 90 m tiles, or ~2.75 km^2, then extrapolated across ~134 km^2, resulting in a training to application ratio of ~0.020 or 2.0% (Zhang et al., 2018). In general, within the area across which the CNN-watershed approach was applied, it was less likely than Mask R-CNN to fail to detect polygonal terrain, but more prone to mistakenly aggregate multiple ice wedge polygons into a single unit. These errors were particularly common at sites characterized by transitional terrain where ice wedge polygons grade into non-polygonal ground. It is reasonable to expect such mistakes in these areas, as microtopography is typically faint and polygons often appear to be bound incompletely by troughs. At one such site (Prudhoe-6), the number of incorrect conglomerate polygons by area delineated by the CNN-watershed algorithm was ~22% (Table 1). This number closely resembles the 21% of human-delineated polygons estimated to go undetected by Mask R-CNN in satellite-based optical imagery (Zhang et al., 2018).

Table S1. Architecture of our CNN.

Layer	Type	Neurons
1	Convolutional	8 arrays of 27×27
2	ReLU†	8 arrays of 27×27
3	Max-pooling	8 arrays of 9×9
4	ReLU	8 arrays of 9×9
5	Fully-connected	64
6	ReLU	64
7	Fully-connected	2
8	ReLU	2
9	Softmax	2

† - ReLu – rectified linear unit

Table S2. Results of manual validation at 100 cm and 25 cm resolution (sites are 1 km^2).

Site	Polygons identified	Polygonal area (%)	*% of polygons by instance*				*% of polygons by area*			
			Whole	Fractional	Conglomerate	Non-polygonal	Whole	Fractional	Conglomerate	Non-polygonal
Barrow-1 (100 cm)	3058	74.3	73.4	18.6	6.2	1.8	65.5	16.6	13.4	4.1
Prudhoe-1 (100 cm)	3019	100	85.6	11.8	2.6	0.0	88.0	79.7	4.0	0.0
Barrow-1 (25 cm)	2870	71.6	89.0	3.8	3.4	3.8	83.4	2.0	9.7	4.8
Prudhoe-1 (25 cm)	3193	100	93.6	3.4	2.4	0.6	94.0	1.6	4.3	0.1

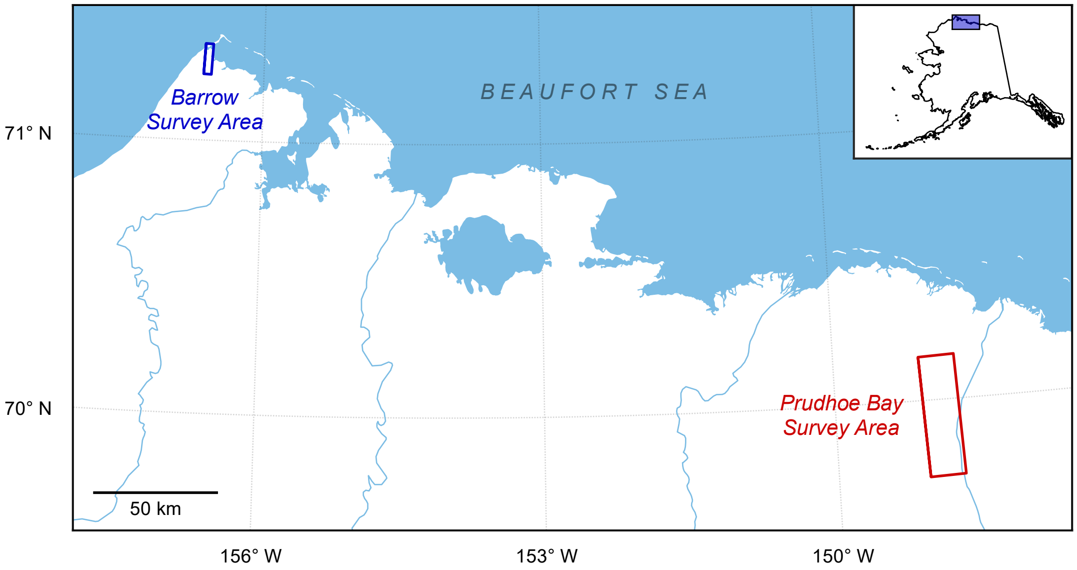

Figure S1. Bounding boxes of airborne lidar surveys~~ in Alaska Albers (NAD83) projection~~.

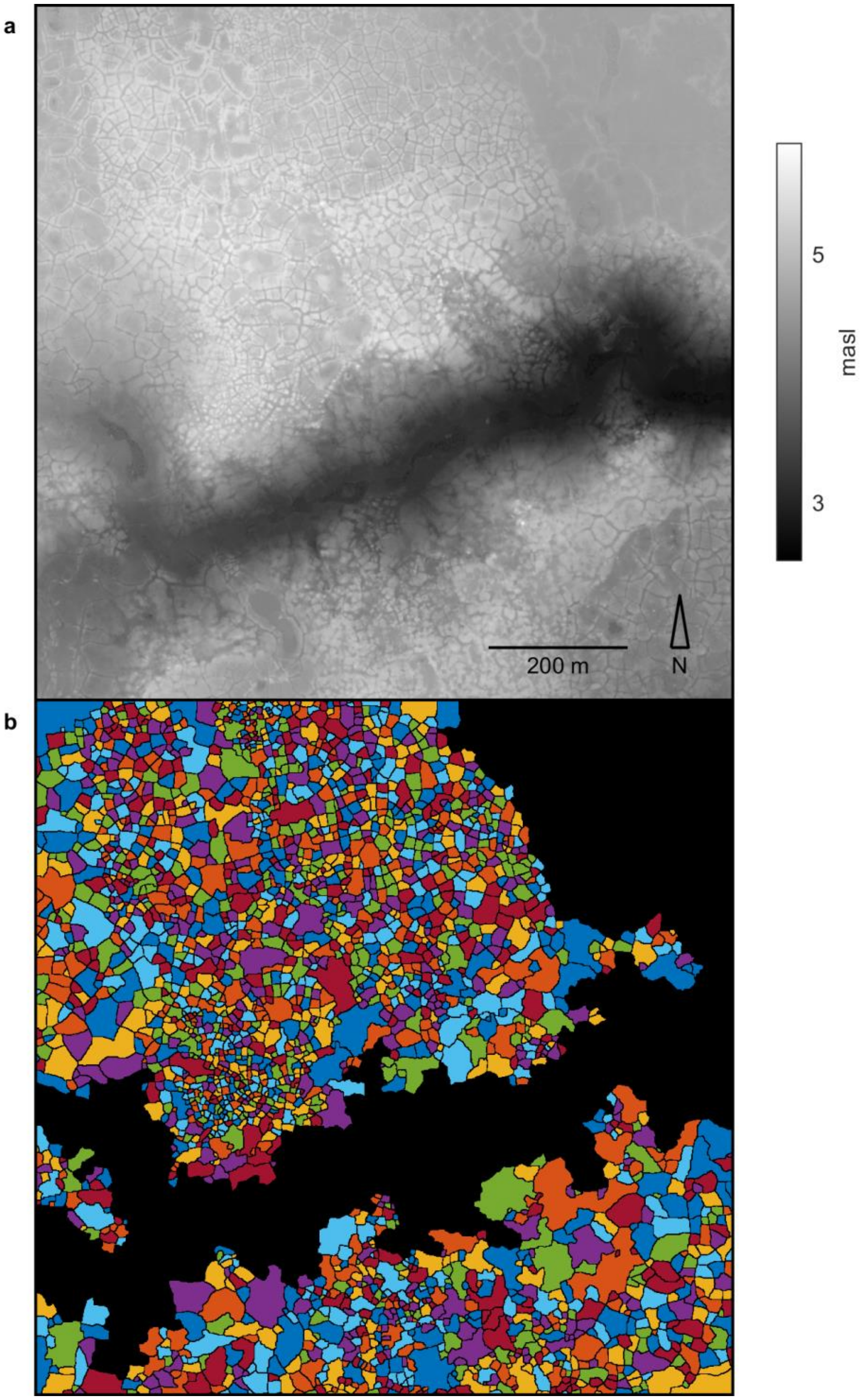

Figure S2. 50 cm DEM (**a**) and polygon delineation (**b**) at site Barrow-1.

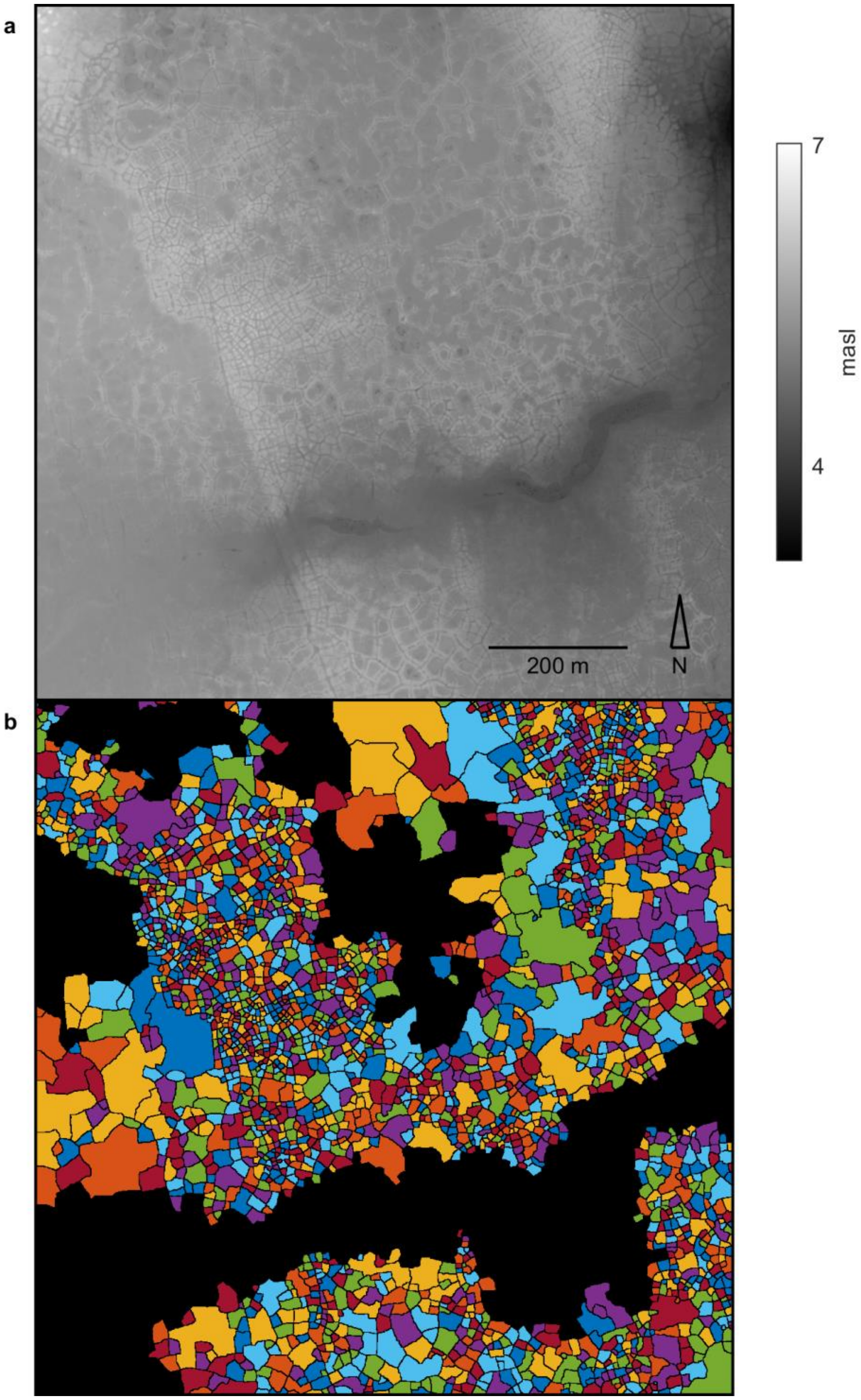

Figure S3. 50 cm DEM (**a**) and polygon delineation (**b**) at site Barrow-2.

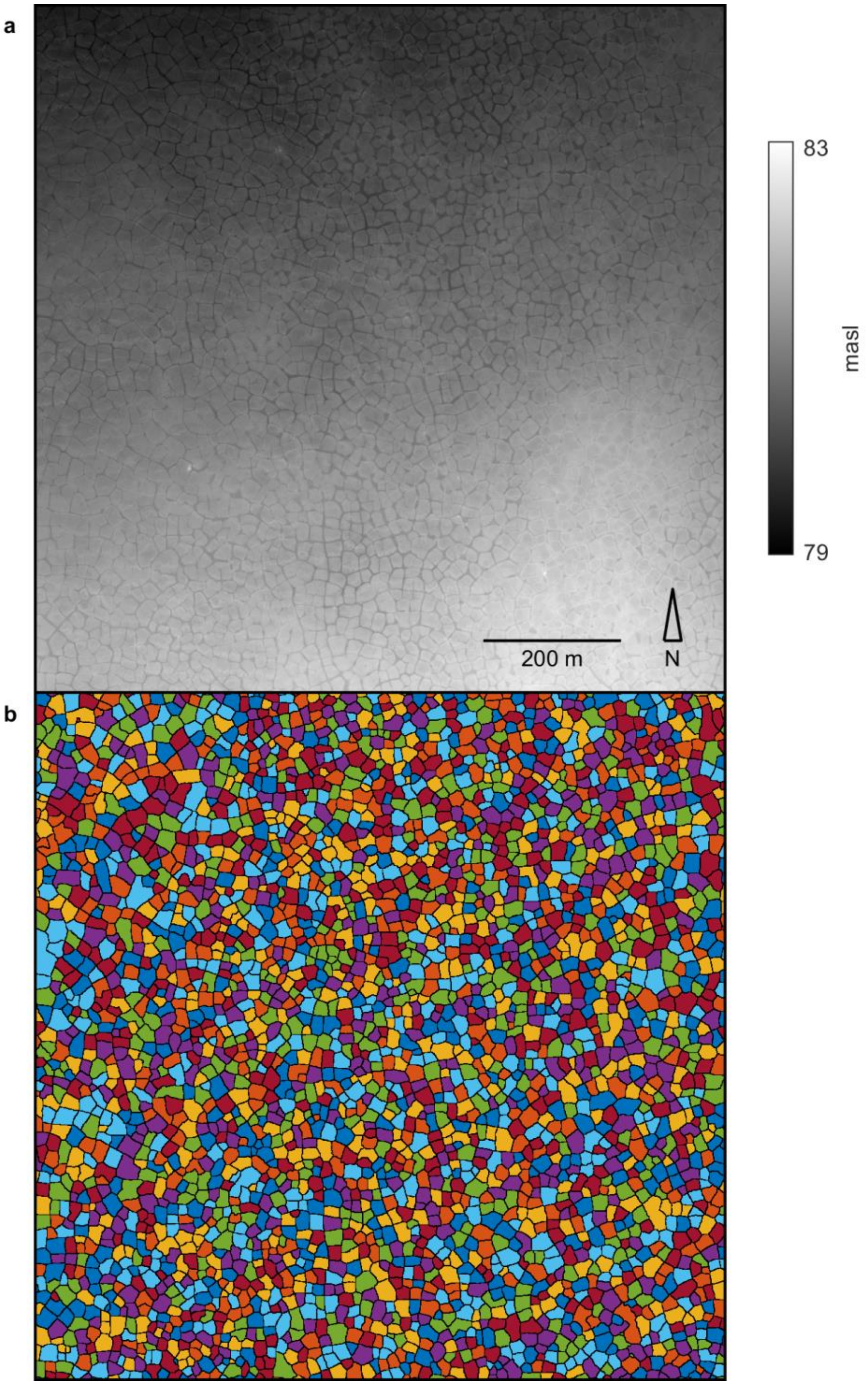

Figure S4. 50 cm DEM (**a**) and polygon delineation (**b**) at site Prudhoe-1.

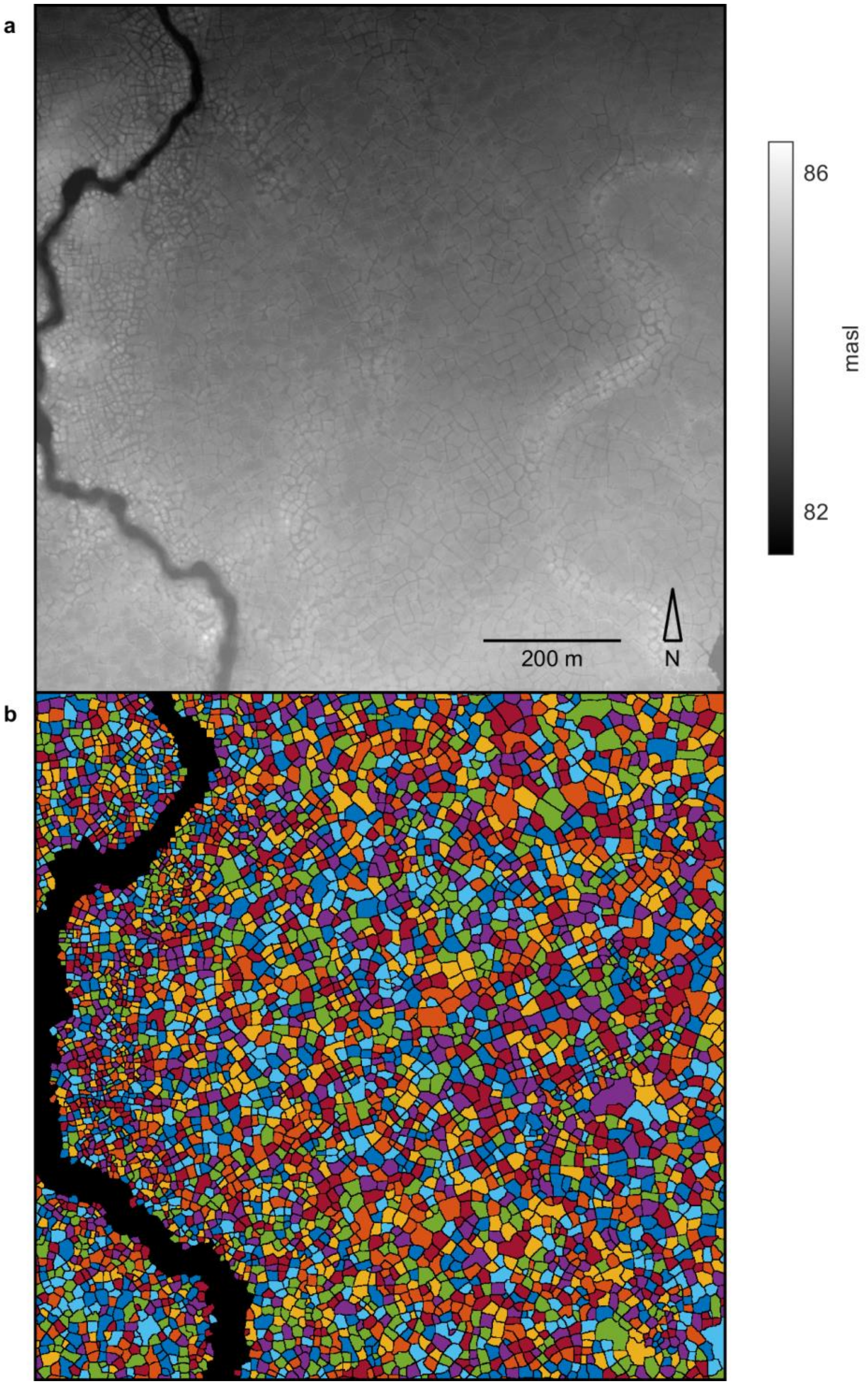

Figure S5. 50 cm DEM (**a**) and polygon delineation (**b**) at site Prudhoe-2.

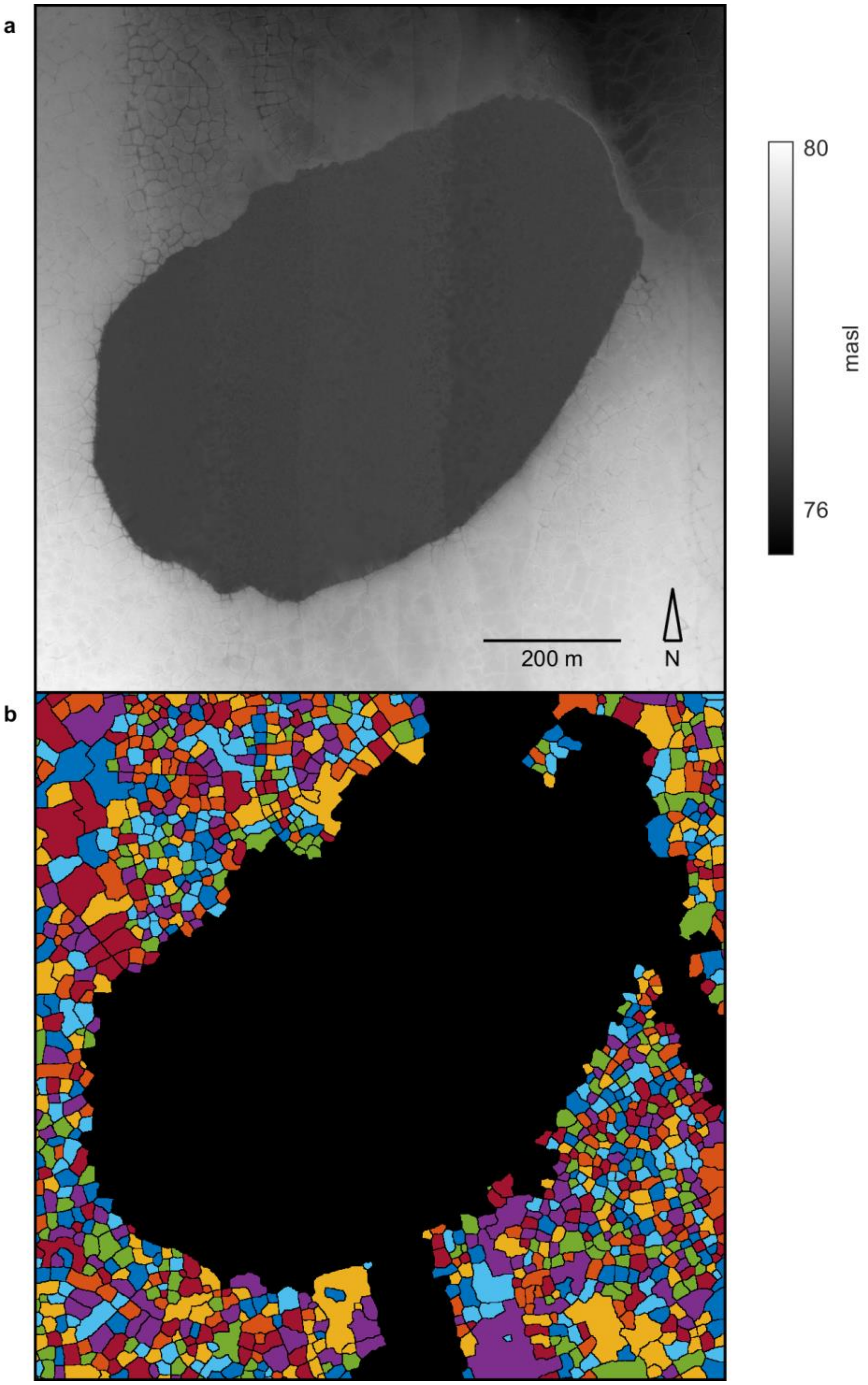

Figure S6. 50 cm DEM (**a**) and polygon delineation (**b**) at site Prudhoe-3.

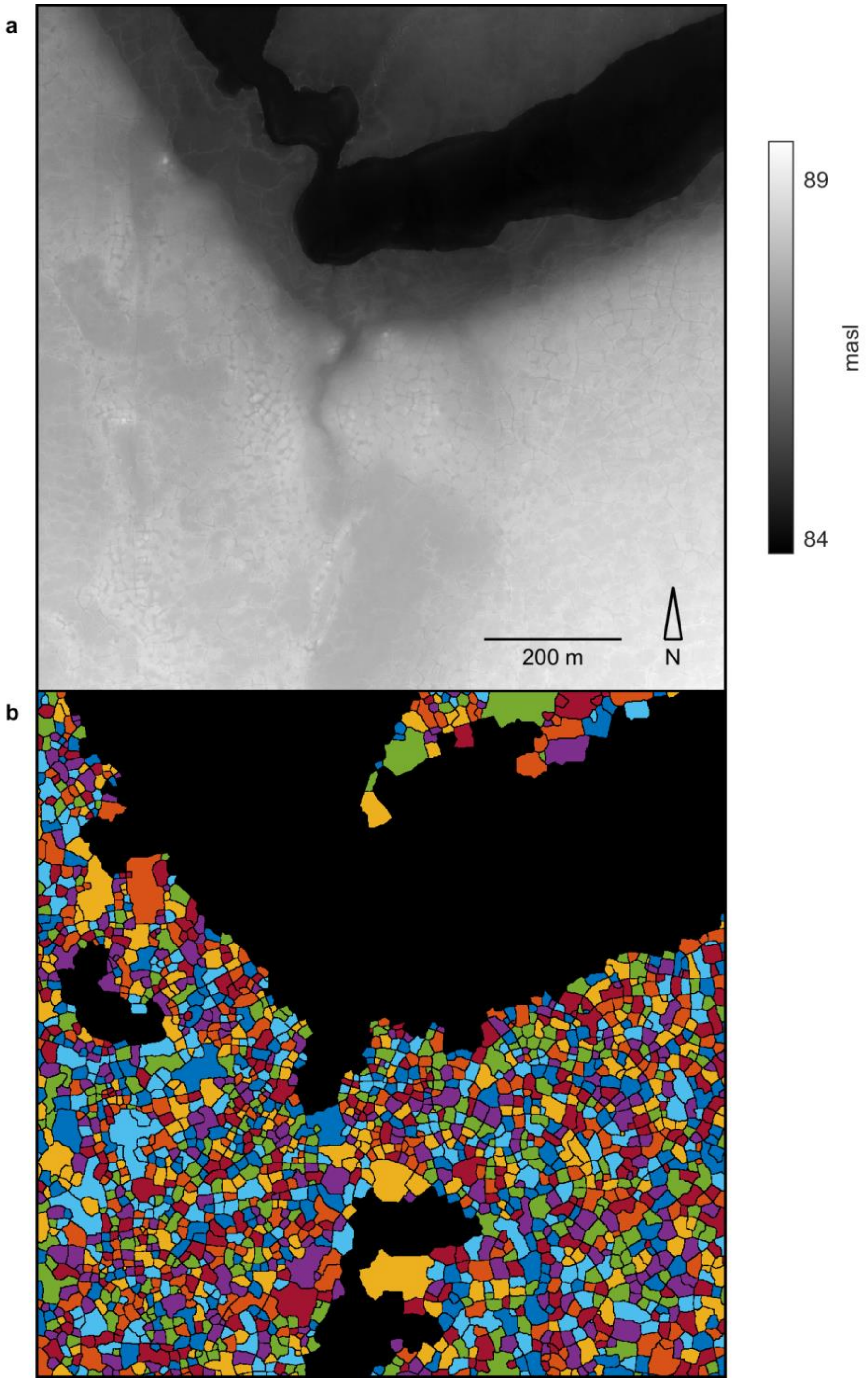

Figure S7. 50 cm DEM (**a**) and polygon delineation (**b**) at site Prudhoe-4.

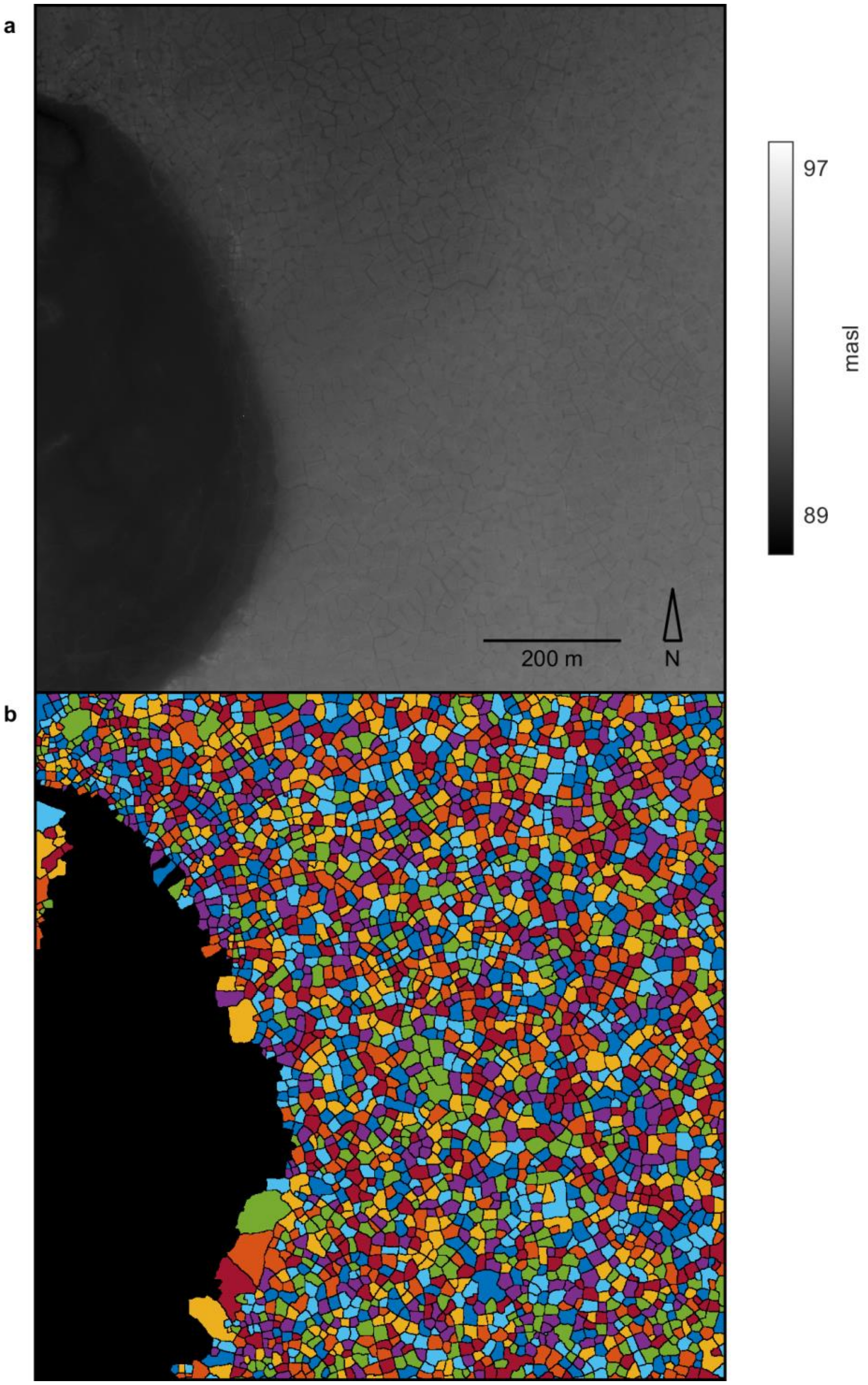

Figure S8. 50 cm DEM (**a**) and polygon delineation (**b**) at site Prudhoe-5.

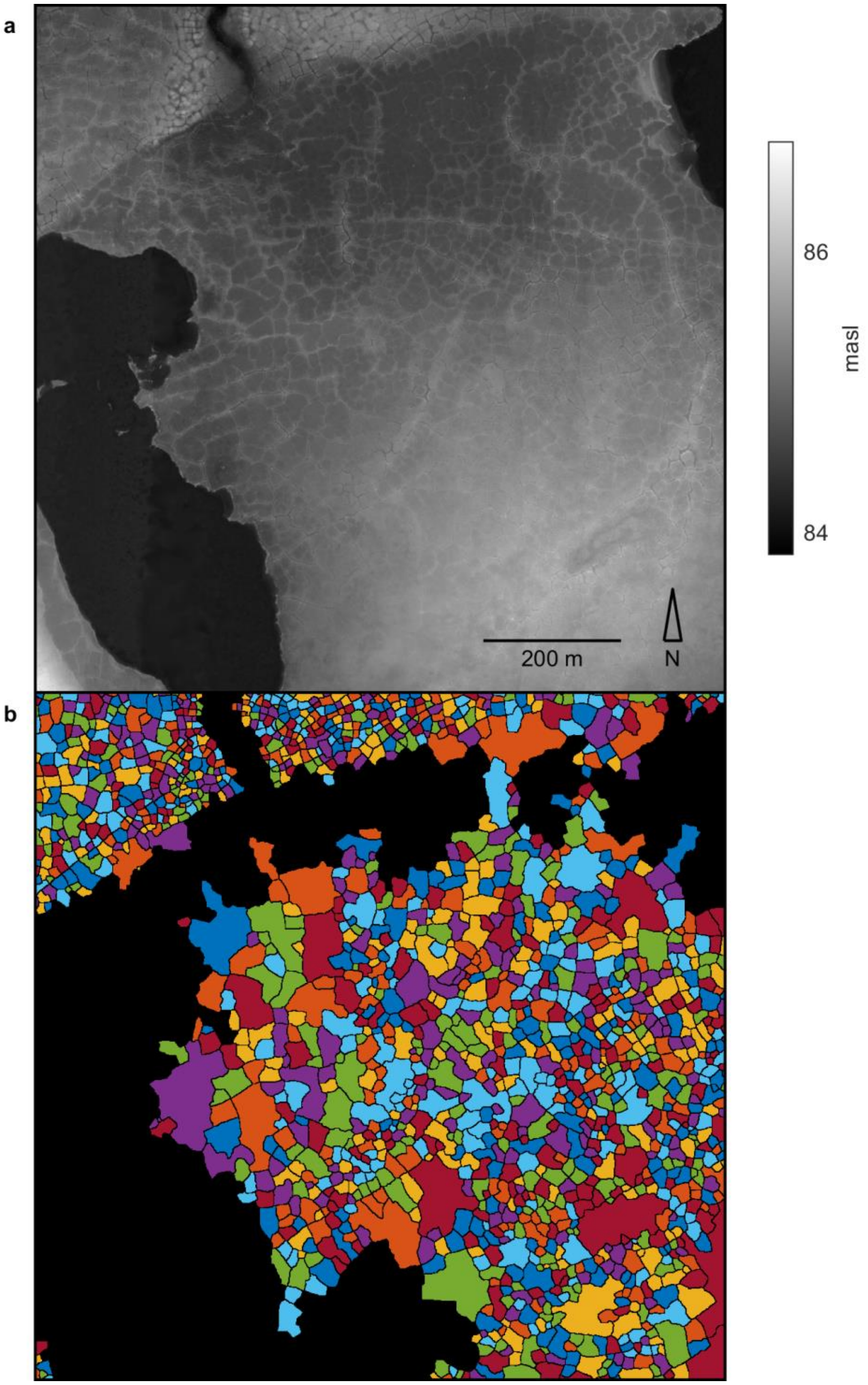

Figure S9. 50 cm DEM (**a**) and polygon delineation (**b**) at site Prudhoe-6.

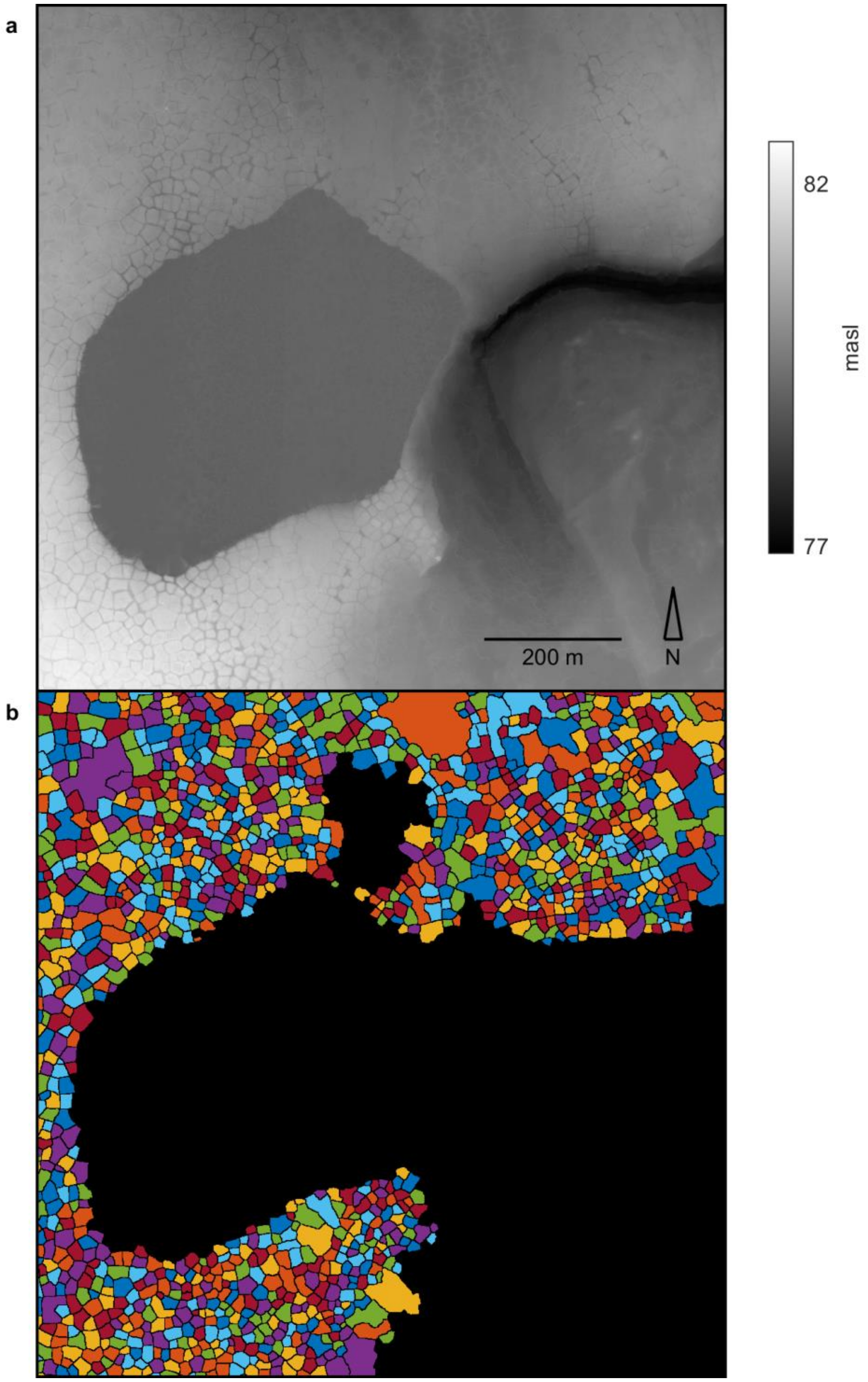

Figure S10. 50 cm DEM (**a**) and polygon delineation (**b**) at site Prudhoe-7.

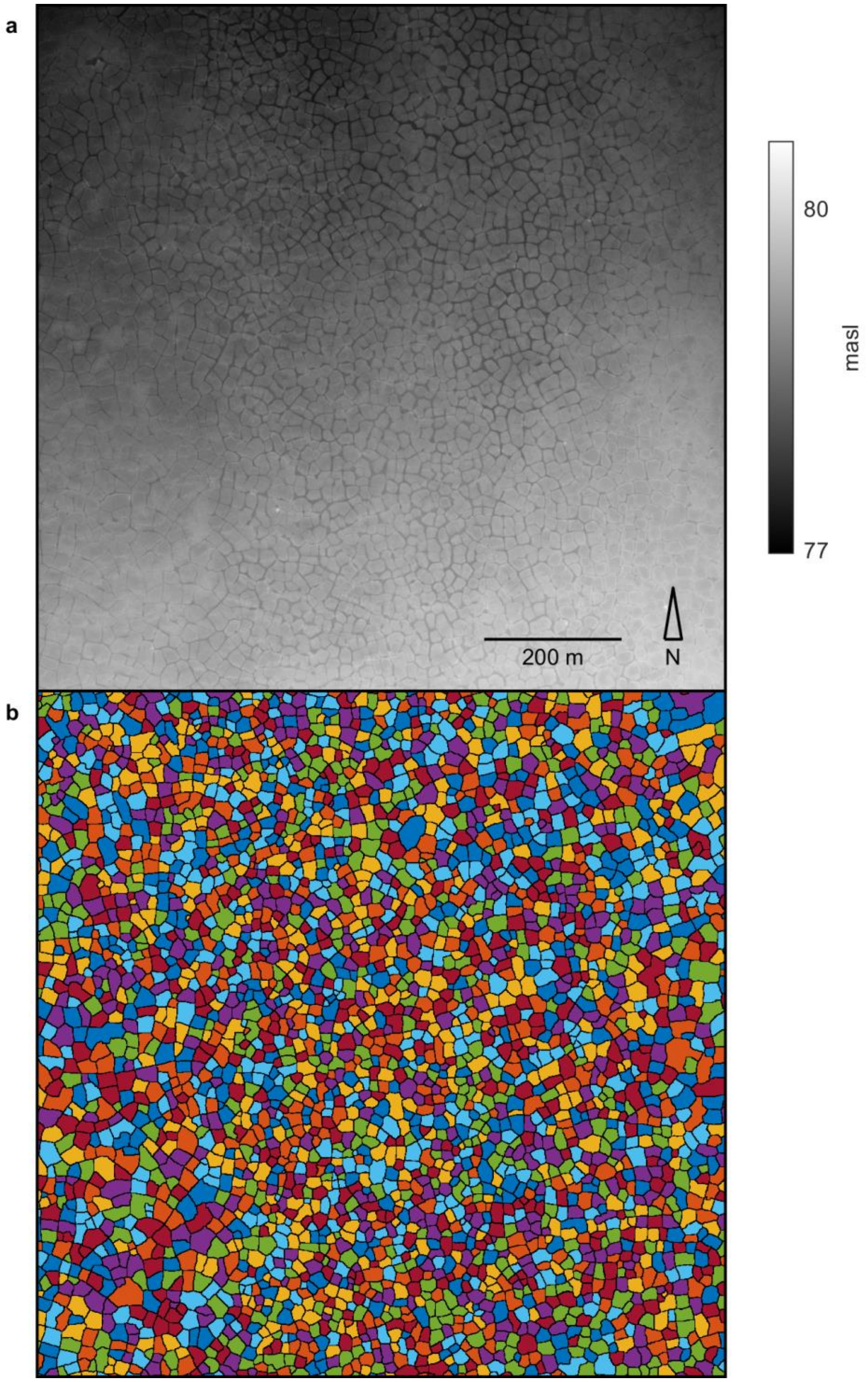

Figure S11. 50 cm DEM (**a**) and polygon delineation (**b**) at site Prudhoe-8.

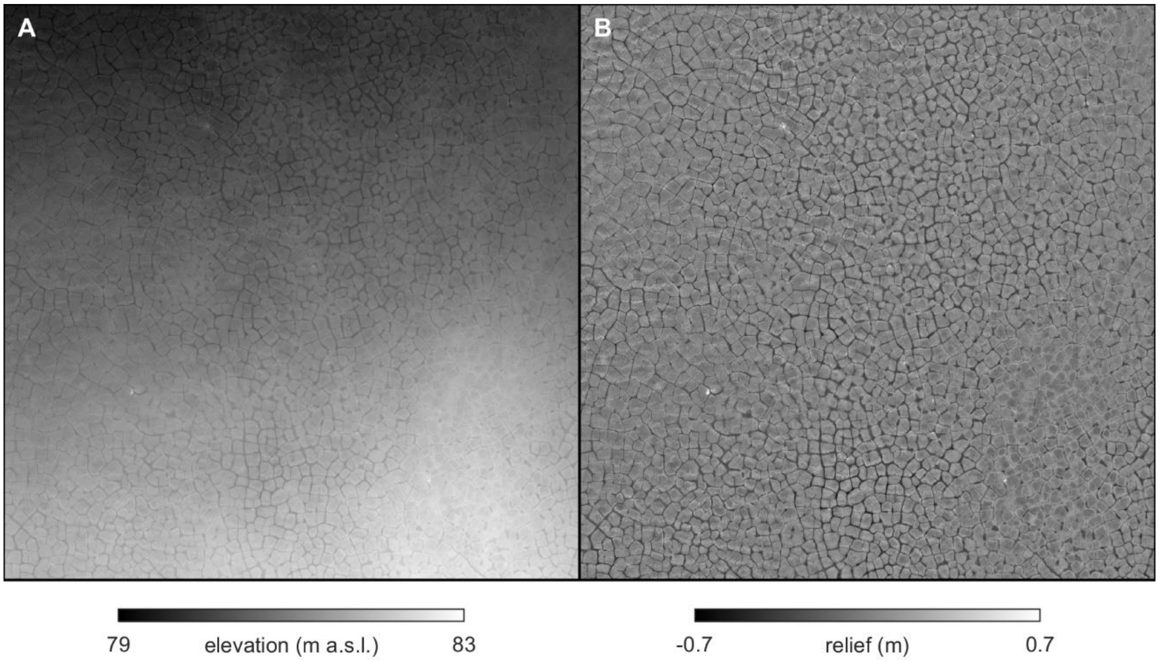

Figure S12. 50 cm DEM of the Prudhoe Bay training site before (**a**) and after (**b**) removing regional trends to isolate microtopography.

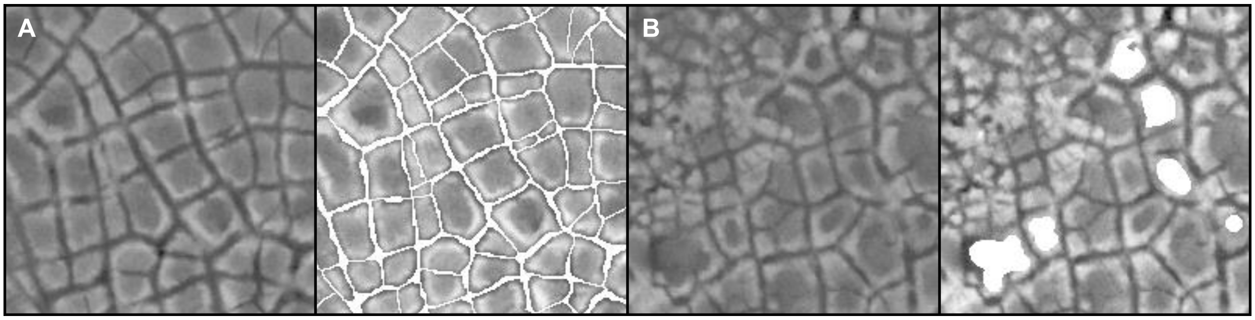

Figure S13. Samples of manually delineated data used to train the CNN, including a tile in which troughs are fully delineated (**a**) and a tile used to supplement the training deck with extra examples of non-trough pixels (**b**).

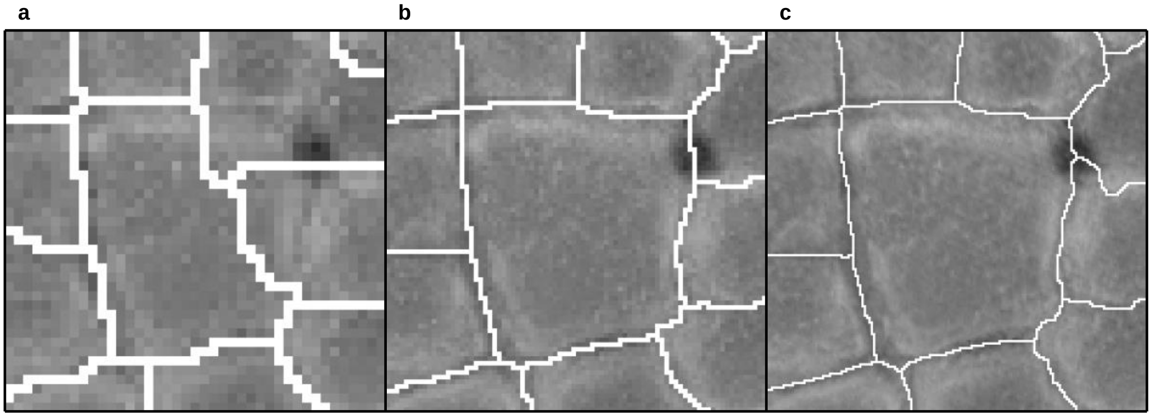

Figure S14. Delineation algorithm on the same ice wedge polygon at 100 cm (**a**), 50 cm (**b**), and 25 cm (**c**) resolution. Each image is 40 m across. Note that anomalously low pixels in the polygon center in (**a**) are mistaken as polygon boundaries, incorrectly fragmenting the polygon.

Text S1. Comparison of training requirements and accuracy between CNN-watershed and Mask R-CNN algorithms.

Due to differences in the training and inference procedures used by each algorithm, training data requirements and accuracy are difficult to compare directly. Nonetheless, in several aspects, performance appears to be similar. In the present study, the CNN-watershed approach is trained initially on data derived from four manually-labeled 100 × 100 m tiles, representing 0.04 km^2. This training data is supplemented with extra examples of boundary and non-boundary features, the convex hulls of which sum to ~0.07 km^2, and the trained model is extrapolated across 10 km^2. The training to application ratio is therefore ~0.011, or 1.1%. In comparison, Mask R-CNN was trained on data from 340 90 × 90 m tiles, or ~2.75 km^2, then extrapolated across ~134 km^2, resulting in a training to application ratio of ~0.020 or 2.0% (Zhang et al., 2018). In general, within the area across which the CNN-watershed approach was applied, it was less likely than Mask R-CNN to fail to detect polygonal terrain, but more prone to mistakenly aggregate multiple ice wedge polygons into a single unit. These errors were particularly common at sites characterized by transitional terrain where ice wedge polygons grade into non-polygonal ground. It is reasonable to expect such mistakes in these areas, as microtopography is typically faint and polygons often appear to be bound incompletely by troughs. At one such site (Prudhoe-6), the number of incorrect conglomerate polygons by area delineated by the CNN-watershed algorithm was ~22% (Table 1). This number closely resembles the 21% of human-delineated polygons estimated to go undetected by Mask R-CNN in satellite-based optical imagery (Zhang et al., 2018).

Table S1. Architecture of our CNN.

Layer	Type	Neurons
1	Convolutional	8 arrays of 27×27
2	ReLU†	8 arrays of 27×27
3	Max-pooling	8 arrays of 9×9
4	ReLU	8 arrays of 9×9
5	Fully-connected	64
6	ReLU	64
7	Fully-connected	2
8	ReLU	2
9	Softmax	2

† - ReLu – rectified linear unit

Table S2. Results of manual validation at 100 cm and 25 cm resolution (sites are 1 km^2).

			% of polygons by instance				*% of polygons by area*			
Site	Polygons identified	Polygonal area (%)	Whole	Fractional	Conglomerate	Non-polygonal	Whole	Fractional	Conglomerate	Non-polygonal
Barrow-1 (100 cm)	3058	74.3	73.4	18.6	6.2	1.8	65.5	16.6	13.4	4.1
Prudhoe-1 (100 cm)	3019	100	85.6	11.8	2.6	0.0	88.0	79.7	4.0	0.0
Barrow-1 (25 cm)	2870	71.6	89.0	3.8	3.4	3.8	83.4	2.0	9.7	4.8
Prudhoe-1 (25 cm)	3193	100	93.6	3.4	2.4	0.6	94.0	1.6	4.3	0.1