# Volumetric Quantification of Active Retrogressive Thaw Slumps in the Arctic

Potential Author lists:

Chunli Dai[1], Melissa K. Ward Jones[2], Jurjen van der Sluijs[3], Ian M. Howat[4], Anna K. Liljedahl[5], Nina Nesterova[6], Bretwood Higman[7], Jeffrey T. Freymueller[8], Steven V. Kokelj[3], Sindhura Sriram[1]

[1] School of Forest, Fisheries, and Geomatics Sciences (FFGS), University of Florida, Gainesville, FL, USA.
[2] Water and Environmental Research Center, University of Alaska Fairbanks, Fairbanks, AK, USA.
[3] Northwest Territories Centre for Geomatics, Government of Northwest Territories, Yellowknife, NT, X1A 2L9, Canada.
[4] Byrd Polar and Climate Research Center, The Ohio State University, Columbus, OH, USA.
[5] Woodwell Climate Research Center, Falmouth, MA, USA.
[6] Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Potsdam, Germany.
[7] Ground Truth Trekking, Seldovia, AK, USA.
[8] Department of Earth and Environmental Sciences, Michigan State University, East Lansing, MI, USA.

Correspondence to: Chunli Dai, chunlidai@ufl.edu

## Abstract (150 words)

Retrogressive thaw slumps (RTS) are crescent-shaped landslides that occur in ice-rich permafrost areas due to the melting of ground ice. Owing to the increased availability of high-resolution (2 m) time-dependent digital elevation models (DEMs) from ArcticDEM, the first volumetric quantification of RTS is carried out for the entire Arctic region. Based on DEM time series analysis and machine learning, we mapped a total of 2850 slumps across the Arctic, with a total volume loss of $(327.6 \pm 0.3) \times 10^6$ m$^3$ during the past decade. The median volume loss is about $(32.08 \pm 0.05)$ million m$^3$/year, yielding an equivalent carbon release of $(3.85 \pm 0.006) \times 10^{-4}$ Pg carbon per year between 2012 and 2022. Across the Arctic, 858 RTS showed positive correlations with the mean summer air temperature. The total volume loss caused by thaw slumping activity increases by $(1.9 \pm 1.6)$ million m$^3$/year for every degree of summer temperature rise.

**Highlights**:
1\ First volumetric quantification of RTS at the pan-Arctic scale.
2\ Estimated volume change per year shows the linkage to climate change.

**Keywords**: Retrogressive thaw slumps; ArcticDEM; Mask R-CNN; Arctic; DEM time series;

**Main Text (3000 words)**

**Introduction**

The Arctic region has experienced a warming trend significantly higher than the rest of the planet (e.g., Serreze and Barry, 2011; Meredith et al., 2019). This phenomenon is commonly referred as 'Arctic Amplification'. The rapid warming is raising concerns about the potential environmental and community impacts in the permafrost regions. Thawing ice-rich permafrost can lead to the formation of thermokarst landscapes, where mass-wasting processes cause substantial land subsidence and lateral soil transport through fluvial or colluvial mechanisms (Olefeldt et al., 2016). Two common types of thermokarst landslides include retrogressive thaw slumps (RTS) and active layer detachment slides (French, 2007). Among these, retrogressive thaw slumps are the most active and prevalent in the Arctic (e.g., Lin and Knudby, 2023). RTS are characterized by their distinctive crescent (or horseshoe) shape, steep ice-rich headwall, downslope slump floor, and deposition of debris tongues downstream from the slump floor (e.g., Witharana et al., 2022). RTS can be categorized as active or stable, where active RTS have newly exposed sediments and poor vegetation cover, and stable RTS have unchanging (sometimes vegetated) slump floors.

The early classification of RTS was mostly conducted with manual or semi-automated digitization using various datasets ranging from field data, aerial photographs to satellite data. From field surveys, Van der Sluijs et al. (2018) analyzed the dynamics of RTS using DEM time series generated by Unmanned Aerial Vehicle (UAVs) photogrammetry and Kokelj et al. (2021) quantified 3D RTS processes using LiDAR DEMs obtained by UAVs. Ward Jones et al. (2019) analyzed the annual dynamics of 12 RTS using a combination of Worldview satellite imagery and field mapping from GPS. Lantz and Kokelj (2008) manually digitized thaw slumps using aerial photographs in the Mackenzie Delta region, Canada. A collection of aerial photographs and satellite imagery were digitized to investigate the influence of climate and landscape factors on thaw slump dynamics in the Western Canadian Arctic (e.g., Segal et al., 2016a; Rudy et al., 2017). Regional thaw slumps were also manually mapped in northwestern Canada from SPOT satellite imagery (Kokelj et al., 2017) and in Banks Island, Canada from Google Earth Engine Timelapse dataset (Lewkowicz and Way, 2019).

With the increased availability of remote sensing data, machine learning and deep learning techniques have been explored for automated RTS classification. Popular machine learning (ML) techniques adopted for landslide and RTS detection include support vector machines (e.g., Ballabio and Sterlacchini, 2012; Mora et al., 2018; Bernhard et al., 2020), random forests (e.g., Nitze et al., 2018; Stumpf and Kerle, 2011; Chang et al., 2019), decision trees (e.g., Tsangaratos and Ilia, 2016). These methods are based on the processing of a set of manually selected features extracted from data. Features for landslide mapping include geomorphometric factors derived from digital elevation models (DEMs) such as terrain slope, aspect, roughness, the direction cosine of eigenvalue ratios, topographical curvatures, as well as other factors such as annual rainfall, lithology, distance from roads, distance from drainage networks, and so on (e.g., Mora et al., 2018; Reichenbach et al., 2018). Nevertheless, the accuracies of these methods depend on the choice of relevant causative factors (e.g., Chang et al., 2019).

Techniques to detect objects in images and generate segmentation masks have rapidly developed over a short period of time in the computer vision community (e.g., Krizhevsky et al., 2012; Girshick et al., 2014; Ren et al., 2015; Li et al., 2017; He et al., 2017). Several recent studies have adopted these deep-learning techniques to map retrogressive thaw slumps. For example, Huang et al. (2020) adopted a convolutional neural network (CNN) algorithm and mapped 220 retrogressive thaw slumps in the Tibetan Plateau from CubeSate images. Huang et al. (2022) further evaluated the transferability of their deep learning model and mapped around 620 RTS in three areas in Canadian Arctic using PlanetScope images aided with ArcticDEM data. Lin and Knudby (2023) mapped 365 RTS in western Canadian arctic regions using PlanetScope images and evaluated the performance of the neural network model Mask R-CNN. Based on Maxar imagery, Yang et al. (2023) tested the performance of neural network models which were trained with 965 RTS across several landscapes in Siberia and Canada. Huang, et al. (2023) retrieved 2494 RTS for the Arctic using ArcticDEM data, deep learning, and a web-based crowdsourcing system.

These previous studies demonstrated continuous efforts in digitizing the shape, size, and distribution of retrogressive thaw slumps (e.g., Segal et al., 2016a; Huang et al., 2020; Kokelj et al., 2021; Nitze et al., 2021; Runge et al., 2022). Nevertheless, there are noteworthy challenges in using remote sensing data for RTS detection, such as leaving slumps undetected (false negatives) or the false inclusion of non-slump landscape features (false positives) (e.g., Nitze et al., 2021; Runge et al., 2022), the limited size of the training data set (e.g., Yang et al., 2023).

While applications at the local or regional scale are many, continental-scale efforts in documenting topographic change and RTS are still needed. The patchy distribution of study areas may bias the scientific consensuses and affect the accuracy of evaluating climate change impact (e.g., Metcalfe et al., 2018). Having a baseline pan-Arctic product would standardize observations and highlight areas to focus future research efforts. A pan-Arctic product would also enable the assessment of large-scale linkages to variables such as climate and surficial geology as recent studies are locally or regionally focused.

The availability and open access of the large collection of time-dependent DEMs (ArcticDEM) offer new opportunities for detecting and quantifying RTS. The ArcticDEM data (Porter et al., 2022) is a large collection of 2-meter resolution Digital Elevation Models (DEMs) of the Arctic, derived from high-resolution (~0.5 m) stereo-paired images by Maxar satellites including Worldview- 1 (launched 2007), 2 (2009), 3 (2014), and GeoEye-1 (2008). DEMs are created using stereophotogrammetry applied to overlapping pairs of images (*Noh and Howat,* 2017), including pairs from the same or different orbital passes (i.e. in-track or cross-track pairs) and combinations of different sensors (*Noh and Howat,* 2019). The latest release of ArcticDEM (ArcticDEM strips version 4.1, Release October 2022) includes 440,949 time-dependent DEM strips between 2007 and 2022 across the Arctic. ArcticDEM data undergo basic automated quality control and filtering for errors due to clouds, shadows, detector saturation (Dai and Howat, 2018), water surfaces, and other sources using algorithms developed by the ArcticDEM project team. To better preserve data, the DEM strips of the new version (4.1) provide the error masks as separate files instead of masking out the classified "errors" as in the previous version, the ArcticDEM version 3 strip data (Porter et al., 2018; Dai et al., 2020a). Here we use the large

volume of ArcticDEM dataset, to generate detailed information including volume estimates for RTS activity.

**Results**

Using the time series analysis of ArcticDEM data, the first volumetric quantification of retrogressive thaw slumps at the pan-Arctic scale was carried out (Fig. 1). Based on change detection algorithms, we estimated the ground surface elevation changes, the time of the change, and the corresponding uncertainties. The time of change maps highlighted the 'evolving' features of the digitized and attributed RTS, where slumping gradually occurred from earlier years to recent years expanding the headwalls (Fig. 1c, Fig. S2e). This unique pattern is named as "*rainbow*" pattern, which distinguishes the RTS from elevation changes from other dynamics. Three bands of data including the elevation change map, total curvature, and the time of change map were processed as input for the machine learning algorithms (He et al., 2017) for automatic RTS classification. With further visual inspection, we identified a total of 2850 active retrogressive thaw slumps across the Arctic.

The detected retrogressive thaw slumps were spatially concentrated in a few regions such as Banks Island, the northwest coast of Victoria Island, Peel Plateau - Mackenzie Delta regions, Novaya Zemlya Rayon, Allaikhovsky District, and northern Siberia. The concentrated clusters of slumps in northwestern Canada were consistent with the wide swath of slumps detected along the Laurentide Ice Sheet margin by Kokelj et al. (2017). A large amount of these slumps were next to river channels or adjacent to lakes. Along the northern coasts of Siberia, Russia, there was a large distribution of lake-side slumps. A few slumps were found near the current glaciers and ice sheet margins in Greenland and Alaska. Next to glaciers in Novaya Zemlya Rayon, there were A total of 171 slumps. We also found a small cluster of slumps in Svalbard.

The amount of volume loss (ground ice and sediment loss) was estimated over the scar area, where materials were transported downslope by thaw slump activity, and ground surface elevation showed a decrease (e.g., Kokelj et al., 2015, 2021). The majority of largest RTS is distributed near Banks Island or Peel Plateau, Canada (Fig. 1). The median ground surface elevation change for all RTS is -3 meters. The median area and median volume detected in this study are 20,600 $m^2$ and 62,243 $m^3$, respectively. The total scar areas are estimated as 81.25 $km^2$ (8125 ha). The RTS with the largest volume loss occurred near Aklavik, Canada, with a volume estimate of 3,584,000 ± 54,000 $m^3$ and a scar area of 249,900 $m^2$ (25 ha). The peak surface elevation change for this largest RTS is about -33 ± 1 m.

The total volume loss for all 2850 slumps is estimated as (327.6 ±0.3) ×$10^6$ $m^3$. With a mass density of 917 kg/$m^3$ (ice density), the volume loss is equivalent to a mass loss of 0.3004 ± 0.0002 gigatonnes between 2012 and 2022. Assuming a carbon content of 12 kg C /$m^3$ in permafrost (e.g., Hugelius et al., 2011; Schuur et al., 2015), the equivalent total carbon release is (4 ± 0.003) ×$10^{-3}$ Pg carbon (where 1 petagram =1 gigatonnes).

To evaluate the impact of climate on RTS activity, the correlation between volume loss per year and the air temperature time series was calculated for individual RTS as well as for all RTS across the Arctic. For example, Fig. 1d shows the estimation of volume loss per year time series for one large RTS in Krasnoyarsk Krai, Russia. Compared to the ERA5 air temperature dataset

(Hersbach et al., 2023), the correlation between the volume loss per year and the summer (July and August) mean air temperature is high with a correlation coefficient of 0.73 and a good significance index ($p<0.05$). Across the Arctic, 858 RTS (30%) showed positive correlations ($>=0.5$), and 76 RTS (3%) showed strong correlations ($>=0.9$) (Fig. 2c). Our results are consistent with current literature, e.g., the mean annual surface temperature was found to be statistically correlated for half of the 19 RTS by Lacelle et al. (2015) and one-third of the 12 RTS by Ward Jones et al. (2019). Due to the heterogeneity of RTS dynamics, the correlation between the total volume loss per year and the mean air temperature (ERA5) is not strong, only about 0.29. Nevertheless, the total volume loss per year shows peaks in 2016 and 2020 and low values in 2013 and 2018, which are consistent with the temperature peaks and troughs in these years (Fig. 2b). The scatter plot of the volume loss per year with the temperature rise (Fig. 3) shows that the volume loss rate increases by ($1.9 \pm 1.6$) million $m^3$/year for every degree of temperature rise.

For the entire Arctic region, our estimated total volume loss per year is around ($32.08 \pm 0.05$) million $m^3$/year. The carbon release from active retrogressive thaw slumps is equivalent to ($3.85 \pm 0.006$) $\times 10^{-4}$ Pg carbon per year. Our estimate of carbon release from RTS is negligible compared to the modeled carbon release (~1 Pg carbon per year) from gradual thawing permafrost in the next century and the fossil-fuel emissions ($9.7 \pm 0.5$ Pg carbon per year in 2012) (Schuur et al., 2015).
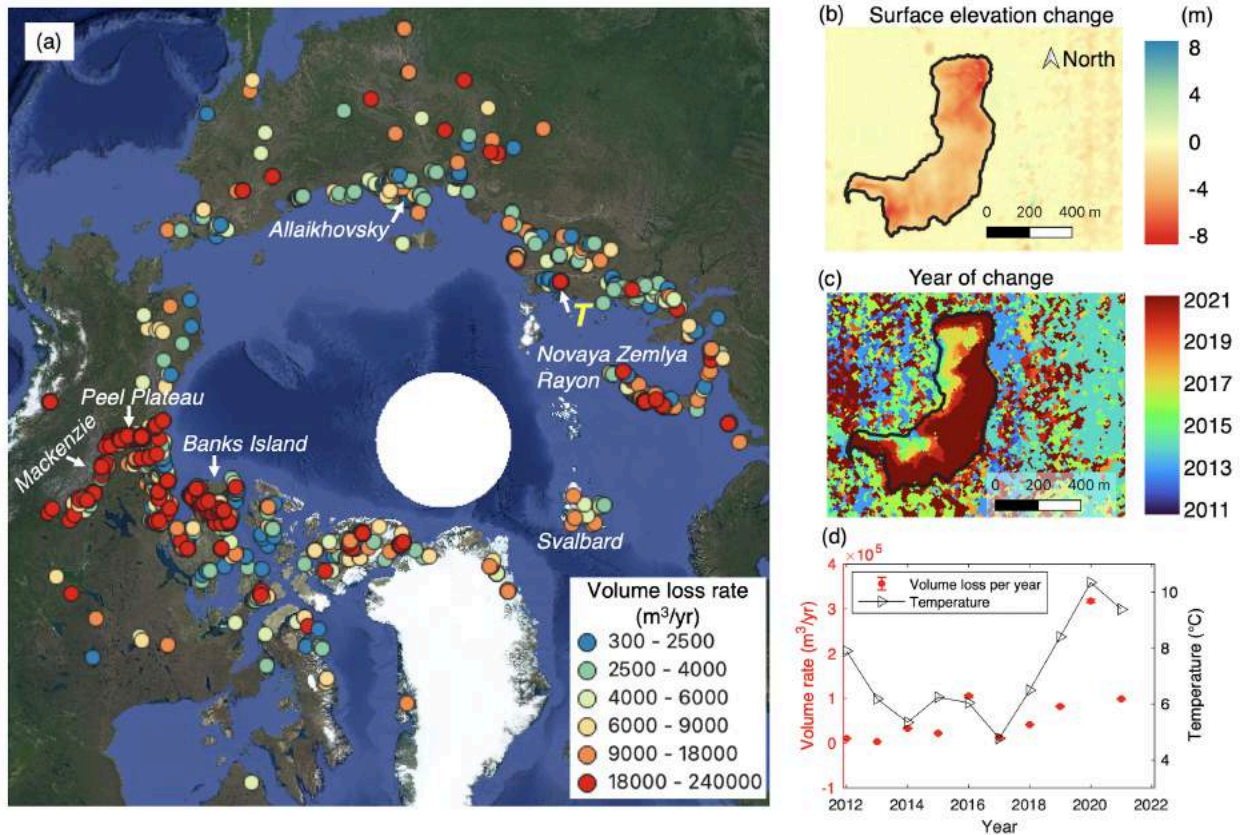


Fig. 1. A total of 2850 active retrogressive thaw slumps were detected in this study. (a) The median volume loss per year (unit $m^3$/year) for each RTS. (b) Surface elevation changes for one

large RTS (75°37'19.2"N 98°39'54.0"E) in Taymyrsky Dolgano-Nenetsky District, Krasnoyarsk Krai, Russia. It is marked as point *T* in (a). (c) The estimated time of change for this RTS. (d) The estimated volume loss per year of this RTS. Triangles are summer (July August) mean surface air temperatures at this site retrieved from the ECMWF Reanalysis version 5 (ERA5) dataset (Hersbach et al., 2023).
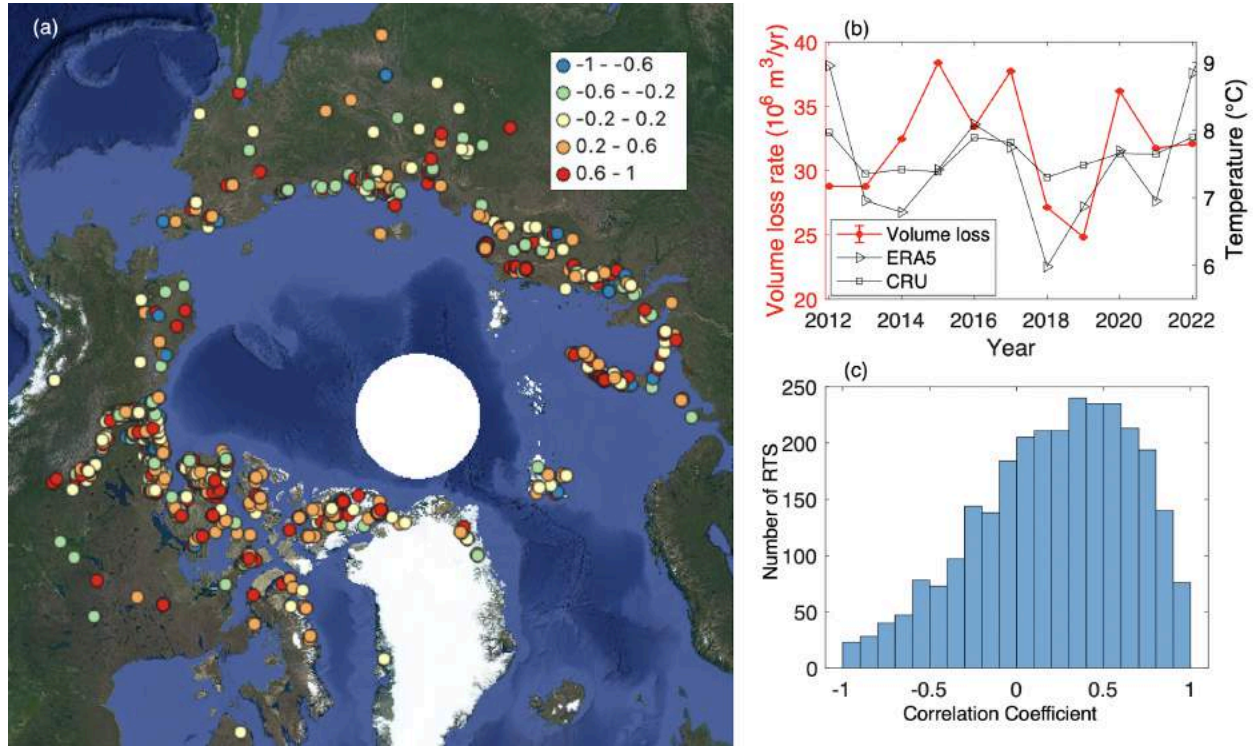


Fig. 2. Linkage between volume loss rate and air temperature. (a)The correlation coefficients between volume loss per year and air temperature for each RTS. Air surface temperatures are from ECMWF Reanalysis version 5 (ERA5) dataset (Hersbach et al., 2023). (b) Total volume loss per year of retrogressive thaw slumps across the Arctic during 2012 and 2022. Black symbols are the average air temperature among all RTS locations during July and August each year. Triangles are from the ERA5 dataset, and squares are from the CRU TS (Climatic Research Unit gridded Time Series, v. 4.07) surface air temperature dataset (Harris et al., 2020). Red dots are the total volume per year for selected 2812 RTS, where the rest (38 RTS) were not included since they only include less than two years of observations. (c) The histogram of correlation coefficients between volume loss rate and air temperature for all RTS in (a).
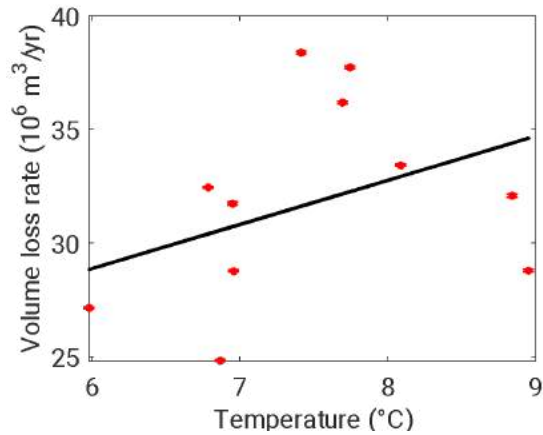
Fig. 3. The volume loss rate increases with temperature rises. The red dots are volume measurements, and the black line is the linear fit. The data are the same as Fig. 2b.

The controlling factors affecting RTS dynamics include temperature, precipitation, and local terrain factors (e.g. Ward Jones et al., 2019; Kokelj et al., 2015). In areas, where temperature dominates the drives for thaw slumping activity, the volume loss rate appears negative correlations with precipitation time series. For example, Ellesmere Island generally shows negative correlations with precipitation. For the selected RTS (Fig. 4b), the correlation with precipitation is -0.6 (p= 0.06), and the correlation with temperature is 0.86 (p =0.001). This negative correlation happens where the precipitation has a negative correlation with the air temperature, e.g., -0.8 (p=0.006).
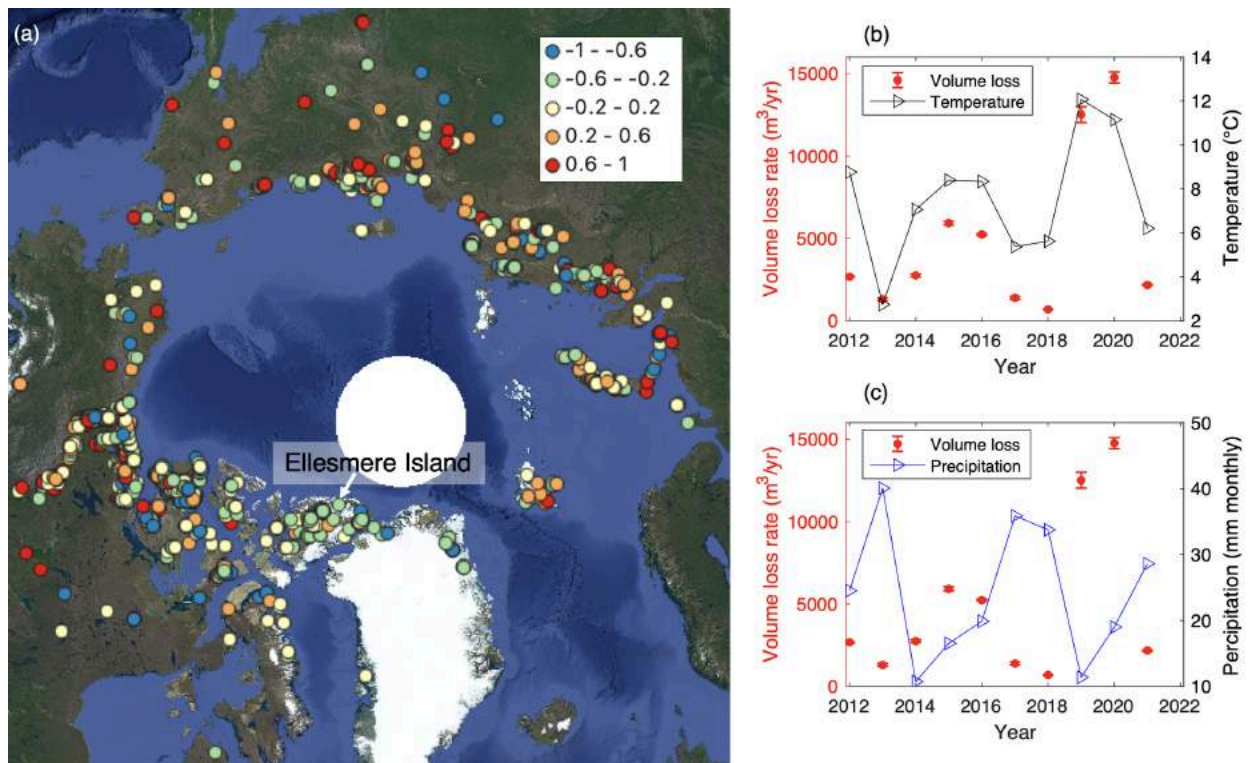
Fig. 4 The correlation of volume loss per year with mean summer (July-August) precipitation. The right panel shows the time series comparison between volume loss per year with temperature (b) and precipitation (c) at one selected RTS (79°47'52.4"N 84°38'50.6"W) in Ellesmere Island, Canada.


**Discussion/Conclusion**

Recent studies have shown a rise in retrogressive thaw slump activities linked to climate change on regional scales. However, knowledge about retrogressive thaw slump activities on a pan-artic scale remained limited. ArcticDEM, created based on stereophotogrammetry from high-resolution optical satellite imagery, provides a large collection of DEMs with vertical precisions as good as decimeters. In this paper, surface elevation changes and associated maps are successfully derived from the time series analysis of DEMs, which revealed the evolving features of RTS through the past decade. Machine learning algorithms are then used to enhance the automatic detection and classification of RTS activities. The training, validation, and test data sets are manually selected based on the surface elevation change maps, DEM curvature, and the time of change maps. We test and compare the performance of the machine learning algorithm (Mask R-CNN) under different scenarios. A pan-Arctic active RTS inventory was generated including RTS outlines, volume estimates, areas, median volume loss per year, as well as detailed surface elevation change maps and time of change maps.

This baseline pan-Arctic product will standardize observations using the new ArcticDEM data and enable us to highlight areas for future research efforts. A few new clusters of RTS were found in Svalbard, Greenland, Southern Alaska, and Novaya Zemlya Rayon. The volume loss per year for retrogressive thaw slump across the Arctic is estimated as $(32.08 \pm 0.05)$ million $m^3$/year, which is equivalent to a mass loss of $0.029422 \pm 0.00004$ Gt/year. The linkage of volume loss per year to summer air temperature is evident, where 30% of thaw slumps showed positive correlations (coefficients $>=0.5$) with temperature.

According to our measurements, the impact of active retrogressive thaw-slumping activity on the global carbon cycle is minimal, $(3.85 \pm 0.006) \times 10^{-4}$ Pg carbon per year. Yet, it provides the first quantification of the pan-Arctic scale carbon release from abrupt thaw processes, which is informative for the modeling of large-scale carbon-climate feedbacks (Schuur et al., 2015). It is noted that the DEM time series analysis presented here only detects active slumps in the past decade during the ArcticDEM time span (2012 to 2022) and cannot identify inactive or ancient slump areas. As van der Sluijs et al. (2022) showed about half of the slumps were inactive in northwestern Canada, for example, about 40% of slumps in Anderson Plain/Tuktoyaktuk Coastlands were active and 60% in Peel Plateau were active between 2004 and 2016.


**Online Methods (No words limit)**

**Step 1: Retrieving surface elevation change from DEM time series.**
*Co-registration of DEMs using Bundle Adjustment*

The DEMs have an internal (pixel-to-pixel) accuracy of approximately 30 cm (1 sigma). Nevertheless, biases in the horizontal and vertical directions give an external (absolute) accuracy of several meters, caused by the errors in sensor models (*Noh and Howat,* 2014; 2015). These biases can be removed through coregistration (Nuth and Kääb, 2011; Noh and Howat, 2014) and the bundle adjustment of translational offsets among repeat DEMs. Validation experiments over unchanging surfaces show that this coregistration reduces the uncertainty in elevation measurements to the random error in the DEMs of approximately 0.3 m~0.5 m (Noh and Howat, 2014; Shean et al., 2016; Dai et al., 2018). DEMs with consistently poor pairwise coregistration residuals are then identified as bad data and are removed.

We improve DEM coregistration through weighted least-squares adjustment of all translational offsets within pairs of DEMs (namely bundle adjustment). For all combinations of $N$ overlapping DEMs, we have $(N-1)N/2$ pairs of relative offsets. We then perform an adjustment to estimate the absolute offsets for each DEM relative to the common reference. For example, given four DEMs, labeled A, B, C and D, there will be 6 pairs of relative offsets, $\delta_{A,B}$, $\delta_{A,C}$, $\delta_{A,D}$, $\delta_{B,C}$, $\delta_{B,D}$, $\delta_{C,D}$, which are related to each DEM's absolute offset $\Delta$ as:

$$\delta_{A,B} = \Delta_A - \Delta_B$$

$$\delta_{A,C} = \Delta_A - \Delta_C$$

$$\delta_{A,D} = \Delta_A - \Delta_D$$

$$\delta_{B,C} = \Delta_B - \Delta_C$$

$$\delta_{B,D} = \Delta_B - \Delta_D$$

$$\delta_{C,D} = \Delta_C - \Delta_D$$

where $\Delta_A$, $\Delta_B$, $\Delta_C$, $\Delta_D$ is each DEM's absolute offset relative to the WGS84 datum.

The observation equation can be written in a matrix form as:

$$y = A\xi + e \qquad (1)$$

where, $y$ is the n by 1 vector of relative displacements, e.g.,
$y = [\delta_{A,B}\, \delta_{A,C}\, \delta_{A,D}\ \ \delta_{B,C}\, \delta_{B,D}\, \delta_{C,D}]^T$, A is the design matrix,
$A = [1 - 1\, 0\, 1\, 0 - 1\, 1\, 0\, 0\ 0\, 0 - 1\ 0\ 1 - 1\, 0\ 1\, 0\, 0\ 0\, 1\, 0 - 1 - 1]$,
$\xi$ is the *m* by 1 vector of parameters to be estimated, e.g., $\xi = [\Delta_A\, \Delta_B\, \Delta_C\ \ \Delta_D]^T$, and $e$ is the *n* by 1 vector of errors. The dispersion of $e$ is $\Sigma = \sigma_0^2 P^{-1}$, where $\sigma_0^2$ is the reference variance, $P$ is the weight matrix. The dispersion matrix can be constructed from the residuals estimated from the coregistration of each pair of images.

Since the least-squares solution of the above equation (1) is ill-conditioned, we add a stochastic constraint, i.e. the absolute offsets $\Delta$ have a zero mean and a standard deviation of 4 m, based on the statistics from the coregistration of 249 DEMs (*Dai et al.,* 2019). Its matrix form is:

$$k_0 = K\xi + e_0 \qquad (2)$$

where $K$ is an $m$ by $m$ unit matrix, $k_0$ is an $m$ by 1 vector of the given constraint, e.g., $k_0 = [0\ 0\ 0\quad 0\ ]^T$. $e_0$ is the $m$ by 1 vector of errors of the constraint. The dispersion of $e_0$ is $\Sigma_0 = \sigma_0^2 P_0^{-1}$, where $P_0$ is assumed as a unit matrix.

The least-squares solution of the above linear system with the stochastic constraint is:

$$\hat{\xi} = \left(N + K^T P_0 K\right)^{-1}\left(c + K^T P_0 k_0\right)$$

where, $N = A^T PA$, and $c = A^T Py$.

The estimated reference variance can be calculated as,

$$\hat{\sigma}_0^2 = \frac{\tilde{e}^T P\tilde{e} + \tilde{e}_0^T P_0 \tilde{e}_0}{n - m + l} \tag{3}$$

where, $\tilde{e}, \tilde{e}_0$ is the estimated error, $\tilde{e} = y - A\hat{\xi}$, $\tilde{e}_0 = k_0 - K\hat{\xi}$, and $l$ is the rank of the constraint equation (2), which equals $m$ here.

Hence, the dispersion matrix of the estimated parameters $\hat{\xi}$ can be calculated as

$$D\left(\hat{\xi}\right) = \hat{\sigma}_0^2\left(N + K^T P_0 K\right)^{-1} \tag{4}$$

The above equations can produce the adjusted absolute offsets for each DEM.
Note that to mitigate long-wavelength errors (e.g., tilt) (e.g., Shean et al., 2016) in DEMs, we divide the DEMs into 2 km by 2 km tiles and carry out the coregistration and bundle adjustment within the small tiles.

*Automated Change Detection Algorithm*
After coregistering DEMs, we develop automatic algorithms to detect the magnitude and time of the change in the ground surface elevation time series. First, we detect the time of the change by differencing the elevation between sequent observation times. A sudden change in the time series will appear as an outlier/spike in the differenced time series. The advantage of this method is that the differenced time series should be around zeros (within the noise and seasonal variability) if no abrupt change occurs. The elevation difference is:

$$y^j - y^{j-1} = e^{j-1}, \quad j = 1, 2, \ldots, n-1, \tag{5}$$

where $y^j$ is the elevation in meters at time $t^j$, $e^j$ is the measurement noise. The largest value for $y^j - y^{j-1}$ will be considered as a potential event of elevation change if it further conforms to a significance criterion, e.g., larger than 3 times the root mean square errors, and the time range of the event would be between $t^{j-1}$ and $t^j$. So the estimated mass wasting time is the median of two sequential measurements that have detected a change, $(t^{j-1} + t^j)/2$. The temporal resolution of the event time estimate is mainly limited by the temporal resolution of satellite imagery acquisitions. After estimating the event time, the elevation change and its uncertainty can be calculated using the linear model based on the Heaviside step function, as demonstrated in Dai and Howat (2017). We then construct maps of event time of change, elevation change, and uncertainty by applying the above algorithm for each pixel on a regular 2 m grid within a 2 km by 2 km tile. Examples of these maps shown in Fig. S1(a–b), display both real surface changes caused by mass wasting and deposits, as well as noise. The elevation change profile (Fig. S1(e)) shows that the magnitude of elevation change increases with the terrain elevation. The detected time of change (Fig. S1(b, f))

demonstrates how this RTS developed retrogressively over the years, where the time of change near the headwall is most recent (2015), and the west of the area had slumping activities in 2013, 2011, and 2010 (earliest where it is furthest away from the headwall). Another characteristic of the time of change map is that the estimated time is locally homogeneous over the depletion zone, but random in other areas.
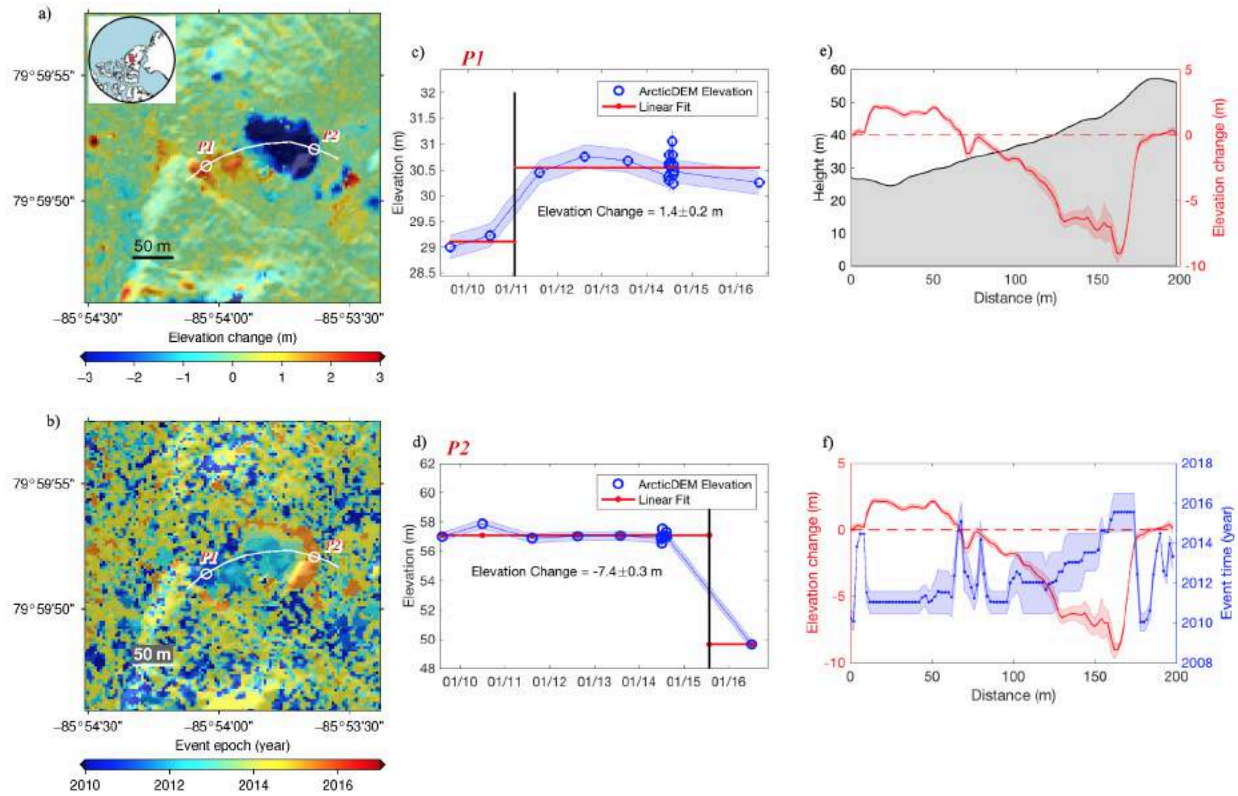


Fig. S1. Dynamics of a retrogressive thaw slump in Eureka, Canada. (a) Surface elevation change estimated from ArcticDEM time series. The background is the hillshade of the ArcticDEM mosaic. The inset shows the location of the site. (b) Estimated elevation change time in years. (c) and (d) shows the elevation time series at points P1 and P2 (79°59'52.1"N, 85°53'38.0"W) in (a), where the x-axis is the date (month/year) and the y-axis is the surface elevation. The blue circles are the terrain elevations measurements, and the black vertical line shows the automatically detected time of elevation change. The red line represents the linear model fit of the measurements. The red shading is the uncertainty of the fitted value through error propagation from the posteriori variance of unit weight. (e) Topography and elevation change across the manually drawn profile (the white line) in (a). The black line and gray shaded area (left y-axis) represent the earliest topography acquired on August 10, 2009, and the red line is the elevation change profile (right y-axis). Red shading is the uncertainty of the estimated elevation change, which varies due to different data availability at different locations. (f) the estimated elevation change time in years (right $y$ axis). The red line and its shading are the same as in (e). DEM(s) created by the Polar Geospatial Center from DigitalGlobe, Inc. imagery.
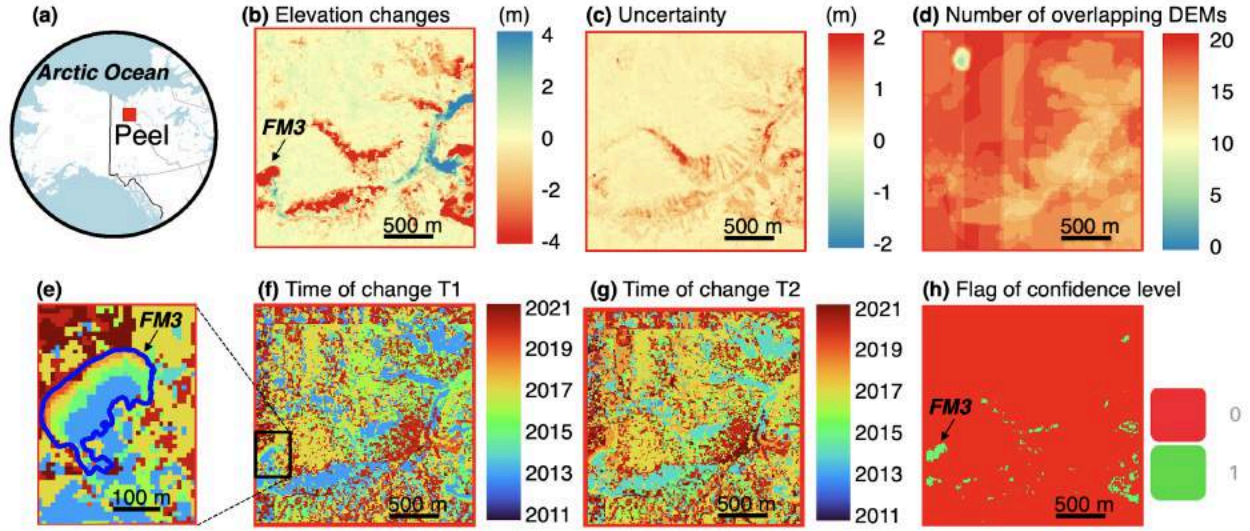
Fig. S2. Sample products in a 2 km by 2 km area in Peel Plateau, Canada. (a) The red square denotes the location of the study area. (b) The surface elevation change map near the slump FM3 (Van der Sluijs et al., 2018). The coordinate of FM3 is 135.274°W, 67.255°N. (c) The uncertainty of the surface elevation change. (d) The number of overlapping DEM strips. (e) A zoom-in view of slump FM3 for the estimated time of change. (f) Estimated time of change, T1. (g) Estimated time of change, T2. (h) Map of flags representing the confidence level of the estimated elevation changes. 1 means high confidence, and 0 means low confidence.

Fig. S2 demonstrates the sample products from our change detection processing, which include 6 data types for each area. The 6 data types are the elevation change map, the uncertainty of the elevation change, the time of the change T1 (date of the closest DEM data before the detected change), the time of the change T2 (date of the closest DEM data after the change), the number of overlapping DEMs in the area, the flag of confidence levels for the estimated elevation change at each pixel.

The confidence level is classified using the following conditions:
1) Estimated time is homogeneous locally;
2) Elevation change uncertainty < 2m;
3) Elevation change < -2 m;
4) The elevation change estimate should be larger than 3 fold the uncertainty, and the total number of overlapping DEMs is at least 4;
5) Total area of the selected cluster> 200 m$^2$.


**Step 2: RTS candidate selection criteria**

From the elevation change maps generated by the change detection algorithm, we then classify potential RTS with a simple thresholding method. The following criteria are used for classifying potential RTS:
    1) Surface elevation changes <= -1 meter, and total area > 2000 m$^2$.
    2) Within the polygon, there is at least one pixel with a high confidence level.
    3) The standard deviation of the time of change (T1) within the selected cluster > 1 year.

Based on only criterion 1, the total number of RTS candidates is 936370 for Region 9 (Victoria, Canada) and 9498 for a 50 km by 50 km area in Peel Plateau, Northwest Territories. Adding criteria 2 and 3, the total number of RTS candidates is reduced by 8 ~ 50 times.

Table S1. The number of RTS candidates:

|  | Region 9 (369 tiles) | Peel (one 50 km by 50 km tile) |
|---|---|---|
| Criterion 1 | 936370 | 9498 |
| Criteria 1-3 | 17184 (50 times less) | 1238 (8 times less) |

Considering the typical sizes of RTS mapped in the literature (e.g., Kokelj et al., 2015; Nitze et al., 2021; van der Sluijs et al., 2022; Lin and Knudby, 2023), RTS were grouped into six classes based on the total area over the scar zone. The six classes are:
1) Mega RTS, scar area $>= 500,000$ m$^2$,
2) Very large RTS, $100,000$ m$^2 <=$ scar area $< 500,000$ m$^2$,
3) Large RTS, $50,000$ m$^2 <=$ scar area $< 100,000$ m$^2$,
4) Median RTS, $10,000$ m$^2 <=$ scar area $< 50,000$ m$^2$,
5) Small RTS, $5,000$ m$^2 <=$ scar area $< 10,000$ m$^2$,
6) Very small RTS, $2,000$ m$^2 <=$ scar area $< 5,000$ m$^2$.
In this study, smaller RTS (with scar areas $<10,000$ m$^2$) are not considered.

**Step 3a: Mask R-CNN for RTS classification with training set 1.**

To detect slumps in images and generate segmentation masks, we adopt the state-of-the-art algorithm Mask R-CNN (Region-Based Convolutional Neural Network) (He et al., 2017), which has been widely recognized as a promising semantic segmentation algorithm across broad scientific applications (e.g., Bhuiyan et al., 2020; Witharana et al., 2020; Ackermann et al., 2022; Martins et al., 2022; Lin and Knudby, 2023). Mask R-CNN extends the previous Faster R-CNN architecture (Ren et al., 2015) to include pixel-level segmentation on top of bounding box recognition. We use open-source detection and segmentation software Detectron2 (Wu et al., 2019), which is the latest implementation of the Mask R-CNN algorithm (He et al., 2017). Deep learning models in this study are trained with Nvidia A100-SXM Tensor Core GPUs on Google Colab.

**Training data preparation**

We collected three bands of data as the input for machine learning. As shown in Fig. S3(a), based on criteria 1-3 only, it is hard to distinguish RTS from other surface changes, such as ground surface changes around gullies. Hence, extra two bands of information are added to input images. Band 2 is the DEM total curvature (Wilson and Gallant, 2000), which shows distinct patterns near gullies and RTS, e.g., the curvature is high and outlines the headwalls of RTS (Fig. S3g), whereas the curvature shows skeleton features along gullies (Fig. S3c). Band 3 is the time of change map, which shows unique "rainbow" patterns for RTS (Fig. S3h, Fig. S2e), whereas it shows random patterns near gullies (Fig. S3d). The 3-band images (Figs. S3e and S3i) are then rescaled to integers between 0 and 255, which are the input images for Mask R-CNN.

Each RTS candidate from Step 2 was prepared in a format compatible with the Detectron2 software. We applied a 100 m buffer around the outline, and then the 3-band images, RTS outlines, and bounding boxes were used as the input for Detectron2. The RTS polygons were labeled as one class 'slump'. The labeling information including coordinates of polygons outlining the boundary of RTS, bounding box, class types, and file size is exported as JavaScript Object Notation (JSON) files in COCO (Common Objects in Context) format to be read by the Detectron2 software. Training and validation data were randomly split with a ratio of 9:1. The training data contains a total of 900 images, and validation data contains a total of 104 images (Fig. S4).
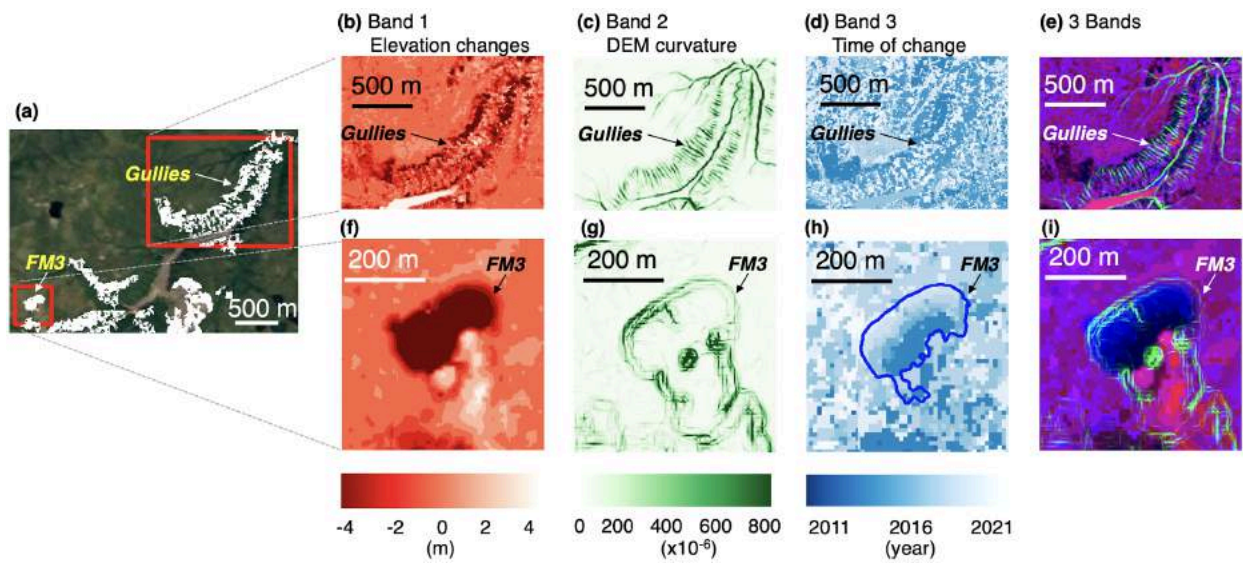


Fig. S3. Example of input images for machine learning. (a) Location of RTS candidates (135.274°W, 67.255°N) in Peel Plateau, Canada. The white pixels are highlighted RTS candidates from criteria 1-3. The domain of (a) is also marked as the white box in Fig. S4b. (b-e) Band 1, 2, and 3 of an image that includes the gullies feature. (f-i) Band 1, 2, and 3 of an image that includes the slump FM3.

**Training Model**

Using the training data set, we fine-tuned a pre-trained Mask R-CNN model (mask_rcnn_X_101_32x8d_FPN_3x.yaml) available in Detectron2's model zoo. Referring to the configuration of training parameters in the literature (Martins et al., 2022; Ackermann et al., 2022), the learning rate (cfg.SOLVER.BASE_LR) is set as 0.00025, and the maximum iteration (cfg.SOLVER.MAX_ITER) is set as 48000, and the threshold for prediction confidence score (cfg.MODEL.ROI_HEADS.SCORE_THRESH_TEST) is set as 0.7.

**Prediction and Evaluation**

We adopt the inventory of retrogressive thaw slumps in the Peel Plateau, Northwest Territories from Segal et al. (2016b) as the ground truth for evaluation. The inventory provides outlines of RTS derived from Quickbird imagery (0.6 m resolution) and SPOT imagery (10 m resolution). We further only selected RTS that show signals from our ArcticDEM data, i.e., the ones that coincide with our RTS candidates from Step 2. The selected 83 RTS in a 50 km by 50 km area in Peel Plateau are used as ground truth (Fig. S4b).

After training, the model prediction is applied to test images for evaluation. The output of prediction for each image is the 2D mask of classified RTS. We remove all clusters with an area of less than 2000 m$^2$. To evaluate the performance of machine learning classification, we adopt the metric -intersection over union (IoU). IoU is defined as the ratio of the intersection of two areas to the union of two areas. The threshold for IoU is set as 0.1. A polygon is classified as a true positive if the IoU is larger than the threshold, and a false positive if otherwise.
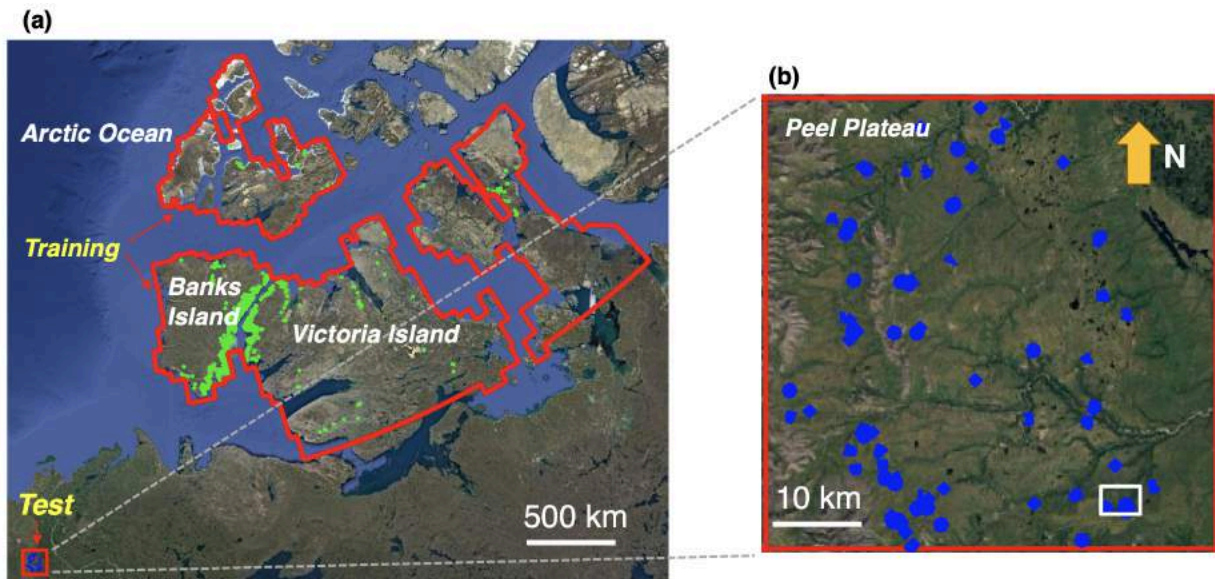


**Fig. S4.** Training and test data 1. The red polygons represent the region of training/validation data sets, including a total of 1004 RTS (green dots). The red square represents a 50 by 50 km test area in Peel Plateau, including 83 RTS (blue dots). The white box marks the location of Fig. S3(a).

As shown in Fig. S4b and Table S2, there are 83 truth positives in this 50 km by 50 km tile area in the Peel Plateau. The criteria 1-3 in Step 2 can successfully recover 77 true positives yet failed to filter out 1161 false positives. The machine learning results can reduce the number of false positives by 23%. We further combine the two solutions, by only selecting candidates that have more than 50% of areas classified as RTS by machine learning. This combination of machine learning and criteria 1-3 reduces the number of false positives by 3 times from 1161 to 357. This combination by requiring the agreement of two methods better than 50% removes a large number of false positives. As demonstrated in Figure S5, for the ground truth Slump FM3 (Van der Sluijs

et al., 2018), the cluster from criteria 1-3 and machine learning show a good agreement, yielding a true positive. For the uppermost cluster, the machine learning indicated pixels (red) are much less than that from criteria 1-3 (white), which indicates bad agreements, leading to a rejection of that cluster.
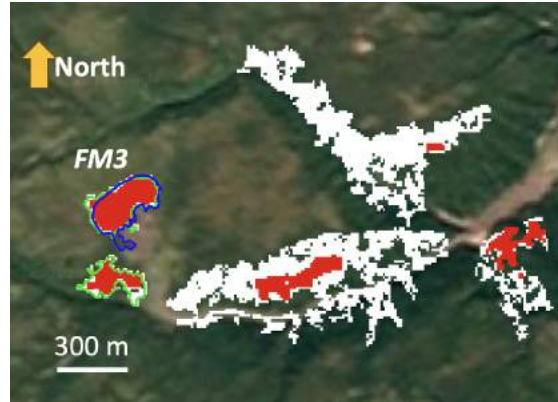


**Fig. S5.** Improvement from the combined solution in Peel Plateau, Canada. White pixels are clusters from criteria 1-3, red are clusters from the machine learning, green are outlines of slumps from the combined solution, and blue is the ground truth outline of Slump FM3. The background is the Google Earth image, shown in the UTM 8N projection.

Table S2. Machine learning performance at Peel Plateau (one 50 km by 50 km tile).

|  | True Positive | False Positive | False Negative | Precision | Recall |
|---|---|---|---|---|---|
| Criteria 1 | 83 | 9415 | 0 | 1% | 100% |
| Criteria 1-3 | 77 | 1161 | 5 | 7% | 94% |
| Mask R-CNN (Model 1) | 76 | 897 | 6 | 8% | 93% |
| Combined | 71 | 357 | 12 | 17% | 86% |

Note: the total number of ground truth 83.


**Step 3b: Mask R-CNN for RTS classification with training set 2.**

After the first iteration of applying the machine learning algorithm on several regions, we manually selected the RTS as a second training data. The manual selection of training data sets was conducted in QGIS, by visually scrutinizing each candidate shapefile on top of the time of change base map. The candidate shapefile that encloses an area with the "rainbow" pattern is classified as a true RTS. The total number of training images is 1660, and the number of validation images is 189 (Fig. S6).
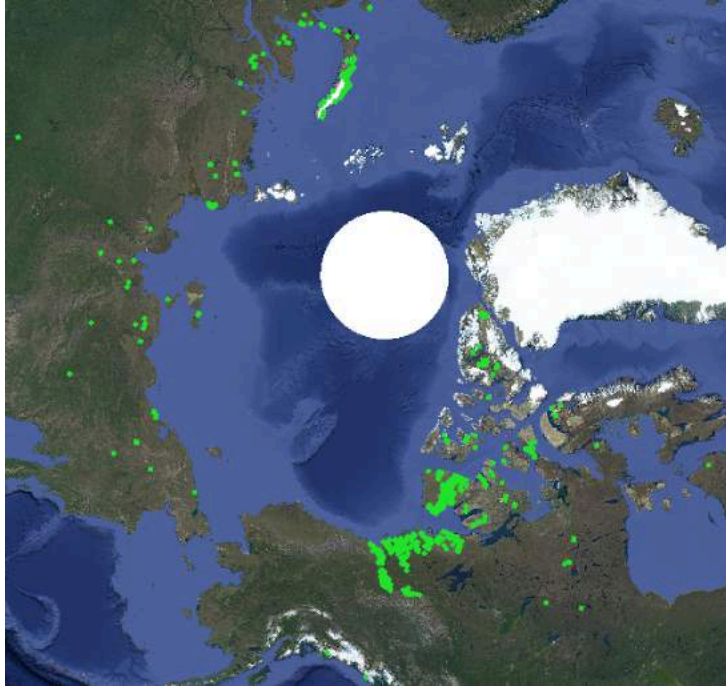
Fig. S6. Training set 2. Green dots represent the training data set 2, with 1660 training and 189 validation images.

**Data augmentation**

Considering the available training data size is relatively small, data augmentation techniques (e.g., Witharana et al., 2020; Ackermann et al., 2022) were applied to minimize overfitting. The adopted data augmentation includes rotation, translation, rescaling, and flipping, which increases the sample size by 25 times. The precision and recall of five different models are compared as shown in Table S3. The overall issue is that the results have too many false positives, which yield a low precision. Data augmentation in fact unfavorably increases the number of false positives, e.g., 524 vs 357 for model 1aug (with augmentation) and model 1 (without augmentation). Using the training data set 2, the number of true positives increases slightly, but the number of false positives also increases. Nevertheless, the combination of five models greatly reduces the number of false positives without significantly reducing the true positives. Yet, the precision is still low, only 24%.

Table S3. Machine learning improvement by merging several results.

| Machine learning models | Peel Plateau | | | | |
|---|---|---|---|---|---|
| | True Positive | False Positive | False Negative | Precision | Recall |
| ML (model 1) | 71 | 357 | 12 | 17% | 86% |
| ML (model 1b) | 71 | 518 | 12 | 12% | 86% |
| ML (model 1aug) | 71 | 524 | 12 | 12% | 86% |
| ML (model 2) | 73 | 673 | 10 | 10% | 88% |
| ML (model 2aug) | 74 | 849 | 9 | 8% | 89% |
| Merge above all | 68 | 210 | 15 | 24% | 82% |

Note: the total number of ground truth is 83 in Peel Plateau. The results here are all combined solutions with criteria 1-3. The configuration of training parameters for all rows is the same except for the model 1b, which has the maximum iteration as 2000 (cfg.SOLVER.MAX_ITER). Model name with 'aug' denotes a model with data augmentation. Models 1, 1b, and 1aug are trained using training data set 1, and models 2 and 2aug are established using training data set 2.

Table S4. The number of candidates from different steps.

| Region Name | Region ID | Step 2 (Criteria 1-3) | Step 3 | Step 4 (Manual selection from step 3) |
|---|---|---|---|---|
| Iceland | 1 | 18,720 | 2,450 | 0 |
| Southwest Greenland | 3 | 69,110 | 7,309 | 1 |
| Northeast Greenland | 5 | 32,005 | 6,339 | 15 |
| Northwest Greenland | 6 | 21,114 | 2,796 | 3 |
| Ellesmere, Canada | 7 | 48,366 | 8,128 | 191 |
| Baffin, Canada | 8 | 37,414 | 5,688 | 15 |
| Victoria, Canada | 9 | 6,674 | 2,253 | 1317 |
| North Mainland, Canada | 10 | 66,888 | 9,636 | 569 |
| North Hudson, Canada | 11 | 22,773 | 5,986 | 10 |
| Svalbard | 14 | 19,241 | 2,533 | 15 |
| Novaya Zemlya, Russia | 15 | 15,297 | 1,859 | 171 |
| Franz Josef, Russia | 29 | 4,615 | 365 | 0 |
| Siberian Islands, Russia | 30 | 9,067 | 1,633 | 103 |
| South Canada | 12 | 304,146 | 50,277 | 101 |
| Cherskly, Russia | 18 | 85,754 | 11,269 | 55 |
| Magadanskaya, Russia | 19 | 323,314 | 26,626 | 13 |
| Kamchatka, Russia | 20 | 195,611 | 17,763 | 3 |
| East Yakutiya, Russia | 21 | 214,518 | 22,321 | 117 |
| West Yakutiya, Russia | 23 | 194,989 | 29,008 | 49 |
| Norilsk, Russia | 25 | 83,792 | 16,843 | 69 |
| Murmansk, Russia | 27 | 95,370 | 19,547 | 18 |
| North Alaska | 34 | 145,953 | 19,154 | 15 |
| Southeast Greenland | 2 | 172,331 | 79,258 | N/A |
| Central East Russia | 22 | 796,816 | 95,662 | N/A |
| Central West Russia | 24 | 245,028 | 51,924 | N/A |
| Petersburg, Russia | 26 | 482,507 | 96,575 | N/A |
| Scandinavia | 28 | 559,753 | 114,719 | N/A |
| South Alaska | 31 | 368,949 | 21,287 | N/A |

Note: Yellow rows are the regions not checked, since they have too many candidates for visual inspection and have no evidence of thaw slump activities. Green rows are the regions that are checked within the vicinity (~10 km) of published RTS.
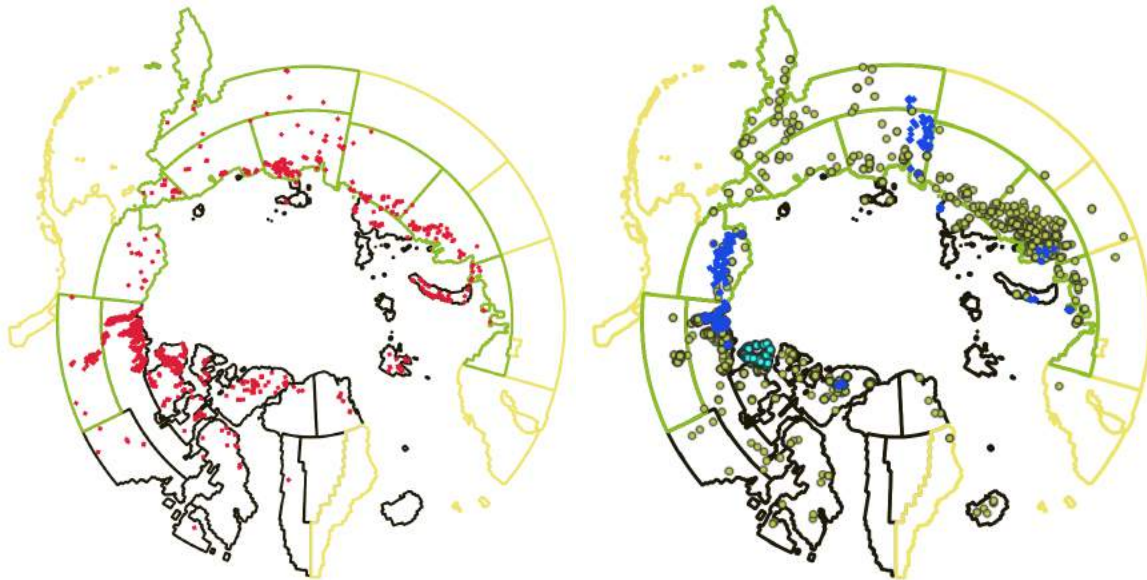


Fig. S7. Locations of RTS detected in this study (*left*) and in the literature (right). Black polygons show regions that have been visually inspected fully, green polygons are the regions that are only checked within the vicinity (~10 km) of known sites, and yellow polygons are the regions not checked. Red dots are RTS in this study, Green dots are RTS by Huang et al. (2023), cyan dots are RTS by Lewkowicz and Way (2019), and blue points are others (Nitze et al., 2018; Nitze et al., 2021; Ramage et al., 2017; Segal et al., 2016b, 2016c, 2017, van der Sluijs and Kokelj, 2023).

**Step 4**
After machine learning, we visually assess the polygons based on Google satellite images and the "rainbow" patterns in the time of change maps derived in this study. Randolph glacier inventory maps (RGI 7.0 Consortium, 2023) were also used to mask out glacier areas. We further use a priori information for references including ground ice content maps (Brown et al., 2002) and published RTS such as 4578 RTS from Lewkowicz and Way (2019), 1203 RTS from Nitze et al. (2021), 395 RTS from Nitze et al. (2018), 286 RTS along Yukon coast by Ramage et al. (2017), 206 RTS in Peel Plateau from Segal et al. (2016b), 2037 RTS in Anderson Plain-Tuktoyaktuk Coastlands regions (Segal et al., 2017, van der Sluijs and Kokelj, 2023), 832 RTS in Banks Island (Segal et al., 2016c), as well as 2494 RTS across the Arctic (Huang et al., 2023). 13 regions (black polygons in Fig. S7) were visually inspected thoroughly for the entire domains, while 9 regions (green polygons in Fig. S7) were inspected only around the vicinity of published RTS. The final selection of RTS is listed in Table S4, which includes a total of 2850 RTS.

**Volume estimation**

We divide the entire slump scar area into sub-areas by the time of change in years. For each year, we estimate the volume change per year using this equation:

$$volume\_per\_year = \frac{A\sum_{i}^{N}h_i}{\Delta t}$$

Where A is the area of a single pixel (e.g., 4 m²), $h_i$ is the height change at each pixel, and N is the total number of pixels within the scar area of a given year. $\Delta t$ is the time span in years. It's noted that satellite sampling interval may undesirably scale the volume loss rate. For example, if two acquisitions between April to September, the satellite interval will yield a half-year time difference, where it actually means the change occurred in one summer year. Hence we use the number of summer years as the time difference instead of using the precise satellite acquisition dates. Here, $\Delta t$ represents the count of summers (June 20 to September 22) within the measurements' time span. If an acquisition occurs before the summer of a year, the time is labeled as the previous year.

Following the method in Dai and Howat (2017), the uncertainty of volume per year is calculated as:

$$\sigma_v^2 = \frac{A^2\sum_{i}^{N}\sum_{j}^{N}C_{ij}}{\Delta t}$$

Where $\sigma_v$ is the standard deviation of the estimated volume per year. $C_{ij}$ is the covariance between height error at pixel $i$ and pixel $j$. The covariance can be empirically calculated using the periodogram method (Welch, 1967; Dai and Howat, 2017; Fig. S8).
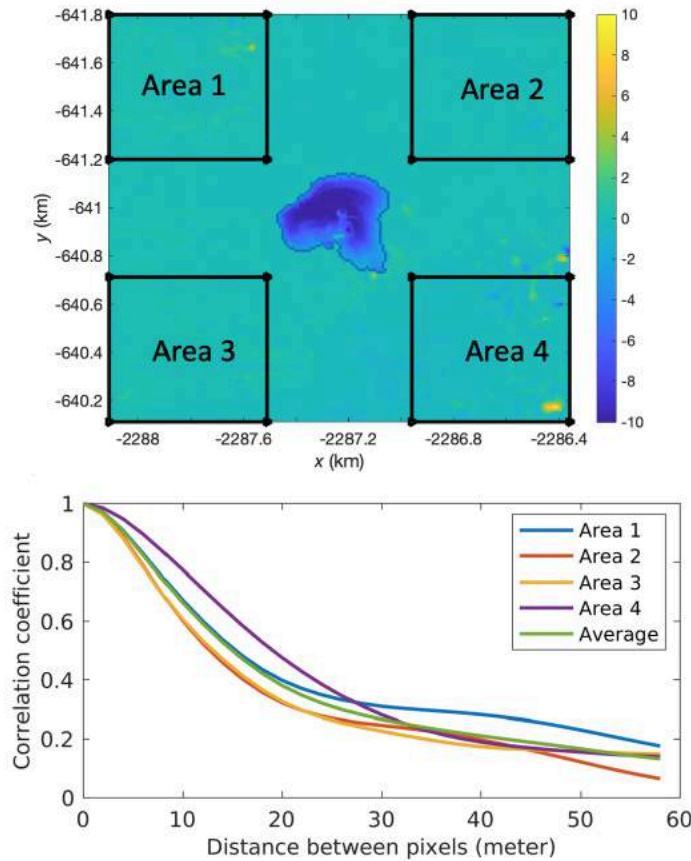
Fig. S8 The calculation of covariance between pixels for an RTS in Kitikmeot Region, Canada. (a) surface elevation changes in meters. Four boxes represent areas used for calculating covariances. $x$ and $y$ axes are polar stereographic coordinates. (b) the correlation coefficient between pixels using the periodogram method (Welch, 1967; Dai and Howat, 2017) among the selected four areas. The green line is the average of the correlation coefficients of four areas. The volume of this slump (68.32449°N, -119.3465°W) is estimated as 764,470 m$^3$ ± 12,525 m$^3$.

**Climate Data.**

To estimate Arctic air temperature, we used the surface air temperature (at 2-meter height) estimates from the hourly ECMWF Reanalysis version 5 (ERA5) data with 0.25° grid size (Hersbach et al., 2023). CRU TS (Climatic Research Unit gridded Time Series, v. 4.07) surface air temperature dataset (Harris et al., 2020) was also used for comparison. The precipitation data were from the hourly ERA5 dataset (total precipitation) (Hersbach et al., 2023). The total precipitation is the accumulated liquid and frozen water, comprising rain and snow, that falls to the Earth's surface. The mean values are calculated for all RTS sites during the selected summer months (July and August).

**Author contributions**

C.D. carried out the study and drafted the manuscript. J. S., M.W.J., and N. N. provide field data and ground truth information for regions respectively. S.S. and C.D. carried out the visual inspection of RTS candidates. XX contributed to the data analysis and interpretation of results. All authors revised the manuscript.

**Data availability**

ArcticDEM data are available via www.arcticdem.org (last accessed July 19, 2023).
, and AWS at. The surface elevation change maps and classified RTS shapefiles are available at.

**References**

Ackermann, M., Iren, D., Wesselmecking, S., Shetty, D. and Krupp, U., 2022. Automated segmentation of martensite-austenite islands in bainitic steel. *Materials Characterization*, *191*, p.112091.

Bernhard, P., Zwieback, S., Leinss, S. and Hajnsek, I., 2020. Mapping retrogressive thaw slumps using single-pass TanDEM-X observations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, pp.3263-3280.

Bhuiyan, M.A.E., Witharana, C. and Liljedahl, A.K., 2020. Use of very high spatial resolution commercial satellite imagery and deep learning to automatically map ice-wedge polygons across tundra vegetation types. *Journal of Imaging*, *6*(12), p.137.

Brown, J., Ferrians, O., Heginbottom, J.A. and Melnikov, E., 2002. Circum-Arctic map of permafrost and ground-ice conditions, version 2. *Boulder, Colorado USA, National Snow and Ice Data Center*.

Chang, K.T., Merghadi, A., Yunus, A.P., Pham, B.T. and Dou, J., 2019. Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques. *Scientific reports*, *9*(1), p.12296.

Dai, C. and Howat, I.M., 2017. Measuring lava flows with ArcticDEM: Application to the 2012–2013 eruption of Tolbachik, Kamchatka. *Geophysical Research Letters*, *44*(24), pp.12-133.

Dai, C., Durand, M., Howat, I.M., Altenau, E.H. and Pavelsky, T.M., 2018. Estimating river surface elevation from ArcticDEM. *Geophysical Research Letters*, *45*(7), pp.3107-3114.

Dai, C., Howat, I.M., Larour, E. and Husby, E., 2019. Coastline extraction from repeat high resolution satellite imagery. *Remote sensing of environment*, *229*, pp.260-270.

Dai, C., Howat, I.M., Freymueller, J.T., Vijay, S. and Jia, Y., 2020(a). Characterization of the 2008 phreatomagmatic eruption of okmok from ArcticDEM and InSAR: deposition, erosion, and deformation. Journal of Geophysical Research: Solid Earth, 125(6), p.e2019JB018977.

Dai, C., Higman, B., Lynett, P.J., Jacquemart, M., Howat, I.M., Liljedahl, A.K., Dufresne, A., Freymueller, J.T., Geertsema, M., Ward Jones, M. and Haeussler, P.J., 2020(b). Detection and assessment of a large and potentially tsunamigenic periglacial landslide in Barry Arm, Alaska. *Geophysical research letters*, *47*(22), p.e2020GL089800.

Girshick, R., Donahue, J., Darrell, T. and Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).

Harris, I., Osborn, T.J., Jones, P. and Lister, D., 2020. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific data*, *7*(1), p.109.

CRU TS Version 4.07, https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.07/ge/ (2024), Access on Jan 20, 2024.

He, K., Gkioxari, G., Dollár, P. and Girshick, R., 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J-N. (2023): ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), DOI: 10.24381/cds.adbb2d47 (Accessed on Jan 29, 2024).

Huang, L., Lantz, T.C., Fraser, R.H., Tiampo, K.F., Willis, M.J. and Schaefer, K., 2022. Accuracy, efficiency, and transferability of a deep learning model for mapping retrogressive thaw slumps across the Canadian Arctic. *Remote Sensing*, *14*(12), p.2747.

Huang, L., Willis, M.J., Li, G., Lantz, T.C., Schaefer, K., Wig, E., Cao, G. and Tiampo, K.F., 2023. Identifying active retrogressive thaw slumps from ArcticDEM. *ISPRS Journal of Photogrammetry and Remote Sensing*, *205*, pp.301-316.

Kokelj, S.V., Tunnicliffe, J., Lacelle, D., Lantz, T.C., Chin, K.S. and Fraser, R., 2015. Increased precipitation drives mega slump development and destabilization of ice-rich permafrost terrain, northwestern Canada. *Global and Planetary Change*, *129*, pp.56-68.

Kokelj, S.V., Lantz, T.C., Tunnicliffe, J., Segal, R. and Lacelle, D., 2017. Climate-driven thaw of permafrost preserved glacial landscapes, northwestern Canada. *Geology*, *45*(4), pp.371-374.

Kokelj, S.V., Kokoszka, J., van Der Sluijs, J., Rudy, A.C., Tunnicliffe, J., Shakil, S., Tank, S.E. and Zolkos, S., 2021. Thaw-driven mass wasting couples slopes with downstream systems, and effects propagate through Arctic drainage networks. *The Cryosphere*, *15*(7), pp.3059-3081.

Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

Hugelius, G., Virtanen, T., Kaverin, D., Pastukhov, A., Rivkin, F., Marchenko, S., Romanovsky, V. and Kuhry, P., 2011. High‑resolution mapping of ecosystem carbon storage and potential effects of permafrost thaw in periglacial terrain, European Russian Arctic. *Journal of Geophysical Research: Biogeosciences*, *116*(G3).

Lacelle, D., Brooker, A., Fraser, R.H. and Kokelj, S.V., 2015. Distribution and growth of thaw slumps in the Richardson Mountains–Peel Plateau region, northwestern Canada. *Geomorphology*, *235*, pp.40-51.

Lantz, T.C. and Kokelj, S.V., 2008. Increasing rates of retrogressive thaw slump activity in the Mackenzie Delta region, NWT, Canada. *Geophysical Research Letters*, *35*(6).

Lewkowicz, A.G. and Way, R.G., 2019. Extremes of summer climate trigger thousands of thermokarst landslides in a High Arctic environment. *Nature communications*, *10*(1), p.1329.

Li, Y., Qi, H., Dai, J., Ji, X. and Wei, Y., 2017. Fully convolutional instance-aware semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2359-2367).

Lin, Y. and Knudby, A.J., 2023. A transfer learning approach for automatic mapping of retrogressive thaw slumps (RTS) in the western Canadian Arctic. *International Journal of Remote Sensing*, *44*(6), pp.2039-2063.

Martins, L., Guede-Fernández, F., Valente de Almeida, R., Gamboa, H. and Vieira, P., 2022. Real-Time Integration of Segmentation Techniques for Reduction of False Positive Rates in Fire Plume Detection Systems during Forest Fires. *Remote Sensing*, *14*(11), p.2701.

Meredith, M., Sommerkorn, M., Cassotta, S., Derksen, C., Ekaykin, A., Hollowed, A., Kofinas, G., Mackintosh, A., Melbourne-Thomas, J., Muelbert, M.M.C. and Ottersen, G., 2019. Polar regions. chapter 3, ipcc special report on the ocean and cryosphere in a changing climate.

Metcalfe, D.B., Hermans, T.D., Ahlstrand, J., Becker, M., Berggren, M., Björk, R.G., Björkman, M.P., Blok, D., Chaudhary, N., Chisholm, C. and Classen, A.T., 2018. Patchy field sampling biases understanding of climate change impacts across the Arctic. *Nature ecology & evolution*, *2*(9), pp.1443-1448.

Mora, O.E., 2015. *Morphology-Based Identification of Surface Features to Support Landslide Hazard Detection Using Airborne LiDAR Data*. The Ohio State University.

Mora, O.E., Lenzano, M.G., Toth, C.K., Grejner-Brzezinska, D.A. and Fayne, J.V., 2018. Landslide change detection based on multi-temporal Airborne LiDAR-derived DEMs. *Geosciences*, *8*(1), p.23.

Nitze, I., Grosse, G., Jones, B.M., Romanovsky, V.E. and Boike, J., 2018. Remote sensing quantifies widespread abundance of permafrost region disturbances across the Arctic and Subarctic. *Nature communications*, *9*(1), p.5423.

Nitze, I., Heidler, K., Barth, S. and Grosse, G., 2021. Developing and testing a deep learning approach for mapping retrogressive thaw slumps. *Remote Sensing*, *13*(21), p.4294.

Noh, M.J. and Howat, I.M., 2014. Automated Coregistration of Repeat Digital Elevation Models for Surface Elevation Change Measurement Using Geometric Constraints, IEEE T. Geosci. Remote, 52, 2247–2260.
Noh, M.J. and Howat, I.M., 2015. Automated stereo-photogrammetric DEM generation at high latitudes: Surface Extraction with TIN-based Search-space Minimization (SETSM) validation and demonstration over glaciated regions. *GIScience & Remote Sensing*, *52*(2), pp.198-217.
Noh, M.J. and Howat, I.M., 2017. The surface extraction from TIN based search-space minimization (SETSM) algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing*, *129*, pp.55-76.
Noh, M.J. and Howat, I.M., 2019. Applications of High-Resolution, Cross-Track, Pushbroom Satellite Images With the SETSM Algorithm. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(10), pp.3885-3899.

Nuth, C. and Kääb, A., 2011. Co-registration and bias corrections of satellite elevation data sets for quantifying glacier thickness change. *The Cryosphere*, *5*(1), pp.271-290.

Olefeldt, D., Goswami, S., Grosse, G., Hayes, D., Hugelius, G., Kuhry, P., McGuire, A.D., Romanovsky, V.E., Sannel, A.B.K., Schuur, E.A.G. and Turetsky, M.R., 2016. Circumpolar distribution and carbon storage of thermokarst landscapes. *Nature communications*, *7*(1), p.13043.

Ramage, J.L., Irrgang, A.M., Herzschuh, U., Morgenstern, A., Couture, N. and Lantuit, H., 2017. Terrain controls on the occurrence of coastal retrogressive thaw slumps along the Yukon Coast, Canada. *Journal of Geophysical Research: Earth Surface*, *122*(9), pp.1619-1634.

RGI 7.0 Consortium, 2023. Randolph Glacier Inventory - A Dataset of Global Glacier Outlines, Version 7.0. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. doi:10.5067/f6jmovy5navz. Online access: **https://doi.org/10.5067/f6jmovy5navz**

Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M. and Guzzetti, F., 2018. A review of statistically-based landslide susceptibility models. *Earth-science reviews*, *180*, pp.60-91.

Ren, S., He, K., Girshick, R. and Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, *28*.

Rudy, A.C.A., Lamoureux, S.F., Kokelj, S.V., Smith, I.R. and England, J.H., 2017. Accelerating thermokarst transforms ice‑cored terrain triggering a downstream cascade to the ocean. *Geophysical Research Letters*, *44*(21), pp.11-080.

Schuur, E.A., McGuire, A.D., Schädel, C., Grosse, G., Harden, J.W., Hayes, D.J., Hugelius, G., Koven, C.D., Kuhry, P., Lawrence, D.M. and Natali, S.M., 2015. Climate change and the permafrost carbon feedback. *Nature*, *520*(7546), pp.171-179.

Segal, R.A., Lantz, T.C. and Kokelj, S.V., 2016a. Acceleration of thaw slump activity in glaciated landscapes of the Western Canadian Arctic. *Environmental research letters*, *11*(3), p.034025.

Segal RA, Lantz TC, and Kokelj SV, 2016b. Inventory of active retrogressive thaw slumps in the Peel Plateau, Northwest Territories. Northwest Territories Geological Survey, https://doi.org/10.46887/2015-020

Segal RA, Lantz TC, and Kokelj SV, 2016c. Inventory of active retrogressive thaw slumps on eastern Banks Island, Northwest Territories. Northwest Territories Geological Survey, https://doi.org/10.46887/2015-021.

Segal, RA, Lantz, TC, Kokelj, SV, Brietzke, CK, Reichheld, S, and Martin, AF, 2017. Inventory of retrogressive thaw slumps in the Tuktoyaktuk Coastlands, Western Anderson Plain, and Mackenzie Delta region, Northwest Territories. Northwest Territories Geological Survey, https://doi.org/10.46887/2015-022

Serreze, M.C. and Barry, R.G., 2011. Processes and impacts of Arctic amplification: A research synthesis. *Global and planetary change*, *77*(1-2), pp.85-96.

Shean, D.E., Alexandrov, O., Moratto, Z.M., Smith, B.E., Joughin, I.R., Porter, C. and Morin, P., 2016. An automated, open-source pipeline for mass production of digital elevation models (DEMs) from very-high-resolution commercial stereo satellite imagery. *ISPRS Journal of Photogrammetry* and Remote Sensing, 116, pp.101-117.

Stumpf, A. and Kerle, N., 2011. Object-oriented mapping of landslides using Random Forests. *Remote sensing of environment*, *115*(10), pp.2564-2577.

Tsangaratos, P. and Ilia, I., 2016. Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi Perfection, Greece. *Landslides*, *13*, pp.305-320.

Porter, C., Morin, P., Howat, I., Noh, M.J., Bates, B., Peterman, K., Keesey, S., Schlenk, M., Gardiner, J. and Tomko, K., ArcticDEM. version 3. 2018. Harvard Dataverse, 1.

Porter, Claire; Howat, Ian; Noh, Myoung-Jon; Husby, Erik; Khuvis, Samuel; Danish, Evan; Tomko, Karen; Gardiner, Judith; Negrete, Adelaide; Yadav, Bidhyananda; Klassen, James; Kelleher, Cole; Cloutier, Michael; Bakker, Jesse; Enos, Jeremy; Arnold, Galen; Bauer, Greg; Morin, Paul, 2022, "ArcticDEM − Strips, Version 4.1", https://doi.org/10.7910/DVN/C98DVS, Harvard Dataverse, V1, [Date Accessed Nov. 3, 2022].

Van der Sluijs, J., Kokelj, S.V., Fraser, R.H., Tunnicliffe, J. and Lacelle, D., 2018. Permafrost terrain dynamics and infrastructure impacts revealed by UAV photogrammetry and thermal imaging. Remote Sensing, 10(11), p.1734.

van der Sluijs, J., Kokelj, S.V. and Tunnicliffe, J.F., 2022. Allometric scaling of retrogressive thaw slumps. *The Cryosphere Discussions*, pp.1-30.

van der Sluijs, J., Kokelj, S.V., 2023. A detailed inventory of retrogressive thaw slump affected slopes using high spatial resolution digital elevation models and imagery, Peel Plateau and Anderson Plain – Tuktoyaktuk Coastlands, Northwest Territories. Northwest Territories Geological Survey, https://doi.org/10.46887/2023-013

Wilson, J.P. and Gallant, J.C., 2000. Terrain Analysis, Principles and Applications, Jhon Wiley&Sons. *Inc. Usa*. pp. 57.

Witharana, C., Bhuiyan, M.A.E., Liljedahl, A.K., Kanevskiy, M., Epstein, H.E., Jones, B.M., Daanen, R., Griffin, C.G., Kent, K. and Jones, M.K.W., 2020. Understanding the synergies of deep learning and data fusion of multispectral and panchromatic high resolution commercial satellite imagery for automated ice-wedge polygon detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, *170*, pp.174-191.

Witharana, C., Udawalpola, M.R., Liljedahl, A.K., Jones, M.K.W., Jones, B.M., Hasan, A., Joshi, D. and Manos, E., 2022. Automated Detection of Retrogressive Thaw Slumps in the High Arctic Using High-Resolution Satellite Imagery. *Remote Sensing*, *14*(17), p.4132.

Yang, Y., Rogers, B.M., Fiske, G., Watts, J., Potter, S., Windholz, T., Mullen, A., Nitze, I. and Natali, S.M., 2023. Mapping retrogressive thaw slumps using deep neural networks. *Remote Sensing of Environment*, *288*, p.113495.

Wessel, P. and Smith, W.H., 1991. Free software helps map and display data. *Eos, Transactions American Geophysical Union*, 72(41), pp.441-446.

Wu, Y., Kirillov, A., Massa, F., Lo, W.Y. and Girshick, R., 2019. Detectron2. Available online: https://github.com/facebookresearch/detectron2 (last accessed August 2, 2023).