Supplement to "Arctic shoreline displacement with open satellite imagery and data fusion: A pilot study 1984–2022"

Tua Nylén[1,2], Mikel Calle[1], Carlos Gonzales-Inca[1]

[1]Department of Geography and Geology, University of Turku, Turku, FI-20014, Finland
[2]Department of Geosciences and Geography, University of Helsinki, Helsinki, FI-00014, Finland
[3]Turku Collegium for Sciences, Medicine and Technology (TCSMT), University of Turku, Turku, FI-20014, Finland

Correspondence to: Tua Nylén (tua.nylen@utu.fi)

Study areas

We used two contrasting coastal environments in the European Arctic to calibrate and validate the procedure for extracting shoreline displacement time series. First, the procedure was iteratively calibrated, tested and validated in the Tanafjorden area, at the Barents Sea coast of Norway (Figure S1). Second, the final procedure was validated in the Ny-Ålesund area at the Greenland Sea coast of the Svalbard archipelago (Figure S1).

The first study area in Tanafjorden (Deanuvuotna in northern Sámi language, Tenovuono in Finnish; c. 70.5° N, 28.5° E; Figure S1), is located in the Fennoscandian Barents Sea coasts of northern Finnmark, Norway. The study area included south-western parts of the Nordkinn Peninsula (Nordkinnhalvøya), the Tana River delta, and north-western parts of the Varanger Peninsula (Varangerhalvøya). The area was selected since it was relatively easy to access during the COVID pandemic and provided varying landscape, including a major delta. We used sea level data from the nearest tide gauge, located in Honningsvåg c. 100 km from the study area and operated by the Norwegian Mapping Authority. The study area is characterised by steep fjord landscapes, the highly dynamic delta, lack of glaciers and sea ice, and mesotidal conditions.

The Tanafjorden coast is in most parts dominated by steep cliffs. Where rivers enter the fjords, the topography is generally flatter and regular water level fluctuations create tidal flats and salt marshes. The largest river in the area is Tana River (Deatnu in northern Sámi language, Tenojoki in Finnish). It is the border river between Norway and Finland and one of the largest rivers in Norway. While there are some snowbeds in the mountains (patches of snow that melt late in the summer), there is no sea ice (data by National Snow and Ice Data Center), no glaciers (Andreassen et al., 2012), and no coastal permafrost in the area (Gisnås et al., 2017). Rivers entering the Barents Sea and their brackish deltas have an ice cover from mid-October to mid-May.

The Tanafjorden study area has a humid subarctic climate. The annual mean temperature in Tanafjorden is around 1 °C and the mean annual precipitation around 500 mm (Norwegian Centre for Climate Services). The coast has a seasonal snow cover and is mainly free of snow from mid-June to mid-October. The area is characterized by a semidiurnal lunar tide with a period of 12 hours and 25 minutes and mean tidal range of 2.3 meters during spring tide (Norwegian Mapping Authority).

The second study area is located around the town of Ny-Ålesund in Svalbard (c. 78.9° N, 11.9° E; Figure S1). The study area covers southern parts of the Mitrahalvøya peninsula, Krossfjorden, southern parts of Haakon VII Land, Kongsfjorden, and northern parts of the strait of Forlandsundet and the island of Prins Karls Forland. Data from the Norwegian Mapping Authority tide gauge located in the town of Ny-Ålesund was used for examining sea level. The coasts area characterized by glaciers, seasonal sea ice, polar climate, and microtidal conditions.

In the Ny-Ålesund study area there are nine glaciers that have in the past flowed onto the ocean forming ice shelves. The ice shelves have retreated during the past decades, and a few of them have shifted from marine-terminating to land-terminating glaciers. The retreat has created new shorelines and pristine landscapes (Kavan & Strzelecki, 2023). The ocean is generally ice-free throughout the year, but parts of the nearshore freeze in winter and spring (Gerland et al., 2022; National Snow and Ice Data Center). There is widespread permafrost in the coast (Humlum et al., 2003).

In Longyearbyen, located 110 km south-east from Ny-Ålesund, the annual mean temperature is -5.9 °C and the measured annual mean precipitation 200 mm (Norwegian Centre for Climate Services). The mean tidal range during spring tide is 1.3 meters (Norwegian Mapping Authority).

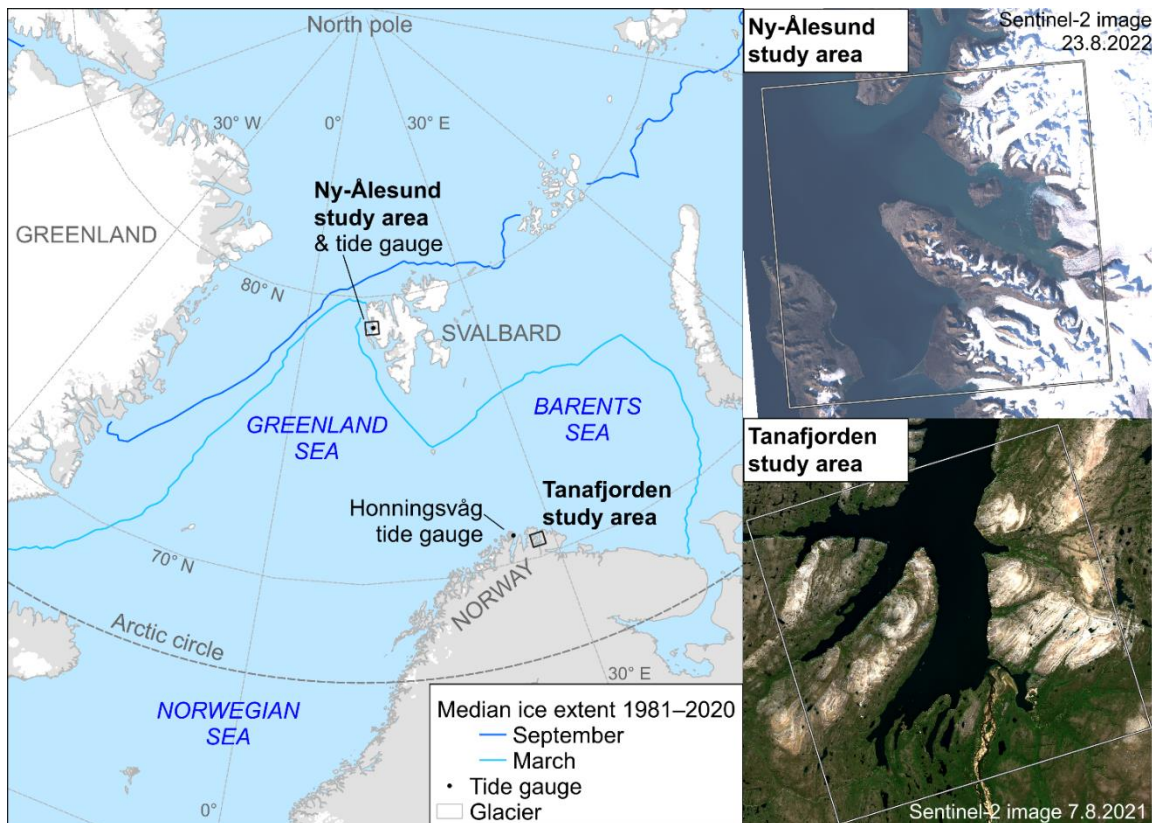



Figure S1. The two study areas: Tanafjorden in northern Fennoscandia, the Norwegian Barents Sea coast, and Ny-Ålesund in the high-arctic Svalbard. The size of each study area is 2500 km² (50 km * 50 km in the local UTM coordinate reference system). Spatial data: National Snow and Ice Data Center and Copernicus Sentinel.

Iterative process-building for Google Earth Engine

Initial procedure-building iterations were done on a desktop, using original Landsat and Sentinel-2 image files for the
50 Tanafjorden study area, and desktop software like R and QGIS. During these iterations, we examined the usability of individual image collections, compared level 1 (top-of-atmosphere) and level 2 (bottom-of-atmosphere) products, optimized the image collection filtering criteria, examined the suitability of existing pan-Arctic spatial data for data fusion, performed preliminary testing of water detection methods and identified remaining classification and performance challenges that should be dealt with during further iterations. The tested water detection methods included single-band thresholding
55 methods, multispectral indices and several unsupervised and supervised classification algorithms.

Each iteration included an examination of the impact of the changes on land cover classification accuracy, accuracy of the fitted shoreline and on processing time. The results of these preliminary test are not reported in this paper.

Based on the results of these first iterations, we constructed the initial processing chain on Google Earth Engine (GEE). Using GEE was necessary for upscaling the analyses to the entire Arctic coast and to the entire period of 40 years. On
60 desktop, the size of files accumulating on the local hard drive would have prevented larger-scale processing. Computing restrictions would have limited the analyses to the few highest-quality images and make it impossible to utilize all suitable data. Using a dense time series would be needed to remove noise and the impact of short-term water level fluctuations. Currently, it is not possible to build automated time series analyses based on the Copernicus Open Access Hub. Since 2019, most of the Sentinel images are archived in the Long Term Archive and need to be ordered and then downloaded after a wait
65 of some hours. This creates a notable lag in any process and requires intervention by the user.

Subsequent iterations aimed at applying the pre-selected processing chain in the GEE JavaScript API, comparing five alternative water detection methods, applying data fusion, algorithm fusion and decision fusion, and solving remaining classification and performance challenges. This paper reports the most consequential results of these iterations, as well as the results of the final, proposed procedure. The procedure utilized the petabyte-size satellite image collections and existing
70 spatial data held in GEE. Those datasets were complemented by uploading to GEE pre-calculated pan-Arctic layers that describe the one-kilometer and two-kilometer zones around the modern shoreline. They were used for determining the area of interest for the analyses. The results were exported to Google Drive for storage, sharing and further analysis.

Proposed procedure

We generally aimed at creating shoreline displacement data at the pan-Arctic scale. With GEE, analyses can be run globally,
75 and GEE divides export tasks into tiles (i.e., sub-areas) that are processed in parallel. The outputs can later be merged to create pan-Arctic coverage. In addition, we aimed at providing reproducible code for working with single study areas, which is a more typical application of the methodology. Several study area sizes were tested before suggesting an optimal area of 50 km * 50 km. Smaller analysis areas would be recommended when working on the desktop, and also in GEE the outputs

for smaller study areas would be quicker to export. Larger study areas, such as 110 km * 110 km corresponding to Sentinel-2

80   tile size, were found to be too large for efficient processing and exports.

The final, proposed procedure for creating time series of coastal landcover and shorelines without user input can be divided into nine steps (Figure S2). The steps are described below in detail.

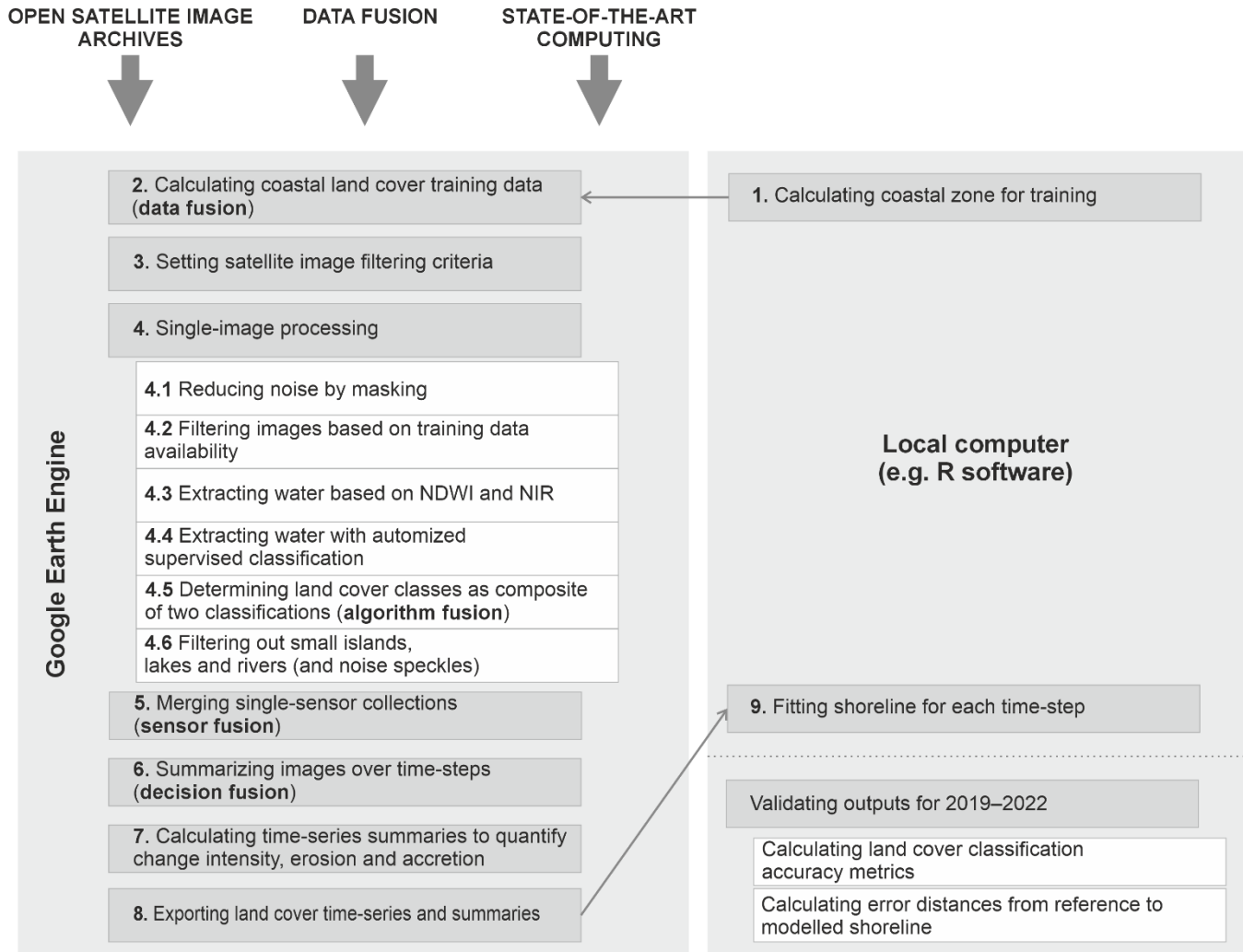



Figure S2. Overview of the proposed procedure, divided into nine steps and leading to validation of the results.

85   **Steps 1 and 2: Calculating coastal land cover training data**

The procedure calculated training data for supervised classification of satellite images automatically from existing spatial data. The goal was to identify pixels from each satellite image, that have always been land or sea for the past 40 years. As the coast is expected to be dynamic in some areas, pixels located very close to the modern-day shoreline were not used for training. Preliminary results indicated that land areas with deep mountain shadows are difficult to classify correctly. Thus,

90 we created a three-class training dataset: water, land and mountain shadows. We classified images first into these three classes and then merged the land and mountain shadow classes.

First, all pixels were classified either as land or as water based on the MERIT Hydro permanent water dataset (Yamazaki et al., 2019) hosted in GEE. Second, we determined the most dynamic coastal zone as a 1000-meter buffer polygon of the OpenStreetMap coastline (FOSSGIS e.V., 2023). The buffer polygon was created and simplified (to reduce computing cost)

95 in the R program and then uploaded into GEE as an asset. Third, land and water pixels closer than 1000 meters to the OpenStreetMap coastline were masked out from the training data.

Fourth, mountain shadows were calculated separately for each image from the ArcticDEM, a two-meter digital elevation model of the Arctic (Porter et al., 2018), also hosted in GEE. The elevation model was first cleaned up from outliers located in the sea, by using the MERIT Hydro dataset. The mountain shadows were calculated with the ee.Terrain.hillShadow

100 function, using image-specific sun aspect and zenith information. Since most of the problematic mountain shadows are both located very close to the shoreline and are relatively stable, none of the mountain shadow pixels were masked out from the training data layer. Finally, the land/water and mountain shadow layers were combined into one three-class training data layer (Figure S3).

In GEE, training points were randomly picked from the training data layer separately for each satellite image. A stratified

105 random sampling strategy (ee.Image.stratifiedSample function) was used to extract 500 points from each class within the image footprint.

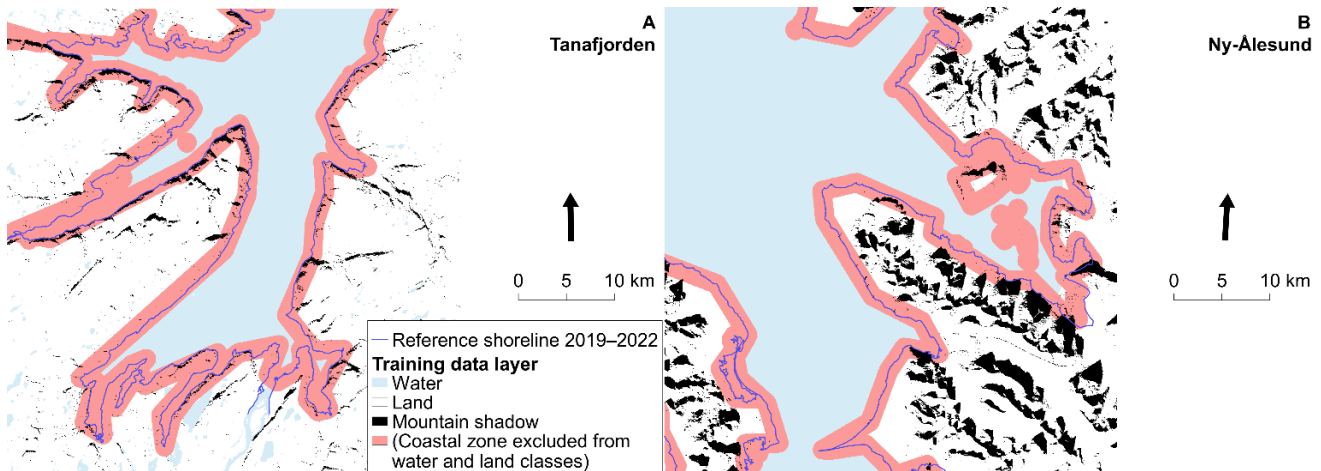



Figure S3. Automatically generated training data layer with three classes: water, land and mountain shadows. The mountain shadows have been calculated for hypothetical images taken on August 31 at 10 am UTC in Tanafjorden and at 1 pm UTC in Ny-
110 **Ålesund (typical acquisition times).**

Step 3: Setting satellite image filtering criteria

Open multispectral satellite image time series produced by different sensors were examined and compared, to see which time series produced useful and accurate information on long-term shoreline changes. The image collections were also merged to

increase data availability and the results from the merged collection were compared to single-sensor results. Since the
shoreline extraction was based on post-classification decision fusion, data harmonization between sensors was not necessary.

Five multispectral image collections were utilized in this study through GEE. Four of these were Landsat surface reflectance collections by the US Geological Survey (level 2, collection 2, tier 1). They were based on the Thematic Mapper sensor (TM) onboard Landsat 4 and Landsat 5, the Landsat 7 Enhanced Thematic Mapper Plus sensor (ETM+) and the Landsat 8 Operational Land Imager sensor (OLI; Table S1). The fifth collection was the Sentinel-2 Multispectral Instrument (MSI) surface reflectance collection (level 2) of the Copernicus program. The Landsat 4 collection held in GEE was empty for the study areas.

During previous iterations we examined the possibility of completing the tier 1 Landsat 5 and Landsat 7 collections with tier 2 images. This was tempting, since it would increase the number of observations, as sometimes the number of observations was too low for creating a summary for an individual time-step. The categorization of images into tiers is based on the root mean square error of the horizontal accuracy. For Landsat 7, for example, tier 2 images have a root mean square error of more than 12 meters. The results suggested that using the tier 2 images added considerable errors to the results and cannot be generally recommended. However, for specific uses and single study areas including those lower-quality images could be an option to consider, if tier 1 collection sizes are too small for time-step summaries.

Together, the four non-empty collections in GEE covered a period from March 1984 to the present. The data of Landsat 7 was considerably deteriorated after the scan line corrector failure in May 2003, but the Landsat 7 collection was still analysed. In the two study areas, we worked with time series that covered a period from June 1984 to September 2022 (Table S1). The spatial resolution of utilized bands was 30 meters in all Landsat products and 10 or 20 meters in the Sentinel product (Table S1).

Table S1. Satellite image collections and bands utilized in the case study for Tanafjorden (T) and Ny-Ålesund (N).

Collection	Appr. local acquisition time (UTC)		Corresponding sun angles 1.6.-30.9.				Filtered collection time span		Spatial resolution	
			Azimuth		Zenith		First	Last	Green, NIR (blue, red)	SWIR1 and SWIR2
	T	N	T	N	T	N				
Landsat 5 TM	9:45 am	11:45 am	171-177°	183-195°	17-43°	8-34°	7.6.1984	29.9.2011	30	30
Landsat 7 ETM+	10:00 am	12:30 pm	176-180°	199-202°	17-43°	8-34°	15.9.1999	26.9.2022	30	30
Landsat 8 OLI	10:00 am	12:30 pm	176-180°	199-202°	17-43°	8-34°	7.6.2013	29.9.2022	30	30
Sentinel-2 MSI	10:30 am	13:00 pm	185-189°	207-210°	16-43°	7-33°	6.6.2017	30.9.2022	10	20

135

With filtering criteria, we aimed at selecting all images that had valuable information on coastal land cover, but excluding images where illumination was not sufficient, where the Earth's surface was covered by clouds or where the sea was covered by sea ice. The selection of the optimal filtering criteria was done iteratively.

In the final procedure, the satellite image collections were filtered based on two criteria. First, images with a cloud percentage of less than 50 % were included in the selection. This threshold was selected to maximize the number of observations on individual pixels, but to exclude images with a large proportion of noisy pixels. Even though most clouds and cloud shadows could be masked from the images at a later stage, with increasing cloud percentage, also the number of undetected clouds increased, thus leading to increasing number of falsely classified pixels. For Sentinel-2 images the maximum cloud cover was set lower, to 30 %, since the shorter revisit time compared to Landsat satellites meant that there were plenty of available images. Second, the collections were filtered into images taken during the four Arctic summer months, namely June, July, August and September. Despite good light conditions, spring images were excluded due to high snow and sea ice cover. Late autumn images were excluded because of insufficient illumination. In Ny-Ålesund, for example, the sun would be only 8° above the horizon on 1.10. at solar noon, and the polar night begins in mid-October.

At a later stage we also ensured that there were enough valid pixels for supervised classification in each image, and images that did not meet this criterium were further filtered out from the image collections. Without this step, supervised classification would have thrown an error. This was after masking noisy pixels (see step 3), by calculating the number of

valid (not masked) image pixels that overlapped with the land and water pixels in the training data layer. Images, that had less than 100 km² of overlap with either land or water training pixels were dropped out from the image collections.

For pan-Arctic analyses or large areas, data availability can be quantified by calculating the total number of observations on each pixel and exporting this information. This can be calculated before and after masking noisy pixels, and for the entire time series or for specific time-steps. When working with our two study areas, the image collections were also filtered with the boundaries of the study area, to examine and report the number of available images. After masking noisy pixels at a later stage in the process, those images that had very few useful pixels, covering less than 90 km² of the study area, were excluded and the number of remaining images was calculated.

Step 4: Single-image processing

The final processing pipeline for individual multispectral images consisted of following steps:

First, two most useful spectral bands, namely green and near infrared (NIR), and the pixel quality band were selected and renamed. Sun angle metadata of the images, namely sun azimuth and zenith, were calculated and renamed.

Second, the amount of noise was reduced by masking individual pixels that had been flagged as clouds, cloud shadows, fill pixels and image edge noise, using pixel quality information and reflectance values. For Landsat images, the pixel quality attribute band, generated from the CFMASK algorithm, was used. For Sentinel-2 images, pixel quality was determined from the scene classification band.

Third, the image was classified into water and land pixels based on the best-performing multispectral water index. We chose the normalized difference water index, NDWI, for the final procedure. The preliminary tests indicated that the classification accuracy of NDWI was among the highest (very similar results with other indices) and it utilizes bands that had highest spatial resolution (10 meters in Sentinel-2 images). Binary land and water classes were calculated from NDWI values using a threshold value of 0. Previous iterations used Otsu's and Huang's methods to identify the optimal threshold value. We noticed, however, that the difference in the outputs was minimal and the added computing cost of determining the threshold was notable.

For comparison, the images were also classified for this paper based on the modified difference water index (MNDWI), the Automated Water Extraction Index (AWEI) and WIBI:

- Normalized difference water index, NDWI = (green − NIR) / (green + NIR)
- Modified normalized difference water index, MNDWI = (green − SWIR1) / (green + SWIR1)
- Automated Water Extraction Index, AWEI = blue + 2.5*green - 1.5*(NIR + SWIR1) − 0.25*SWIR2
- WIBI = NDWI − NDBI, where NDBI = (SWIR1 − NIR) / (SWIR1 + NIR) (Ismail et al. 2022)

Fourth, we used single-band reflectance information to improve the classification of water and snow or ice. After examining the validation results in the Ny-Ålesund study area, we identified a need to improve this distinction. Water is difficult to distinguish from snow or ice based on NDWI alone since they all have a high visible to infrared reflectance ratio. However, the difference in reflectance of water and snow or ice in any single wavelength is notable. Thus, we determined sensor-

185 specific thresholds for the NIR wavelength for separating water and snow or ice pixels (Table S2). Pixels that had a high NDWI but also a NIR band reflectance above the threshold, were assigned to the snow or ice class. The threshold was set to 10–20 % of the average maximum green band reflectance in the glaciated Ny-Ålesund area. The average maximum reflectance was calculated from the median composite (1984–2022) separately for each sensor for Ny-Ålesund (and for Tanafjorden for reference). Consequently, the fixed NIR thresholds listed in Table S2 were applied to all images. For this

190 paper, we then combined the land and ice classes into one land class to simplify the visualisation of the outputs. In the future, the procedure will be made more applicable to coasts that are surrounded by all-year sea ice by adopting a three-class land cover classification with classes land, water, and ice.

Fifth, we extracted water based on supervised classification with the random forest algorithm. The automatically generated training data layer (land and water outside the 1-km coastal zone, and mountain shadows) was used for extracting the

195 random training sample of 1500 points (500/class). In the final procedure, the random forest classifier was trained using information from two bands only: NIR and NDWI. This band selection among all common bands was based on examination of variable importance in the random forest classifier and the spatial resolution of the bands: bands with 10 meter resolution in the Sentinel-2 images were preferred. The number of bands was reduced to minimum to minimize computing cost. After the initial classification, the land and mountain shadow classes were merged.

200 Sixth, we calculated the final land cover layer as a composite of the two classifications, NDWI and random forest classes. The image-specific mountain shadow layer was used for determining the composite. This algorithm fusion aimed at improving the overall classification accuracy by taking advance of the contrasting strengths of the two classification approaches: NDWI provided good results in sunlit and shallow areas, while the supervised classification was more accurate in mountain shadows.

205 Seventh, we filtered out small islands, lakes and rivers. This was done by calculating the size of each cluster of pixels belonging to the same class. Small clusters were then assigned to the other class, e.g., small clusters of water pixels (pond) to the land class.

While masking of noisy pixels is important when working with individual images, summarizing a large number of images does reduce the impact of noise. However, our tests showed that masking is necessary also for the quality of the time series

210 summary, particularly for areas and time-steps that have been observed only a few times. Outputs derived from unmasked image collection would contain a number of misclassified pixels, particularly cloudy water pixels classified as land. Our tests also indicated the random forest algorithm cannot be reliably trained for land and water classification for images that had a small number of valid pixels within the coastal zone, i.e., the area-of-interest.

Table S2. Maximum NIR band pixel values for each image collection within the study areas and the threshold value (c. 10–20 % of the Ny-Ålesund maximums) used for mapping snow and ice that have a high reflectance in individual bands but NDWI values similar to water.

Sensor	Maximum NIR band reflectance of the median composite 1984–2022		Snow/ice threshold
	Tanafjorden	Ny-Ålesund	
Landsat 5	40 000	54 500	10 000
Landsat 7	38 000	54 500	10 000
Landsat 8	37 000	56 000	10 000
Sentinel-2	7 200	10 400	1 000

Step 5: Merging single-sensor image collections (sensor fusion)

Single-sensor image collections were processed separately due to sensor-specific settings. The outputs (coastal land cover classification accuracy and missing value statistics) of those single-sensor collections were reported for this paper. However, the final procedure merged all classified images to one large collection, to maximize the number of observations on each pixel.

Step 6: Summarizing images over time-steps (decision fusion)

To reduce the impact of noise remaining in single images on classification accuracy, and the impact of short-term water level changes on the shoreline, we summarized dozens of images taken within a specific time frame. The conceptual framework for this temporal decision fusion was taken from Laegner et al. (2019). In other words, the satellite images were grouped into time-steps to ensure that there were as many observations as possible to make reliable decisions on long-term mean conditions and changes.

In the selection of the time-step length, we aimed at balancing the number of time-steps and the number of available images per time-step. As a result, the images were divided into eight five-year time-steps based on their acquisition time: 1984–1988, 1989–1993, 1994–1998, 1999–2003, 2004–2008, 2009–2013, 2014–2018 and 2019–2022. Five years was considered long enough to have at least 10 acceptable images per time-step also in the 80s and 90s, when the Landsat satellite revisit time was long. On the other hand, five years was considered short enough to gather enough details about the dynamic changes in the coastal landscape.

To make the final class assignment of a pixel, we calculated the number of observations where the pixel was classified as land during one time-step and divided it with the number of valid observations (c.f. Laegner et al. 2019). Pixels with less than five valid observations during the time-step were classified as missing data. Pixels that had been classified as land in at least 50 % of the individual images were assigned to the land class and the remaining pixels were assigned to the water class.

The procedure then created one time series image with eight bands, each corresponding to one of the time-step land cover layers.

Step 7: Calculating time series summaries to quantify change intensity, erosion and accretion

We calculated three time series summary variables in GEE to quantify change intensity, erosion and accretion. First, we calculated the probability of the pixel to belong to the water class (sum of time-steps in water class / number of time-steps with valid observations). This variable highlighted pixels that either experienced true coastal change during the 40-year period, or where misclassified during one or several time-steps.

Second, we calculated the long trend between two 20-year periods: 1984–2003 and 2004–2022. This was done by dividing the land cover time series into two halves and summarizing them in the same way as single images for five-year time-steps (step 5). The long trend was then calculated by subtracting the second half land cover layer from the first half land cover layer. This variable identified pixels that on average shifted from the land class to the water class and vice versa, as well as pixels with no change.

Third, we quantified change intensity by dividing the number of class shifts (land to water or water to land) within the time series by the number of time-step observations. This variable highlighted pixels that changed class several times during the time series, including pixels in highly dynamic coastal or estuarine areas or intertidal zone, or pixels that were often misclassified due to some random or systematic reasons. These three summary layers were then combined into one summary image.

We also demonstrated how the time series can be used to calculate further summaries outside GEE, by calculating four additional summary layers in R. We calculated the first and last time-step when a pixel was classified as land and the first and last time-step when a pixel was classified as water. The two "last time-step" summaries described the timing of permanent changes in the coastline, such as cliff erosion (last land time-step) or delta accretion (last water time-step). The two "first time-step" summaries described the timing of both permanent changes and those changes that were later reversed, such as sand bars migrating through the delta.

Step 8: Exporting the land cover time series and time series summaries

The land cover time series image, the observation count time series and the summary image were then exported from GEE to Google Drive as two GeoTIFF raster images. They were exported at the 30-meter resolution that corresponds to the pixel size of the Landsat collections. GeoTIFFs exported from GEE do not store missing value information: missing values are exported as zeros. Thus, the land cover class values were encoded to 1 for land and 2 for water.

As GEE performs the actual server-side processing on-demand, the exporting (or visualization) step is the one that reveals if the procedure is efficient enough to be processed. For this reason, the majority of the process-building iterations were dedicated to lowering the computing cost. The starting point was that the preselected procedure was too computationally heavy and resulted in errors related with exceeded memory limits and computation timing out. Therefore, the scaling of the

code was improved in the final procedure to minimize the computing time and memory usage. This was mainly done by avoiding calculations that use large amounts of pixel information simultaneously, such area reduceRegion calculations for large regions and with high resolution and avoiding using complex geometries. The number of bands used for calculations was also optimized to reduce computing cost.

GEE handles exports for large areas by automatically splitting them into tiles. The number of these tiles is determined by the area of the export region and the pixel size. However, a typical application of this procedure is that the users have a specific study area, and the export is, accordingly, restricted to this area. For this paper, the results were exported for the two study areas, each with a size of 2 500 km² (50 km in both east-west and north-south directions) or c. 2 800 000 pixels. The processing and export of the eight-band time series for such an area took c. 20–50 minutes depending on the processing queue. The time series and the summary layer could be processed in parallel. Processing would be even faster for smaller study areas.

Step 9: Fitting shoreline for each time-step

Tests for fitting the shoreline to the land cover data at the 30-meter resolution in GEE failed. Raster data is converted to vector data in GEE using the ee.Image.reduceToVectors function. This approach is basically based on having all pixels of the area-of-interest in the memory simultaneously, to determine the borders between pixel value classes. For large areas, such as our two study areas, the number of pixels for the memory. This would have been a possible solution if the study area size could have been decreased or the pixel size could have been increased, but none of these options were acceptable for our goals.

Instead, we fitted the shoreline to the land cover raster data outside GEE. We chose to fit the shoreline as an isoline ("contour") running between the land (1) and water (2) pixel values, using the rasterToContour function of the R package raster (Hijmans et al. 2023). The resulting isolines followed the pixel borders and therefore had an unnaturally jagged shape (Figure S4). They were smoothed to better represent the natural shapes of the shoreline (Figure S4). The lines were smoothed using the ksmooth Gaussian kernel smoothing method implemented in the smooth function of the R package smoothr (Strimas-Mackey 2023). The smoothness parameter was optimized in previous iterations and was set to 2. The smoothed lines were then simplified (lowered the number of polyline vertices to minimize file size), with a one-meter tolerance value using the st_simplify function from the R package sf (Pebesma et al. 2023).

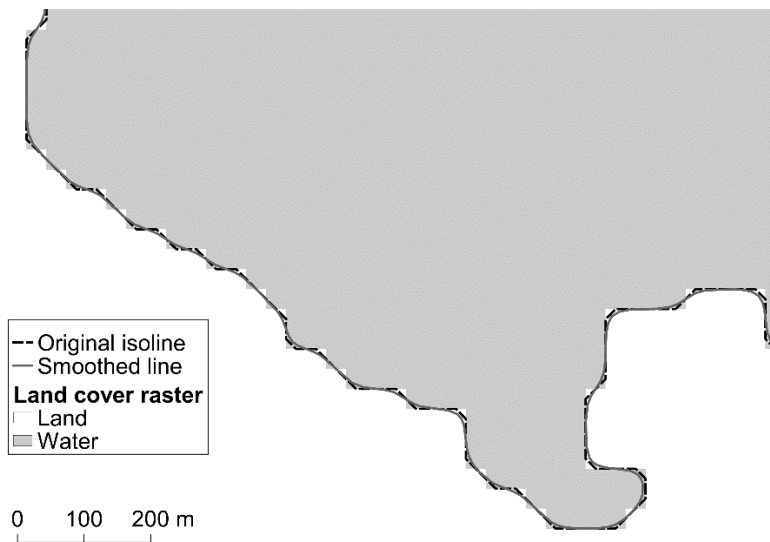



Figure S4. Example of fitting the shoreline to the classified raster image. The shoreline was first fitted as an isoline and then smoothed using a Gaussian kernel smoothing, to reduce jagged shapes.

Validation data of modern shoreline

Independent validation data were created for examining the accuracy of the land cover classification and shoreline detection. Two types of validation data were created, both corresponding to mean sea level conditions during c. 2019–2022. First, the coastal zones (excluding lakes and ponds) of the two study areas were randomly sampled and the random points were manually classified into land and sea. Second, the shorelines of the study areas were digitized by hand. These validation data were created by the same coastal geomorphologist to maximize the homogeneity of the data. The interpretations were mainly based on hand-picked remote sensing images. For Tanafjorden, the interpretation was also based on in situ observations from October 2021.

Ortophotograph mosaics and selected Sentinel-2 multispectral satellite images were used to digitize the shoreline and the land cover class of the random validation points. In some challenging areas, the interpretations were confirmed from geotagged photographs found through the internet. The ortophotograph was published for Tanafjorden by the Norwegian Mapping Authority in 2016 (updated 2020) and for Svalbard by the Norwegian Polar Institute (no date information; https://geodata.npolar.no/). They both had a high spatial resolution (1–2 m) compared to Landsat and Sentinel-2 images and were accessed using a web map service (WMS).

Highest-quality Sentinel-2 level 2 bottom-of-atmosphere images from the summers 2019, 2021 and 2022 were used in the interpretation (Table S3). They were cloud free images with good light conditions. We selected images that were taken in three different sea level conditions: low tide, mean sea level and high tide. This helped us determine the position of the mean shoreline as accurately as possible also in flat intertidal areas. For Tanafjorden, we also used an image acquired during field

work on October 20th, 2021 (Table S3) to help link the *in situ* observations with the bird's eye view. The October image was taken at low tide during spring tide conditions, showed the almost entire intertidal zone exposed and corresponded well with the *in situ* observations (made between 7:15–13:30 UTC). The images were retrieved from the Copernicus Open Access Hub as local files and visually interpreted in QGIS.

Table S3. Details of Sentinel-2 images from 2019–2022 utilized as reference for creating validation data. Sea levels (Norwegian Mapping Authority Kartverket) at the time of the image acquisition are reported relative to the mean sea level 1996–2014. Tanafjorden sea levels are calculated for Gávesluokta, Tana from Honningsvåg tide gauge observations (Figure S1). Ny-Ålesund levels are from the Ny-Ålesund tide gauge. The coasts of the study areas were cloud-free in the selected images.

	Date	Time (UTC)	Cloud cover	Sea level (cm)
Tanafjorden				
S2A L2A 20210807T103031 N0301 R108 T35WNU	7.8.2021	10:30	2 %	-75
S2A L2A 20190723T101031 N0213 R022 T35WNU	23.7.2019	10:10	6 %	-3
S2B L2A 20210703T102559 N0301 R108 T35WNU	3.7.2021	10:25	5 %	+57
Tanafjorden: Coincidental with *in situ* observations				
S2A L2A 20211020T101041 N0301 R022 T35WNU	20.10.2021	10:10	59 %	-125
Ny-Ålesund				
S2A L2A 20220821T132719 N0400 R024 T33XVH	21.8.2022	13:27	9 %	-34
S2A L2A 20220823T131731 N0400 R124 T33XVH	23.8.2022	13:17	7 %	-8
S2A L2A 20220715T124711 N0400 R138 T33XVH	15.7.2022	12:47	41 %	+45

In Tanafjorden, in situ observations of coastal landforms, land cover and shoreline were made on October 19th, 20th and 22nd 2021. The whole area had received a thin snow cover a few days earlier. The field work was done in spring tide conditions, with a semidiurnal tidal fluctuation of c. 260 cm (Norwegian Mapping Authority Kartverket). Observations were made approximately during low tide, and thus the instantaneous shoreline was close to the seaward limit of the intertidal zone.

Timestamped positions of instantaneous shoreline, high water mark, and other distinct coastal landforms were recorded in the field with a handheld Garmin GPS receiver (horizontal accuracy within 5 m). In total, 59 points were marked (Figure S5), including 30 high water mark positions and 16 instantaneous shoreline positions. Observations were made in accessible sites that represented different coast types. The visited shoreline types ranged from tidal sand flats and pebble beaches to bedrock shorelines, and from low-energy to medium-energy coasts depending on their position along the fjord-open sea transition (Table S4).

Geotagged and timestamped photographs and videos were taken with a handheld Garmin GPS receiver and a Samsung smartphone to document the location of the instantaneous shoreline, landforms, vegetation, soil, rock type, infrastructure, and weather conditions at the coast (Figure S5). In total, 448 geotagged photographs and videos were taken.

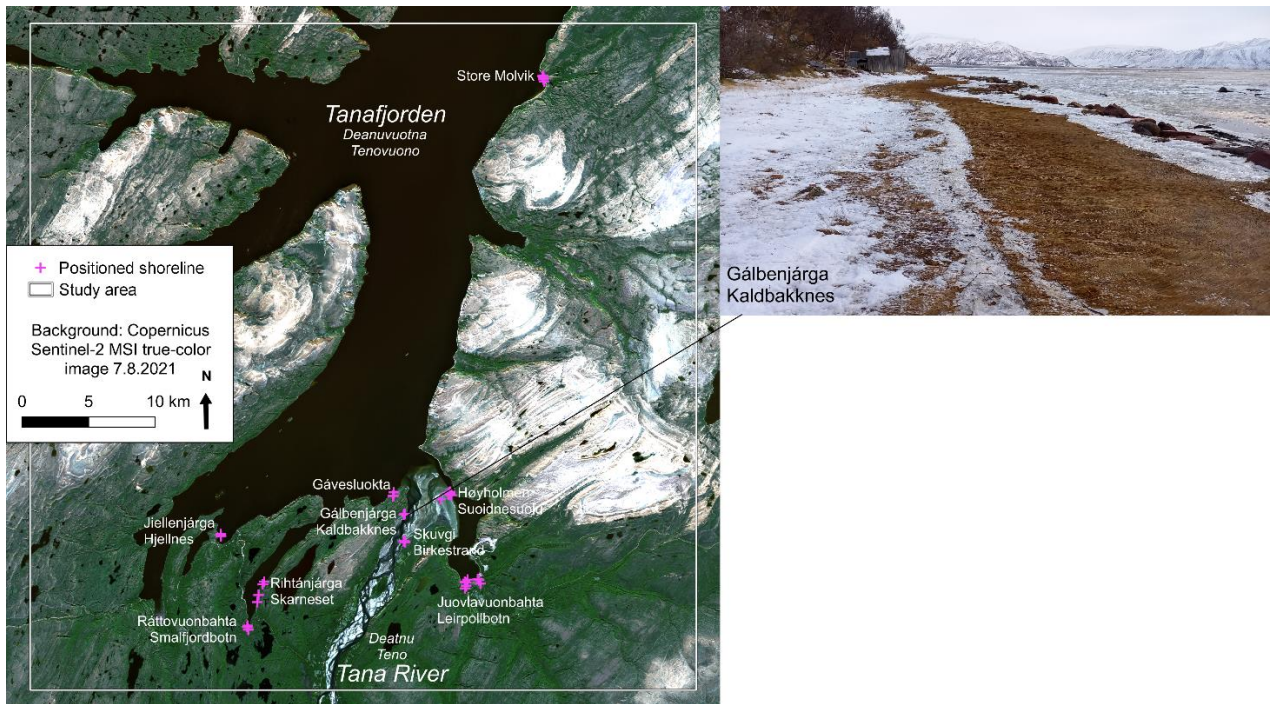



Figure S5. Left: Locations of observed instantaneous shorelines and high-water marks. The observations were positioned with a handheld GPS receiver on October 19th, 20th and 22nd 2021. Right: Example of a geotagged photograph in Gálbenjárga, Tana. It shows visible evidence of the location of the shorelines at different sea levels. The instantaneous shoreline during falling tide is located behind a belt of boulders to the right. The height of the recent high tide (c. 120 cm above mean sea level) is visible as a boundary in snow and sea ice cover at the center of the photograph. The high-water mark can be seen from shoreface curvature and vegetation and runs in front of the boat shelter to the left. Photograph: Tua Nylén 22.10.2021.

345

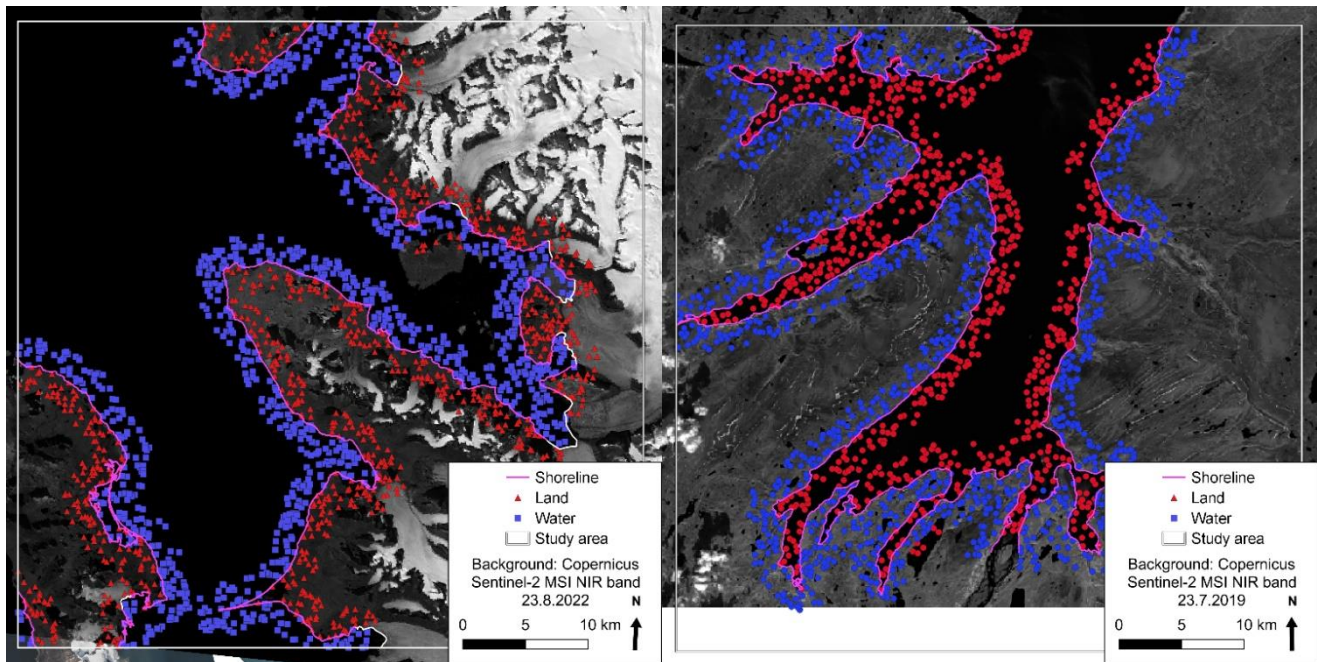
Table S4. Coast types and subtypes found in in situ observation sites. Note that the main types are non-exclusive: there is variation of coast types within named sites. See Figure S5 for a site map.

Main type	Subtype	Energy class	Site
Bedrock shoreline	Sandstone (grey)	Medium	Store Molvik
	Quartzite (pink - red - brown)	Low	Gálbenjárga - Kaldbakknes
	Mudstone (dark grey)	Medium	Gávesluokta
	Sandstone (red - brown - purple)	Low	Rihtánjárga - Skarneset
	Undifferentiated sedimentary (brown)	Medium	Jiellenjárga - Hjellnes
Pebble or cobble beach	Pebble to cobble	Medium	Store Molvik
	Pebble	Low	Juovlavuonbahta - Leirpollbotn
	Pebble	Low	Gálbenjárga - Kaldbakknes
	Cobble	Medium	Gávesluokta
	Cobble	Low	Rihtánjárga - Skarneset
	Pebble to cobble	Low	Ráttovuonbahta - Smalfjordbotn
	Cobble	Medium	Jiellenjárga - Hjellnes
Tidal sand flat and sandbar	Tidal sand flat	Low	Høyholmen - Suoidnesuolu
	Sandbar	Low	Skuvgi - Birkestrand
	Tidal sand flat	Low	Juovlavuonbahta - Leirpollbotn
	Tidal sand flat	Low	Ráttovuonbahta - Smalfjordbotn

350   The mean sea level shoreline was digitized by hand for both study areas (Figure S6). It was first digitized for Tanafjorden and for the sites that had been observed *in situ*. The shoreline was then extended outwards to cover the entire study area and then the Ny-Ålesund study area. The digitizing was performed at a fixed scale (1:5000) to minimize variation in the level of detail along the coast. The interpretation was primarily based on the hand-picked Sentinel-2 images with known dates and known sea level, but the orthophoto and geotagged photographs were used to check the interpretation and to understand the

355   landforms better. Particularly in Ny-Ålesund, there were large differences between the orthophotograph and images from 2022 due to the shifting position of the glaciers. The border between marine-terminating glaciers and sea were digitized based on the Sentinel-2 image from 23.8.2022 and compared to the median ice margin during 2019–2022. In total, 373 km of mean sea level shoreline was digitized for Tanafjorden and 264 km for Ny-Ålesund.

   To create point data for validating land cover classifications, a 2000-meter buffer was drawn around the digitized shoreline.

360   Small islands, lakes, ponds and rivers were excluded from the buffer area. Within the buffer area, 2000 random points were created. These were manually classified into water and land (in mean sea level conditions), based on visual interpretation of remotely sensed reference data and in situ observations. In Tanafjorden, we classified 1193 points as land and 807 points as land (Figure S6). In Ny-Ålesund, there were 952 land points and 1048 water points.

365 **Figure S6. Validation data for (A) Tanafjorden and (B) Ny-Ålesund.**

Validating coastal land cover and shoreline data against independent data

For this paper, the land cover classification and the fitted shorelines for the most recent time-step (2019–2022) were validated in the two study areas. Results for the earlier time-steps were not validated for this paper.

First, the classification results were validated separately for single sensors (1-3/time-step) and the merged multi-sensor
370 collection, and separately for the five initial classification methods (NDWI, MNDWI, AWEI, WIBI, random forest). Each classification was compared to the land cover validation data (2000 manually classified points per study area) with a cross-tabulation approach by calculating a confusion matrix.

We quantified the accuracy of each classification by calculating the overall accuracy, the number of correctly classified validation points divided by the total number of validation points. For the two classification methods selected for the final
375 iterations, namely NDWI and supervised random forest classification, we also examined producer's (correctly classified points / total number of points of true class X) and user's accuracy (correctly classified points / total number of points assigned to class X) for the two land cover classes. We also prepared maps of the spatial distribution of the misclassified points.

Second, the accuracy of the fitted shorelines was assessed at equally spaced cross-shore transects. This was done separately
380 for the two classification methods selected for the final iterations, namely NDWI and supervised random forest classification. We calculated the distance from the reference shoreline to the modelled shoreline and reported the summary statistics of these error distances as well as produced maps of the spatial distribution of the largest error distances. To

calculate the error distances, we used the Points Along Geometry function (QGIS Vector general) to create one point at 30-meter intervals along the reference shoreline. Similarly, we created points at 5-meter intervals along the modelled shoreline. We then calculated the distance between one reference shoreline point to the nearest point along the modelled shoreline by using the Distance to nearest hub (line to hub) function (QGIS Vector analysis).

Data availability

We processed a total of 624 images of the Tanafjorden study area to create the shoreline time series (Table S5). More than 30 images were found suitable for each time-step (Table S5). The Landsat 7 image collections was largest in the Tanafjorden study area (Table S5). As the Sentinel-2 mission has a shorter revisit time compared to Landsat satellites, the number of available Sentinel-2 images was high from 2017 onwards. This resulted in a notably higher number of available images for time-steps 2014–2018 and 2019–2022 (Table S5).

When the procedure was transferred to the Ny-Ålesund study area, the data availability was rather different from that in mainland Norway (Table S5). Although a total of 656 images were processed, the image collections for three early time-steps, namely 1984–1988, 1994–1998 and 2004–2008, had very little images and were considered too small for reliable time-step summaries (Table S5). Still, there was enough data to create a shoreline time series with five time-steps and covering a time period of 30 years. Image availability in Landsat 5 and Landsat 7 collections was in general low in the Ny-Ålesund study area (Table S5). Since the overlap of ground tracks of the Sentinel-2 satellites increases towards the north pole, the availability of Sentinel-2 images was even higher for Svalbard than for Tanafjorden (Table S5).

For the most recent time-step, there were at least 26 observations and up to 332 observations on each pixel. For the first six time-steps, there were large areas with less than 5 valid observations on a pixel, that is, below the threshold of being included in the time-step summary. In Tanafjorden, most of those were mountain tops with patches of snow that melt very late in the summer. The pixel quality algorithm (Landsat CFMASK algorithm) had often falsely classified them as clouds. The same general pattern of observation counts is visible throughout the time series in Tanafjorden. It was evident that the cloud detection algorithms unnecessarily decreased data availability, while using a cloud mask was necessary for the outcome. In Ny-Ålesund, there were more random data shortages, depending on tile divisions.

Table S5. Number of analysed satellite images for A. the Tanafjorden study area and B. the Ny-Ålesund study. The considered images had been filtered to those that had overlap with the study area, had at most 50 % (30 % for Sentinel-2) cloud cover and were taken during the northern summer months (June to September). After masking clouds, cloud shadows, fill pixels and edge noise, the images were further filtered to those that had at least 100 km² overlap with the land and water classes in the training data layer (4 % of the study area). For Ny-Ålesund, the original tier 1 image collections for time-steps 1984–1988, 1994–1998 and 2004–2008, with less than 10 observations, were considered too small for reliable shoreline detection. * = total number of images excluding time-steps 1984–1988, 1994–1998 and 2004–2008.

Number of images processed

A. Tanafjorden

Time-step	Landsat 5	Landsat 7	Landsat 8	Sentinel-2	ALL
1984–1988	55	-	-	-	55
1989–1993	39	-	-	-	39
1994–1998	33	-	-	-	33
1999–2003	1	46	-	-	47
2004–2008	38	33	-	-	71
2009–2013	24	49	18	-	91
2014–2018	-	52	47	29	128
2019–2022	-	61	44	55	160
Total	190	241	109	84	**624**

B. Ny-Ålesund

Time-step	Landsat 5	Landsat 7	Landsat 8	Sentinel-2	ALL
1984–1988	6	-	-	-	(6)
1989–1993	19	-	-	-	19
1994–1998	9	-	-	-	(9)
1999–2003	0	24	-	-	24
2004–2008	6	10	-	-	(16)
2009–2013	2	24	4	-	30
2014–2018	-	0	99	41	140
2019–2022	-	0	116	296	412
Total	42	58	219	337	**625***

Classification accuracy

Validation of the results for the time-step 2019–2022 against the independent set of 2000 coastal points showed, that the procedure reached an overall classification accuracy of c. 0.99 for the Tanafjorden study area and c. 0.98 for the Ny-Ålesund study area (Table S6). Most accurate classifications were derived from Sentinel-2 images for both study areas, but also the overall accuracies of classifications based on other satellite sensors were higher than 0.85 (Table S6). The size of the image collection was not the only factor influencing classification accuracy since the accuracy of the Landsat 8 classification with

30 images in Tanafjorden was higher than that of the Landsat 7 with 35 images (Table S6). Similarly, the Sentinel-2 classification was more accurate than the classification based on all available images (Table S6).

For Tanafjorden, the overall accuracy of the random forest supervised classification was similar to the accuracies of simple water indices like NDWI (Table S6). In Ny-Ålesund, however, the overall accuracy of the random forest algorithm was notably higher than that of other algorithms for the Sentinel-2 and combined image collections (Table S6). The accuracies of water index-based classifications were close to each other with no clear winner (Table S6).

Compared to water indices, the random forest supervised classification generally improved the producer's accuracy of the land class (improved classification of true land pixels) and the user's accuracy of the water class (more pixels classified as water were truly water), sometimes at the cost of other accuracy metrics (Table S6).

Table S6. Validation results for the 2019-2022 time-step using Landsat 7, Landsat 8 and Sentinel-2 images, against 2000 manually classified land cover points. Overall accuracies (A, C) and user's and producer's accuracies (B, D) are reported for the Tanafjorden and Ny-Ålesund study sites. Results for the sensor fusion ('ALL') and random forest supervised classification ('RF') are highlighted.

A. Tanafjorden	Overall accuracy						
Sensor	RF	NDWI	MNDWI	WIBI	AWEI	Images	NAs
Landsat 7	0.954	0.984	0.984	0.984	0.971	35	0
Landsat 8	0.978	0.989	0.991	0.991	0.983	30	10
Sentinel-2	0.991	0.991	0.988	0.988	0.991	55	0
ALL	0.989	0.990	0.990	0.990	0.989	120	0

B. Tanafjorden	User's and producer's accuracies	
(ALL)	RF	NDWI
Producer's - Water	0.975	0.986
Producer's - Land	0.998	0.992
User's - Water	0.997	0.989
User's - Land	0.983	0.991

C. Ny-Ålesund	Overall accuracy						
Sensor	RF	NDWI	MNDWI	WIBI	AWEI	Images	NAs
Landsat 7	-	-	-	-	-	0	2000
Landsat 8	0.866	0.908	0.912	0.912	0.886	73	0
Sentinel-2	0.988	0.894	0.876	0.876	0.871	272	0
ALL	0.977	0.896	0.883	0.883	0.873	345	0

D. Ny-Ålesund	User's and producer's accuracies	
(ALL)	RF	NDWI
Producer's - Water	0.956	1.000
Producer's - Land	0.999	0.780
User's - Water	0.999	0.834
User's - Land	0.954	1.000

Impact of the environment of initial classification accuracy

We examined the classification accuracy and the impact of environmental characteristics on it further for the best-performing variates of the procedure. We examined these for the combined image collection 2019–2022 (including all Landsat and Sentinel-2 images) and for the two best-performing algorithms, namely random forest supervised classification and NDWI.

In the Tanafjorden study area, the misclassified points were located very close (in practice all < 30 meter) to the actual shoreline, in intertidal zones and in areas shadowed by mountains (Figure S7A-B). The random forest algorithm misclassified 19 water points as land and two land points as water (Figure S7A). These mistakes were mainly caused by mixed pixels at the shoreline, pixel size in the analyses being 30 meters. Many of the misclassified points were also located in the intertidal zone, where the shoreline is sensitive to the tidal sea level (Figure S8A-B). The NDWI algorithm misclassified ten water points as land and nine land points as water (Figure S7B). Since this algorithm was not trained, it put some land areas shadowed by mountains (low primary productivity, higher slope, lower illumination) into the water class. However, it seemed to classify intertidal zones (gravel and sand) and well-illuminated shorelines slightly more accurately than the random forest algorithm (Figure S8B).

In the Ny-Ålesund study area, points misclassified by the random forest algorithm were located further away from the shoreline than in Tanafjorden (Figure S8C-D). The algorithm misclassified one land point as water and 46 water points as land (Figure S7C). All except three of the misclassified points were located at or near the ice margin (Figure S8C). These points were thus heavily influenced by the fluctuating position of the ice margin and the coastal waters with a high suspended sediment load (Figure S8C). These factors had an impact on the position of the transient shoreline in individual images and on the reflectance of sea water, respectively. The other three misclassified points were at the shoreline and influenced by mixed pixels and sea level variation (Figure S8C). In other ways the misclassified points were similar to correctly classified points (Figure S8C). The random forest algorithm was able to classify all bedrock, deep sea and gravel points correctly (Figure S8C). As can be clearly seen from Figure S7D, NDWI (as well as other water indices) generally failed to classify the Ny-Ålesund study area into water and land. This was mainly due to the glaciers, that had a positive NDWI value. This caused 209 land points to be classified as water – as much as 59 % of the points located on ice sheets were misclassified (Figure S8D). Points misclassified with the NDWI algorithm were thus widely spread in the 2-km coastal zone (Figure S8D). Similar to Tanafjorden, the NDWI algorithm performed better than random forest in shallow areas with high sediment loads (the Tana River delta in Tanafjorden and sediment-rich ice-marginal waters in Ny-Ålesund; Figure S8C-D).

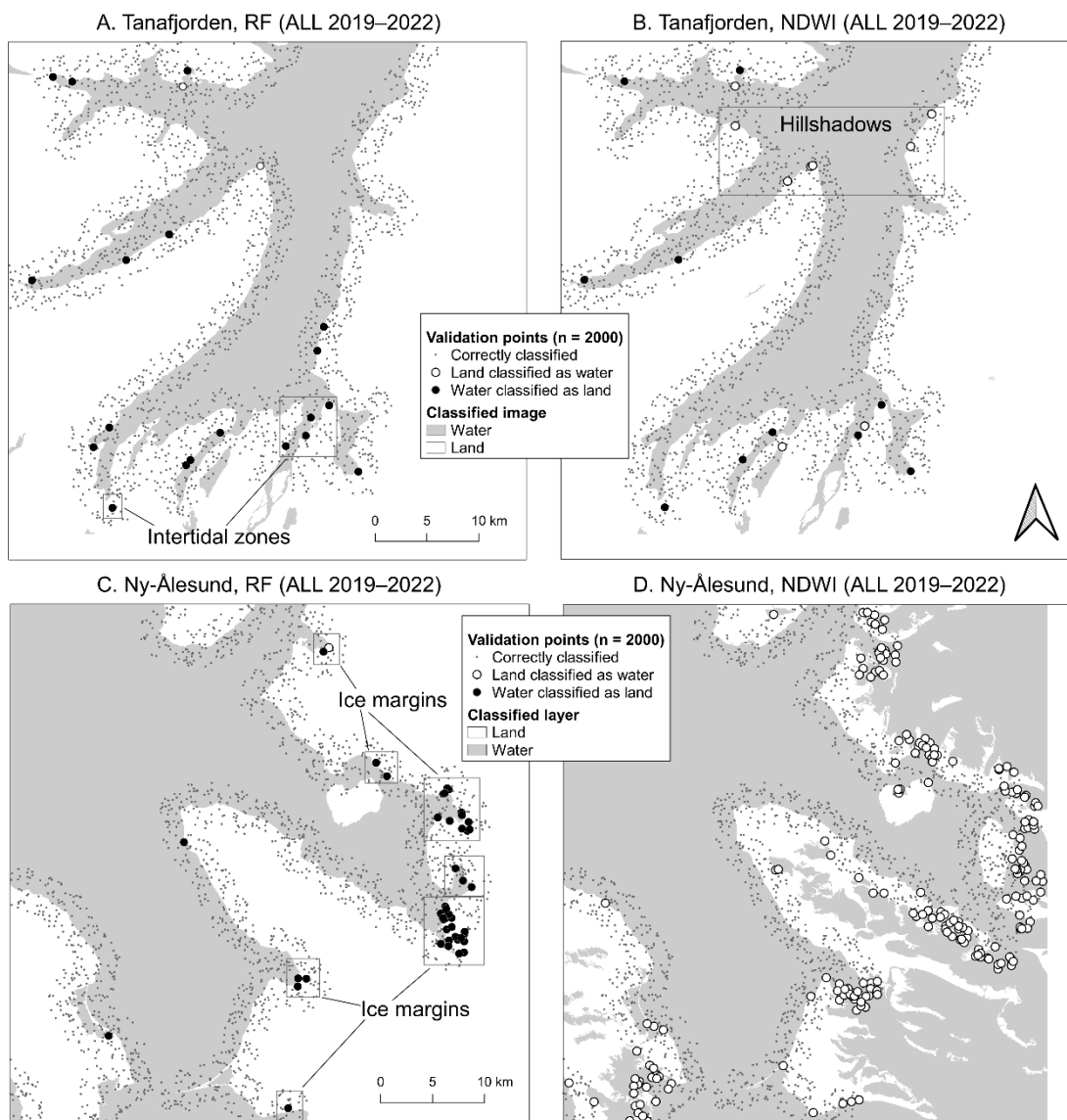



Figure S7. Location of the misclassified points in (A-B) Tanafjorden and (C-D) Ny-Ålesund study areas. The results are shown for
combined satellite image collections ('ALL') for the time-step 2019–2022, and separately for random forest supervised
classification (RF; A, C) and normalized difference water index (NDWI; B, D).

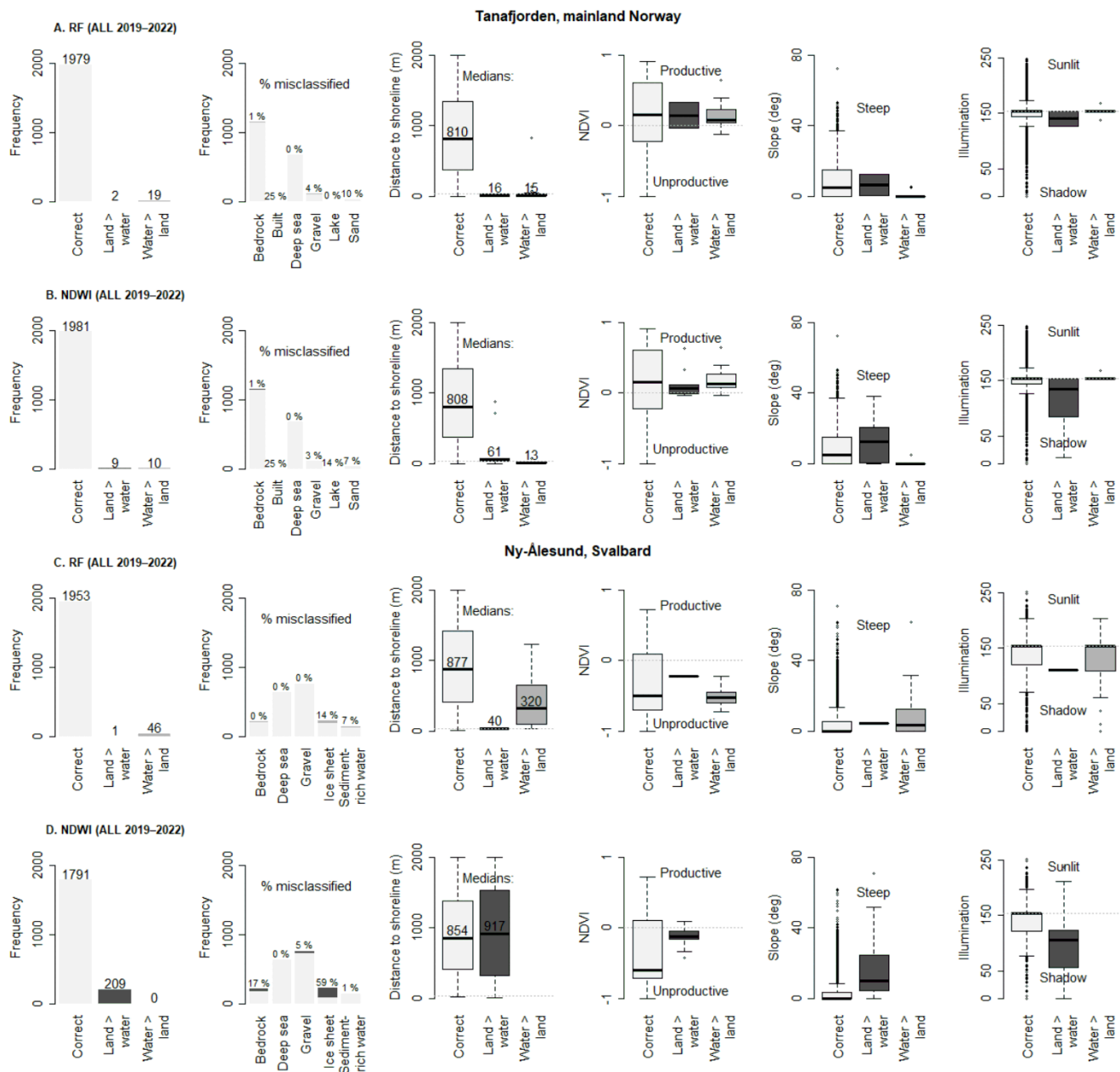



Figure S8. Environmental characteristics of correctly classified and misclassified points in the (A-B) Tanafjorden and (C-D) Ny-Ålesund study areas. Illumination is calculated from the digital elevation model and indicates sunlight availability. Illumination ranges from 0 (no light) to 255 (full illumination), while flat areas like the sea surface have a value of 153.

470

Accuracy of initial shorelines

The accuracy of the fitted shorelines was examined in both study areas and for the random forest and NDWI outputs. We examined also the accuracy of the NDWI shoreline in Ny-Ålesund, even though the result was highly misleading around glaciated areas.

475 The results indicate that the median distance (error distance) from the reference shoreline to the modelled shoreline was less than 20 meters, i.e., roughly half of the Landsat image pixel size of 30 meters (Figure S9A-D). Circa 75 % of the modelled shoreline was closer than 30 meters from the actual shoreline. The median error distances were comparable in Tanafjorden and Ny-Ålesund, while the variability in the error distance was much higher in Ny-Ålesund (Figure S9C-D). Since the error distance was calculated from the reference shoreline to the modelled shoreline – and not vice versa – it largely overestimated

480 the accuracy of the shoreline modelled with the NDWI algorithm in Ny-Ålesund.

In Tanafjorden, the longest error distances were found either in intertidal zones – most of them located in estuaries – or in mountain shadows (Figure S9E-F). The intertidal zones were more problematic with the random forest algorithm and the mountain shadows with the NDWI algorithm (Figure S9E-F). In Ny-Ålesund, the longest error distances associated with the random forest algorithm mainly occurred in ice-marginal areas, but some also in the lagoon of Richardlaguna on the island

485 of Prins Karls Forland and at a pier in the town of Ny-Ålesund (Figure S9G-H). Outside these challenging areas, however, the error distances were generally smaller than in Tanafjorden. While the NDWI shorelines were misleading in large parts of the Ny-Ålesund study area (H), their distances calculated from the actual shoreline were not any longer than the random forest shorelines. Compared to the random forest algorithm, NDWI had more problems in the point of Sarstangen between the main island and the island of Prins Karls Forland.

490

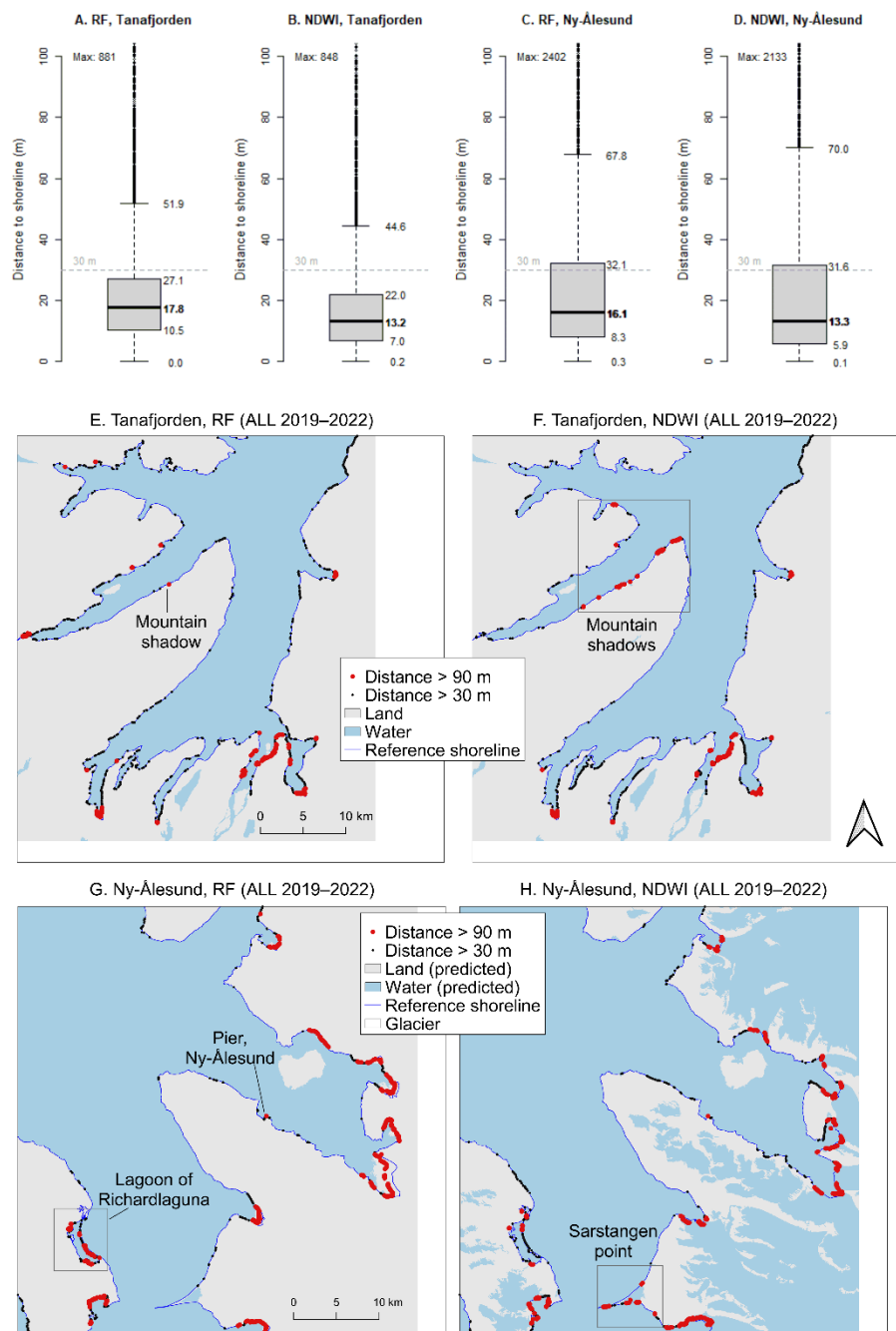



Figure S9. The boxplots (A–D) show the distribution of distances from the reference shoreline to the modelled shoreline in Tanafjorden (A, B) and Ny-Ålesund (C, D). The maps (E–H) highlight the location of long error distances. In Tanafjorden (E, F), the longest error distances are found either in intertidal zones – most of them located in estuaries – or in mountain shadows. In Ny-Ålesund, the longest error distances associated with the random forest algorithm mainly occur in ice-marginal areas, but some also in the lagoon of Richardlaguna on the island of Prins Karls Forland and at a pier in the town of Ny-Ålesund (G). Problems with the NDWI algorithm are clustered into glaciated areas but also in the Sarstangen point between the main island and the island of Prins Karls Forland (H).

495

Correcting remaining classification problems

In the last iteration round, we aimed at correcting the most influential errors in the classification:

- Problems of separating water from ice and snow using spectral indices
- Problems with detecting land in mountain shadows using spectral indices
- Problems with detecting water in shallow areas and coastal water with high load of suspended sediment

These corrections were pursued by combining the advantages of the NIR reflectance differences of water and ice, NDWI classes and random forest classes. Those pixels that had a high NDWI value, but also high NIR reflectance were moved from the water class to the land class. The final classes were calculated as a composite of the two initial class layers. The NDWI class was selected for all sunlit pixels and the random forest class was selected for all mountain shadow pixels. These two changes further improved the classification accuracy to more than 99 % and decreased shoreline error distance to less than 15 meters (Table S7).

Table S7. Comparison of classification accuracy and shoreline error distances between the original and the final iteration. The final iteration used all available image collections and a combination of NDWI and random forest classes, based on pixel illumination.

	Original NDWI		Original RF		Composited class	
	Tanafjorden	Ny-Ålesund	Tanafjorden	Ny-Ålesund	Tanafjorden	Ny-Ålesund
Classification accuracy						
Overall accuracy	0.990	0.896	0.989	0.977	0.993	0.990
User's accuracy Land	0.991	1.000	0.983	0.954	0.997	0.982
User's accuracy Water	0.989	0.834	0.997	0.999	0.988	0.996
Producer's accuracy Land	0.992	0.780	0.998	0.999	0.992	0.996
Producer's accuracy Water	0.989	1.000	0.975	0.956	0.996	0.984
Error distances (m)						
Median error distance	17.8	13.2	16.1	13.3	13.6	9.1
95 % quantile	149.0	90.5	743.9	722.9	78.9	156.8

Further examples of geomorphological applications

Seven summary variables were calculated from the final land cover time series. These were the probability to belong to the water class, long-term trend, change intensity, first and last time-steps in the water class and first and last time-steps in the land class (Figures S10-S13).

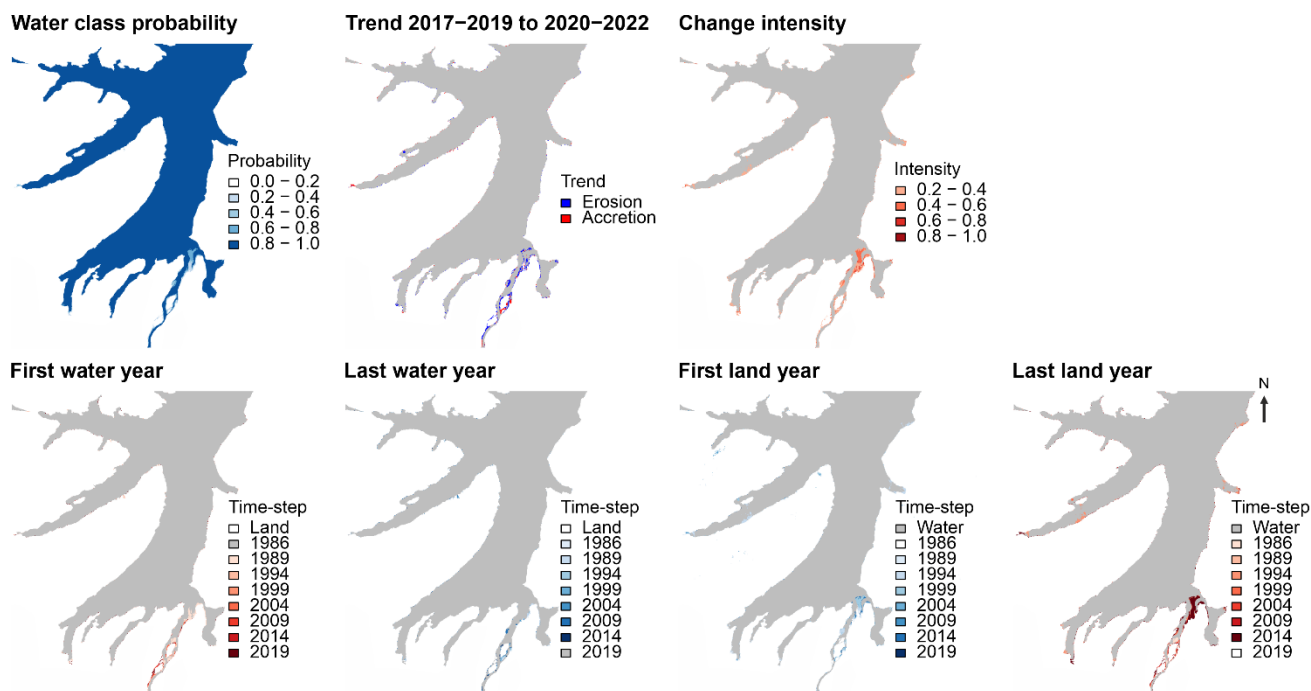
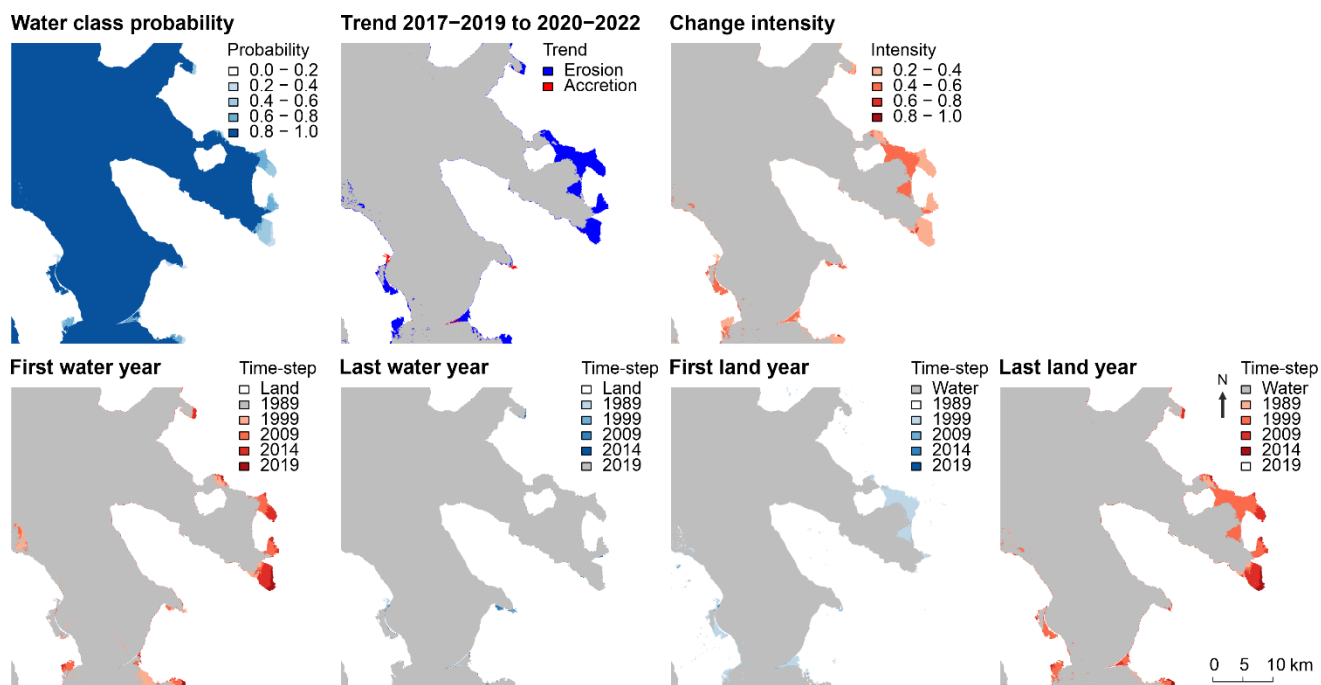



Figure S10. Examples of time series summary variables for geomorphological applications in the Tanafjorden study area.

520   **Figure S11. Examples of time series summary variables for geomorphological applications in the Ny-Ålesund study area.**

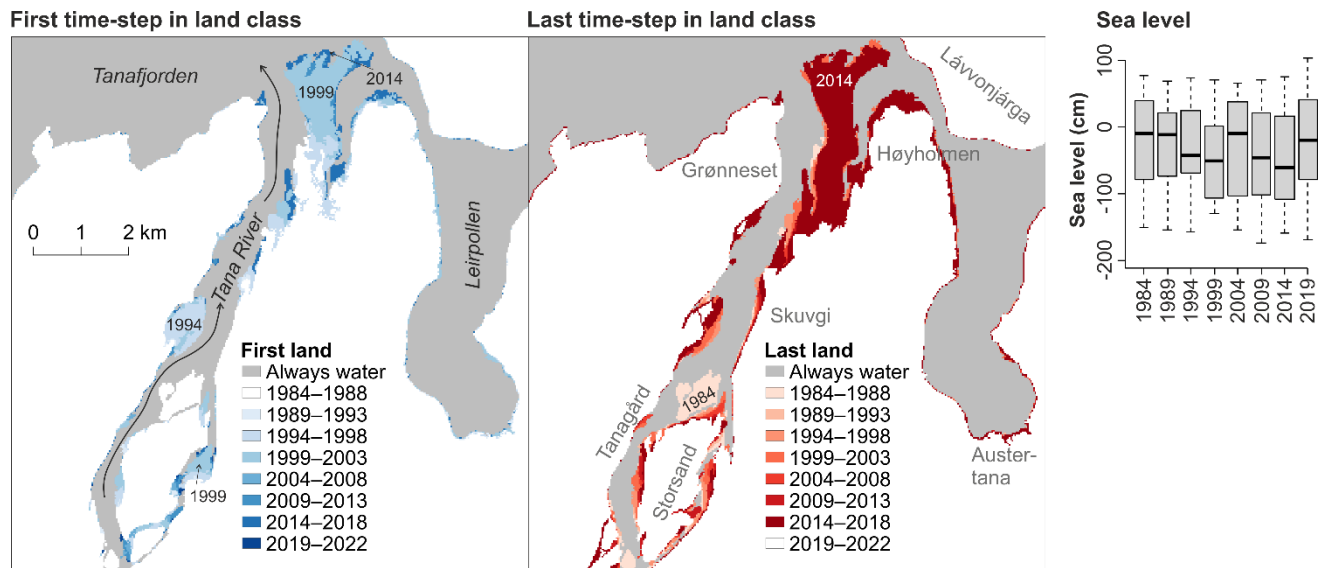



Figure S12. A local example of shoreline displacement in the Tana River delta, which is the hotspot of coastal change in the study area. Not all changes are attributable to erosion and accumulation, however; inter-annual differences in sea level have a notable impact on the shoreline position in the tidal flat of Høyholmen. The sand bar was largely exposed in 1999–2003 and 2009–2018, when the median sea level was relatively low but submerged during other time-steps.

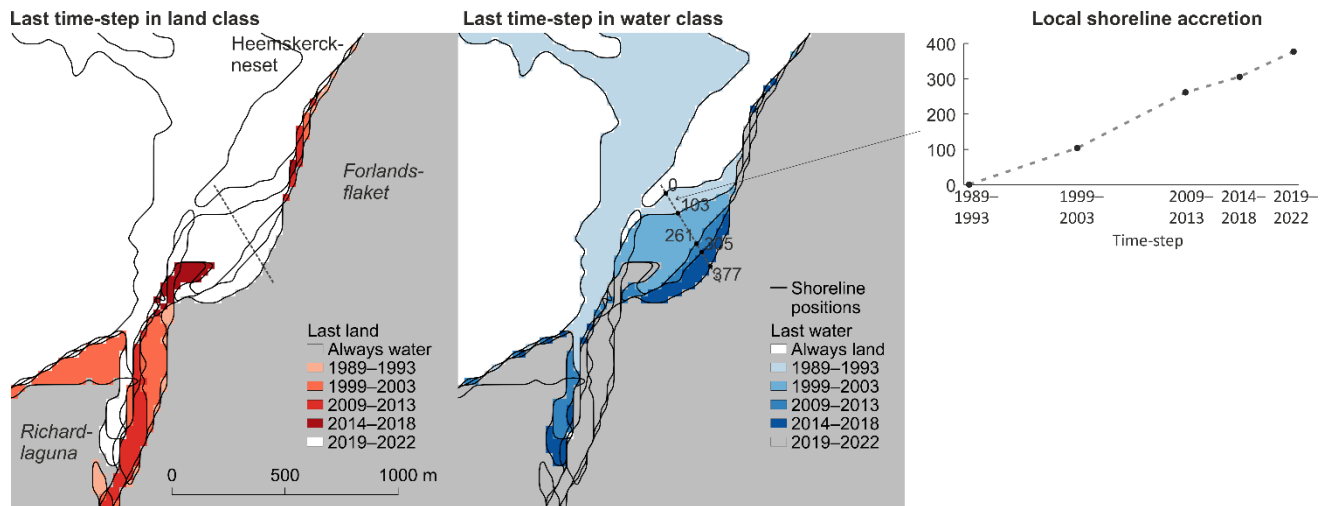



Figure S13. A local example of coastal erosion and accretion and related spit development at the north-east coast of Prins Karls Forland. Accretion rates can be extracted from the shoreline time series to examine.

Remote sensing method literature for coastal land cover

Bartsch, A., Ley, S., Nitze, I., Pointner, G., and Vieira, G.: Feasibility Study for the Application of Synthetic Aperture Radar for Coastal Erosion Rate Quantification Across the Arctic, Front. Environ. Sci., 8, 143, https://doi.org/10.3389/fenvs.2020.00143, 2020.

Bartsch, A., Pointner, G., Ingeman-Nielsen, T., and Lu, W.: Towards Circumpolar Mapping of Arctic Settlements and Infrastructure Based on Sentinel-1 and Sentinel-2, Remote Sens., 12, 2368, https://doi.org/10.3390/rs12152368, 2020.

535   Bishop-Taylor, R., Nanson, R., Sagar, S., and Lymburner, L.: Mapping Australia's dynamic coastline at mean sea level using three decades of Landsat imagery, Remote Sens. Environ., 267, 112734, https://doi.org/10.1016/j.rse.2021.112734, 2021.

Boothroyd, R. J., Williams, R. D., Hoey, T. B., Barrett, B., and Prasojo, O. A.: Applications of Google Earth Engine in fluvial geomorphology for detecting river channel change, Wiley Interdiscip. Rev. Water, 8, https://doi.org/10.1002/wat2.1496, 2021.

540   Ding, Y., Yang, X., Jin, H., Wang, Z., Liu, Y., Liu, B., Zhang, J., Liu, X., Gao, K., and Meng, D.: Monitoring Coastline Changes of the Malay Islands Based on Google Earth Engine and Dense Time series Remote Sensing Images, Remote Sens., 13, 3842, https://doi.org/10.3390/rs13193842, 2021.

Guo, Q., Pu, R., Li, J., and Cheng, J.: A weighted normalized difference water index for water extraction using Landsat imagery, Int. J. Remote Sens., 38, 5430–5445, https://doi.org/10.1080/01431161.2017.1341667, 2017.

545   Guo, W. Y., Wang, X. F., and Xia, X. Z.: Two-dimensional Otsu's thresholding segmentation method based on grid box filter, Optik, 125, 5234–5240, https://doi.org/10.1016/j.ijleo.2014.05.003, 2014.

Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., and Hobart, G. W.: Disturbance-Informed Annual Land Cover Classification Maps of Canada's Forested Ecosystems for a 29-Year Landsat Time Series, Can. J. Remote Sens., 44, 67–87, https://doi.org/10.1080/07038992.2018.1437719, 2018.

550   Hermosilla, T., Wulder, M. A., White, J. C., and Coops, N. C.: Land cover classification in an era of big and open data: Optimizing localized implementation and training data selection to improve mapping outcomes, Remote Sens. Environ., 268, 112780, https://doi.org/10.1016/j.rse.2021.112780, 2022.

Irrgang, A. M., Bendixen, M., Farquharson, L. M., Baranskaya, A. V., Erikson, L. H., Gibbs, A. E., Ogorodov, S. A., Overduin, P. P., Lantuit, H., Grigoriev, M. N., and Jones, B. M.: Drivers, dynamics and impacts of changing Arctic coasts,
555   Nat. Rev. Earth Environ., 3, 39–54, https://doi.org/10.1038/s43017-021-00232-1, 2022.

Ismail, M. A., Waqas, M., Ali, A., Muzzamil, M. M., Abid, U., and Zia, T.: Enhanced index for water body delineation and area calculation using Google Earth Engine: a case study of the Manchar Lake, J. Water Clim. Change, 13, 557–573, https://doi.org/10.2166/wcc.2021.282, 2022.

Laengner, M. L., Siteur, K., and van der Wal, D.: Trends in the Seaward Extent of Saltmarshes across Europe from Long-
560   Term Satellite Data, Remote Sens., 11, 1653, https://doi.org/10.3390/rs11141653, 2019.

Li, J., Knapp, D. E., Lyons, M., Roelfsema, C., Phinn, S., Schill, S. R., and Asner, G. P.: Automated global shallow water bathymetry mapping using Google Earth Engine, Remote Sens., 13, 1469, https://doi.org/10.3390/rs13081469, 2021.

Li, J., Peng, B., Wei, Y., and Ye, H.: Accurate extraction of surface water in complex environment based on Google Earth Engine and Sentinel-2, PLoS ONE, 16, e0253209, https://doi.org/10.1371/journal.pone.0253209, 2021.

565   Liang, J. and Liu, D.: A local thresholding approach to flood water delineation using Sentinel-1 SAR imagery, ISPRS J. Photogramm., 159, 53–62, https://doi.org/10.1016/j.isprsjprs.2019.10.017, 2020.

Liu, X., Fu, D., Zevenbergen, C., Busker, T., and Yu, M.: Assessing Sponge Cities Performance at City Scale Using Remotely Sensed LULC Changes: Case Study Nanjing, Remote Sens., 13, 580, https://doi.org/10.3390/rs13040580, 2021.

Murray, N. J., Bunting, P., Canto, R. F., Hilarides, L., Kennedy, E. V., Lucas, R. M., Lyons, M. B., Navarro, A., Roelfsema, C. M., Rosenqvist, A., Spalding, M. D., Toor, M., and Worthington, T. A.: coastTrain: A Global Reference Library for Coastal Ecosystems, Remote Sens., 14, 5766, https://doi.org/10.3390/rs14225766, 2022.

Muslim, A. M., Foody, G. M., and Atkinson, P. M.: Localized soft classification for super-resolution mapping of the shoreline, Int. J. Remote Sens., 27, 2271–2285, https://doi.org/10.1080/01431160500396741, 2006.

Nitze, I. and Grosse, G.: Detection of landscape dynamics in the Arctic Lena Delta with temporally dense Landsat time series stacks, Remote Sens. Environ., 181, 27–41, https://doi.org/10.1016/j.rse.2016.03.038, 2016.

Nitze, I., Grosse, G., Jones, B., Arp, C., Ulrich, M., Fedorov, A., and Veremeeva, A.: Landsat-Based Trend Analysis of Lake Dynamics across Northern Permafrost Regions, Remote Sens., 9, 640, https://doi.org/10.3390/rs9070640, 2017.

Nitze, I., Grosse, G., Jones, B. M., Romanovsky, V. E., and Boike, J.: Remote sensing quantifies widespread abundance of permafrost region disturbances across the Arctic and Subarctic, Nat. Commun., 9, 5423, https://doi.org/10.1038/s41467-018-07663-3, 2018.

Olthof, I., Fraser, R. H., and Schmitt, C.: Landsat-based mapping of thermokarst lake dynamics on the Tuktoyaktuk Coastal Plain, Northwest Territories, Canada since 1985, Remote Sens. Environ., 168, 194–204, https://doi.org/10.1016/j.rse.2015.07.001, 2015.

Otsu, N.: A Threshold Selection Method from Gray-Level Histograms, IEEE Trans. Syst. Man Cybern., 9, 62–66, https://doi.org/10.1109/TSMC.1979.4310076, 1979.

Pan, F., Xi, X., and Wang, C.: A Comparative Study of Water Indices and Image Classification Algorithms for Mapping Inland Surface Water Bodies Using Landsat Imagery, Remote Sens., 12, 1611, https://doi.org/10.3390/rs12101611, 2020.

Pekel, J.-F., Cottam, A., Gorelick, N., and Belward, A. S.: High-resolution mapping of global surface water and its long-term changes, Nature, 540, 418–422, https://doi.org/10.1038/nature20584, 2016.

Pérez-Cutillas, P., Pérez-Navarro, A., Conesa-García, C., Zema, D. A., and Amado-Álvarez, J. P.: What is going on within google earth engine? A systematic review and meta-analysis, Remote Sensing Applications: Society and Environment, 29, 100907, https://doi.org/10.1016/j.rsase.2022.100907, 2023.

Pickens, A. H., Hansen, M. C., Hancher, M., Stehman, S. V., Tyukavina, A., Potapov, P., Marroquin, B., and Sherani, Z.: Mapping and sampling to characterize global inland water dynamics from 1999 to 2018 with full Landsat time series, Remote Sens. Environ., 243, 111792, https://doi.org/10.1016/j.rse.2020.111792, 2020.

Shi, T. and Xu, H.: Derivation of Tasseled Cap Transformation Coefficients for Sentinel-2 MSI At-Sensor Reflectance Data, IEEE J. Sel. Top. Appl., 12, 4038–4048, https://doi.org/10.1109/JSTARS.2019.2938388, 2019.

Vos, K., Splinter, K. D., Harley, M. D., Simmons, J. A., and Turner, I. L.: CoastSat: A Google Earth Engine-enabled Python toolkit to extract shorelines from publicly available satellite imagery, Environ. Modell. Softw., 122, UNSP 104528, https://doi.org/10.1016/j.envsoft.2019.104528, 2019a.

Vos, K., Harley, M. D., Splinter, K. D., Simmons, J. A., and Turner, I. L.: Sub-annual to multi-decadal shoreline variability from publicly available satellite imagery, Coast. Eng., 150, 160–174, https://doi.org/10.1016/j.coastaleng.2019.04.004, 2019b.

605   Winkler, K., Fuchs, R., Rounsevell, M., and Herold, M.: Global land use changes are four times greater than previously estimated, Nat. Commun., 12, 2501, https://doi.org/10.1038/s41467-021-22702-2, 2021.

Zhang, G., Wu, M., Wei, J., He, Y., Niu, L., Li, H., and Xu, G.: Adaptive Threshold Model in Google Earth Engine: A Case Study of Ulva prolifera Extraction in the South Yellow Sea, China, Remote Sens., 13, 3240, https://doi.org/10.3390/rs13163240, 2021.

Zhu, Z., Wang, S., and Woodcock, C. E.: Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and

610   snow detection for Landsats 4–7, 8, and Sentinel 2 images, Remote Sens. Environ., 159, 269–277, https://doi.org/10.1016/j.rse.2014.12.014, 2015.

Zou, Z., Xiao, X., Dong, J., Qin, Y., Doughty, R. B., Menarguez, M. A., Zhang, G., and Wang, J.: Divergent trends of open-surface water body area in the contiguous United States from 1984 to 2016, Proc. Natl. Acad. Sci. U.S.A., 115, 3810–3815, https://doi.org/10.1073/pnas.1719275115, 2018.

615   Zuo, C., Chen, Q., and Sui, X.: Range Limited Bi-Histogram Equalization for image contrast enhancement, Optik, 124, 425–431, https://doi.org/10.1016/j.ijleo.2011.12.057, 2013.

References

Andreassen, L. M., Winsvold, S. H., Paul, F., and Hausberg, J. E.: Inventory of Norwegian glaciers, NVE Report, 38, 235 pp, Norwegian Water Resources and Energy Directorate, Oslo, Norway, ISBN 978-82-410-0826-9, 2012.

620   FOSSGIS e.V.: Processed coastline data derived from OpenStreetMap [data set], https://osmdata.openstreetmap.de/, 2023.

Gerland, S., Pavlova, O., Marnela, M., Divine, D., Kohler, J., Renner, A. H. H., and Skoglund, A.: Sea ice extent variability in Kongsfjorden, Svalbard during 2003-2021, based on visual observations from the mountain Zeppelinfjellet [data set], https://doi.org/10.21334/NPOLAR.2022.D6D31F5B, 2022.

Gisnås, K., Etzelmüller, B., Lussana, C., Hjort, J., Sannel, A. B. K., Isaksen, K., Westermann, S., Kuhry, P., Christiansen, H.

625   H., Frampton, A., and Åkerman, J.: Permafrost Map for Norway, Sweden and Finland: Permafrost map for Norway, Sweden and Finland, Permafrost and Periglac. Process., 28, 359–378, https://doi.org/10.1002/ppp.1922, 2017.

Humlum, O., Instanes, A., and Sollid, J. L.: Permafrost in Svalbard: a review of research history, climatic background and engineering challenges, Polar Research, 22, 191–215, https://doi.org/10.3402/polar.v22i2.6455, 2003.

Kavan, J. and Strzelecki, M. C.: Glacier decay boosts the formation of new Arctic coastal environments—Perspectives from

630   Svalbard, Land Degrad. Dev., 4695, https://doi.org/10.1002/ldr.4695, 2023.

Porter, C., Morin, P., Howat, I., Noh, M.-J., Bates, B., Peterman, K., Keesey, S., Schlenk, M., Gardiner, J., Tomko, K., Willis, M., Kelleher, C., Cloutier, M., Husby, E., Foga, S., Nakamura, H., Platson, M., Wethington, M., Williamson, C., Bauer, G., Enos, J., Arnold, G., Kramer, W., Becker, P., Doshi, A., D'Souza, C., Cummens, P., Laurier, F., and Bojesen, M.: ArcticDEM, Version 3, https://doi.org/10.7910/DVN/OHHUKH, 2018.

635 Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., and Pavelsky, T. M.: MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset, Water Resources Research, 55, 5053–5073, https://doi.org/10.1029/2019WR024873, 2019.