Product Validation and Intercomparison Report (PVIR)

glaciers cci

Prepared by:	Glaciers_cci consortium
Contract:	4000127593/19/I-NB
Name:	Glaciers_cci+_Ph2_D4.1_PVIR
Version:	0.2
Date:	15.11.2024

Contact:
Frank Paul
Department of Geography
University of Zurich
frank.paul@geo.uzh.ch

Technical Officer:
Anna Maria Trofaier
ESA Climate Office

UNIVERSITY OF OSLO

University of Zurich[UZH]

UCL Mullard Space Science Laboratory

GAMMA REMOTE SENSING

enveo

Document status sheet

Version	Date	Changes	Approval
0.1	24.07.2024	Initial draft	
0.2	15.11.2024	Input from all partners integrated	

The work described in this report was done under ESA contract 4000127593/19/I-NB. Responsibility for the contents resides with the authors who prepared it.

Author team:
Jan Wuite, Thomas Nagler, Gabriele Schwaizer (ENVEO), Lin Gilbert (UCL), Andreas Kääb, Désirée Treichler (GUIO), Frank Paul (GIUZ), Tazio Strozzi (Gamma).

Glaciers_cci Technical Officer at ESA:
Anna Maria Trofaier

Table of Contents

1. Introduction

1.1. Purpose and scope

This document is the Product Validation and Intercomparison Report (PVIR) for the ESA Glaciers CCI+ project Phase 2. The PVIR describes the validation and cross comparison activities of the parameters that are under development in the project, including: validation data, study sites and validation results. The validation and quality assessment methodology are described in detail in the Product Validation Plan (PVP; [RD1]). The validation results presented here focus on glaciers in different regions, including Svalbard, Greenland and High Mountain Asia (HMA). The geophysical parameters and corresponding data products on glaciers are: 1) glacier length changes, 2) glacier facies, 3) glacier surge inventory, 4) elevation change, 5) ice velocity.

1.2. Document structure

This document is structured into an introductory chapter followed by five chapters focussed on the validation and product intercomparison for each of the geophysical parameters:

- glacier length changes
- glacier facies
- glacier surge inventory
- elevation change
- ice velocity

Additionally, Chapter 7 provides the outcome of the geolocation intercomparison. All chapters provide a short introduction followed by an overview of the validation and product intercomparison datasets and results. The last chapter provides a list of the references.

1.3. Reference Documents

No	Doc. Id	Doc. Title	Date	Version
[RD1]	Glaciers_cci+_Ph2_D2.5_PVP	Product Validation Plan (PVP)	08.05.2023	3.0
[RD2]	Glaciers_cci+_Ph2_D4.2_PUG	Product User Guide (PUG)	22.07.2024	0.2

2. Glacier length changes (GIUZ)

2.1. Background

We have derived length changes for about 1400 glaciers >1 km^2 by digitally intersecting glacier outlines from different points in time with a centreline calculated for the maximum extent of each glacier. Extents for the 1990s, 2000s, 2010s and 2020s were taken from Xie et al. (2023), those for 1965 were manually digitized by us from an orthorectified Corona mosaic covering the northern part of the study region and using the RGI 6.0 as a base. If the glaciers are surge-type or not was assigned from the Bhambri et al. (2017) dataset. As useful (same time period) field measurements of length changes do not exist for the glaciers in the Karakoram, we could not validate the measurements directly, but only analyse possible uncertainties. These are related to the digitizing of the terminus, i.e. whether or not it is at the correct position. For debris-covered and/or glaciers in the (down-wasting) quiescent phase after a surge, this position can be very difficult to determine. Figure 2.1 shows mapping conditions for 1965 with outline overlays from Xie et al. (2023). Whereas the terminus of the surging glacier is well visible, those facing north are partly difficult to detect under the snow covert.

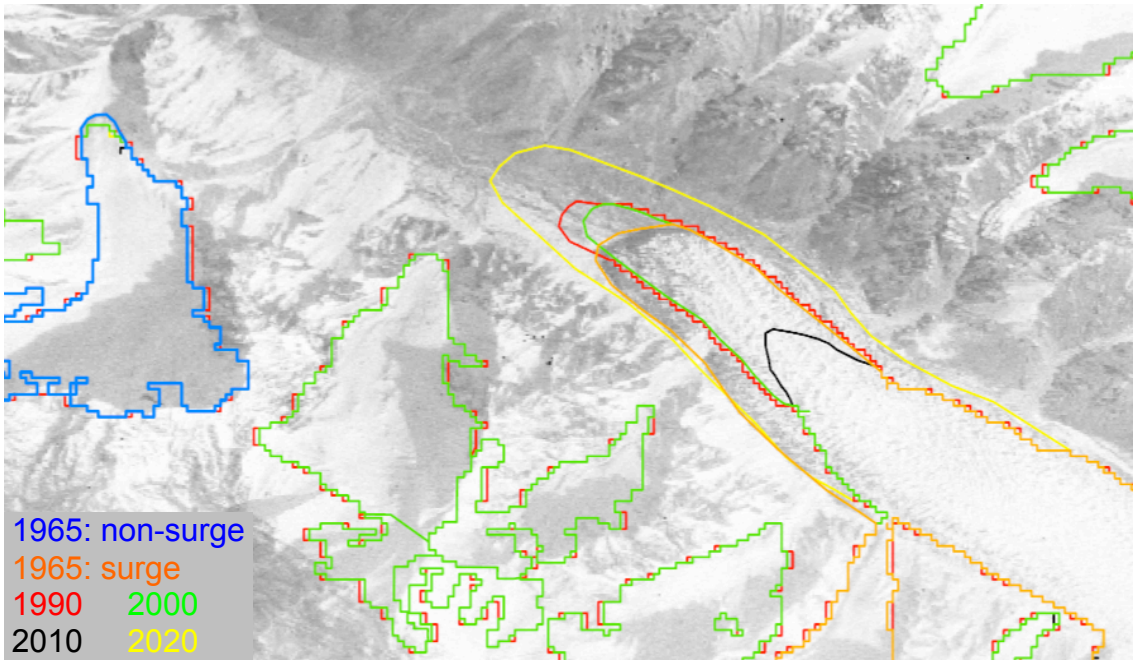

Fig. 2.1: Corona satellite image from 1965 and outlines from Xie et al. (2023) referring to four different mapping dates. Glacier termini are difficult to identify when snow covered.

While the authors of the Xie et al. (2023) study have manually corrected the terminus positions of many glaciers in all four datasets, a large number of mostly small smaller glaciers slipped through and were either not changed at all or wrongly. Figure 2.2 shows examples of well mapped advancing / surging glaciers besides smaller glaciers that seemingly change extent due to different snow conditions. Although we expected about stationary glacier fronts for most of the non surge-type glaciers, also these fluctuate by a few pixels as animations of image quicklooks over many years revealed (Paul 2015). However, we have not analysed here these minor terminus fluctuations, but focus on comparing the observed changes for selected larger glaciers. The manually digitized extents of the 1965 dataset are analysed separately.

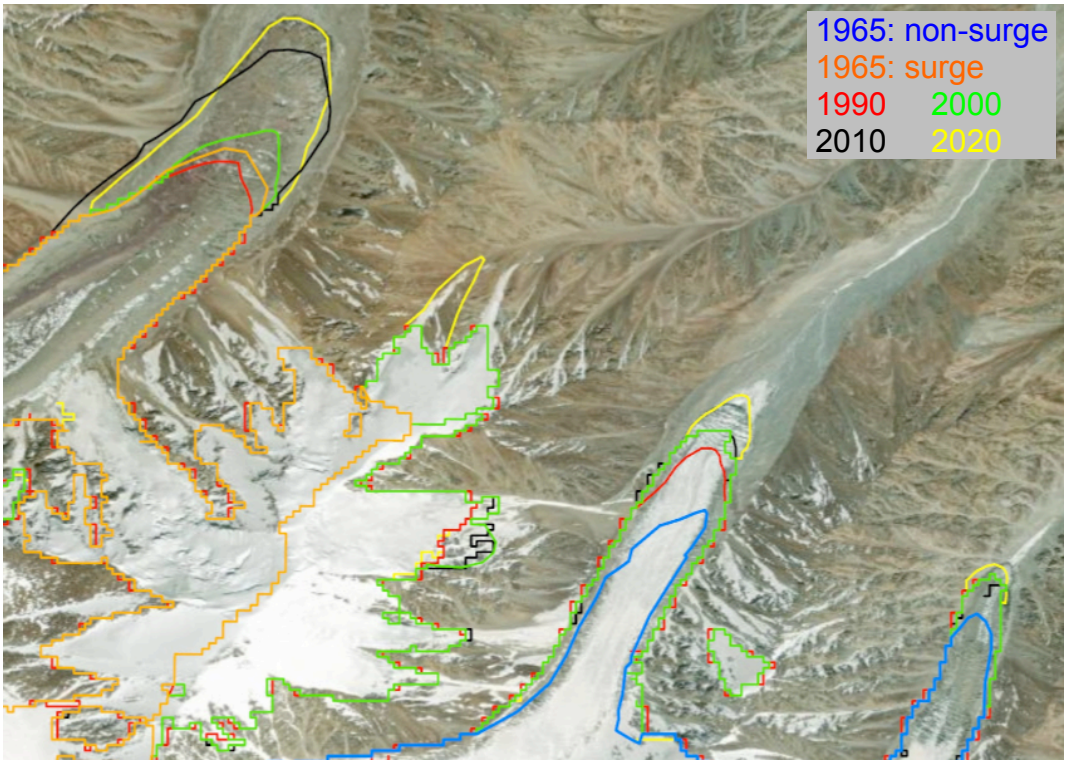

Fig. 2.2: Overlay of glacier outlines from 1965 (blue and orange) with those from Xie et al. (2023) referring to four different mapping dates (see legend). The two glaciers to the lower right have also slowly advanced but were not classified as surge-type.

2.2. Methods for validation

Glacier termini were manually digitized by one analyst for the 1965 dataset and were checked / corrected by a second analyst. The resulting changes in length were used as a measure of uncertainty. For the Xie et al. (2023) dataset we used Landsat 7 and 8 panchromatic images from 2000 and 2020 to determine terminus positions of selected and obviously wrongly mapped glaciers. Also here the differences of the mapped to the real position provides a measure of uncertainty. As the sample of glaciers is comparably small in both selections, we use the range of length differences as a measure of uncertainty rather than the standard deviation from the mean value of the respective sample.

2.3. Validation Results

2.3.1. Outlines for 1965

In Fig. 2.3 we show selected examples of glacier outlines interpreted differently by Analyst 2. The corrections are usually in the range of 100 to 400 m but can reach 2 km under challenging conditions (e.g. different interpretation of disconnected and down-wasting debris-covered glacier parts). In general, extents were overestimated by Analyst 1 so that there is a bias. Largest differences were found for down-wasting glacier tongues after a surge, debris-covered glaciers and those under snow cover, both with very little contrast. In Fig. 2.3 we show some examples of the related different interpretation.

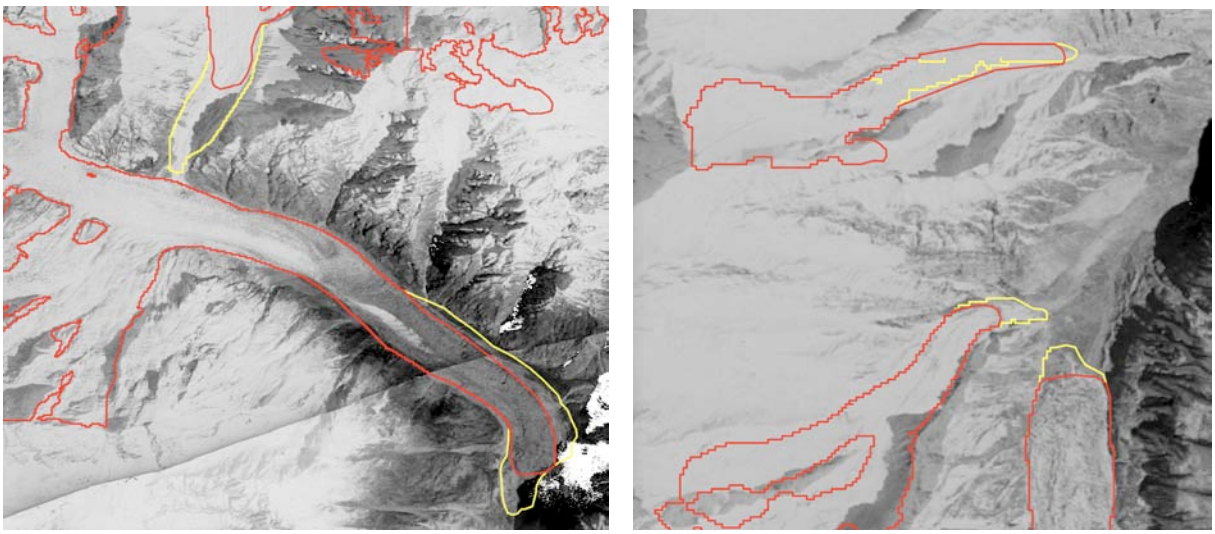

Fig. 2.3: Comparison of manually mapped glacier extents from two analysts for the 1965 Corona scenes. The difference in the extents is about 650/200 m (left) and 100 to 200 m (right).

2.3.2. Outlines by Xie et al. (2023)

Compared to satellite images from the years 2000 and 2020, the extents mapped by Xie et al. (2023) were up to 1 km longer or shorter (in both directions). In Fig. 2.4 we show an example of strong extent overestimation whereas the examples in Fig. 2.5 show both over- and underestimation. The differences are also here in part due to ice masses that were separated from the glacier several years after a surge. Although this has limited impact when the focus is on area changes, it gives wrong results when analysing length changes. For the glaciers analysed here, the differences in terminus position are too large for determination of length changes.

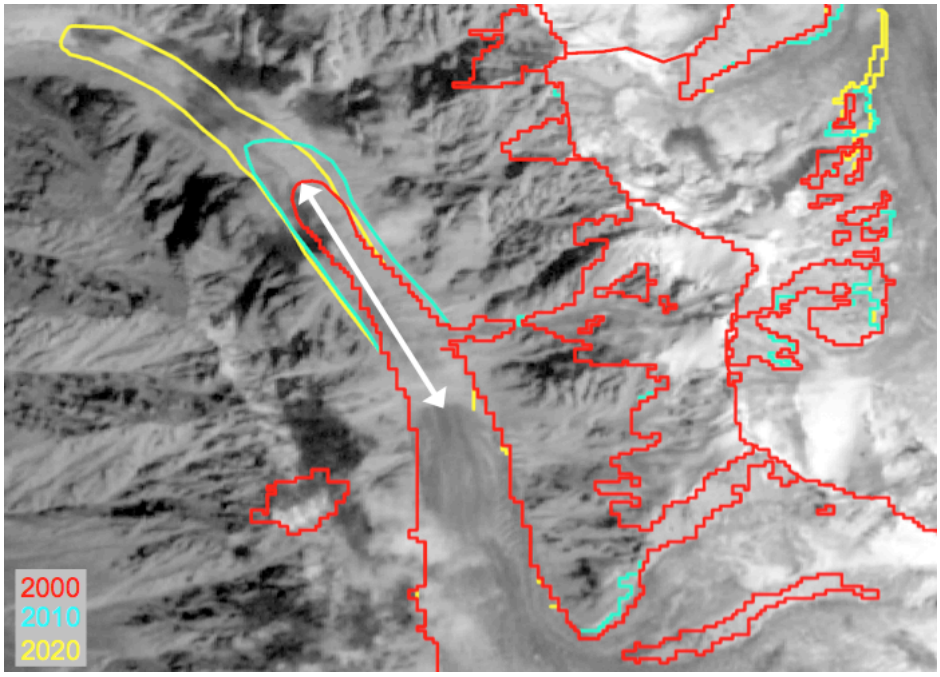

Fig. 2.4: Whereas the 2010 and 2020 extents of First Feriole Glacier is well captured by Xie et al. (2023), the wrong year 2000 extent makes the glacier 1 km longer than it was, reducing its 10-year advance from 1.2 km to 200 m.

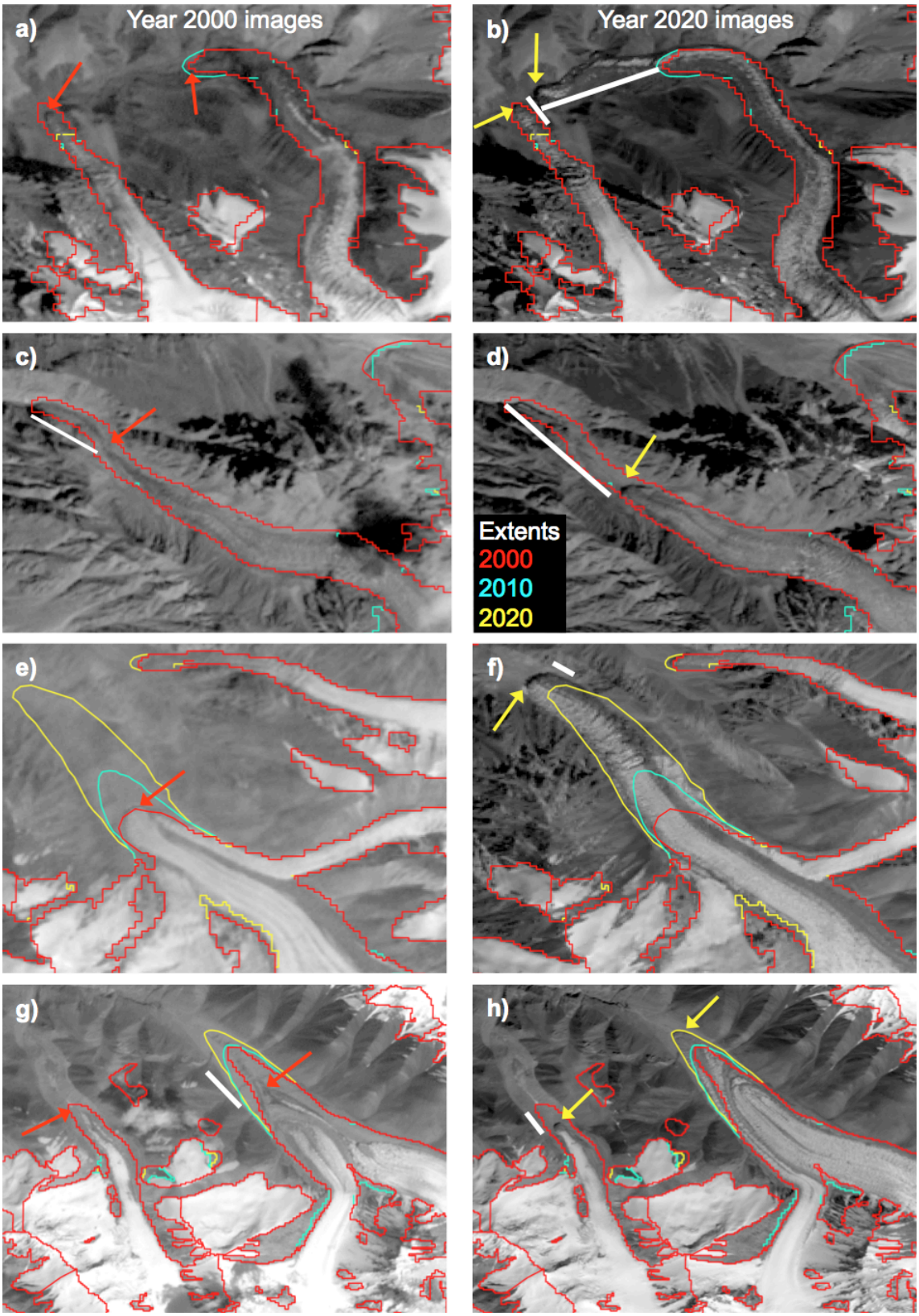

Fig. 2.5: Glacier extents by Xie et al (2023) for 2000, 2010 and 2020 compared to Landsat satellite images from 2000 (left) and 2020 (right). Arrows mark real terminus positions in the respective year, white lines the wrong part of the length change. The real position can be strongly underestimated (panel b) or overestimated (panels d and g). The red line from 2000 is shown on top, i.e. when only a red line is visible the extents from 2010 and 2020 are below.

2.4. Conclusions and Recommendations

Our analysis revealed that the small fluctuations or retreat of many glaciers could not be correctly determined with the dataset by Xie et al. (2023). The advances of surging glaciers could be correctly determined for many glaciers, but terminus positions were also wrong (up to a kilometre) in several cases and could thus not be used. As related termini were often well visible, it seems that terminus positions have not been corrected systematically or the analysts had other problems in identifying the termini. Overall, the dataset by Xie et al. (2023) is not good enough to detect the length changes we are interested in with the required certainty. Moreover, for surge-type glaciers a ten-year interval is too long to capture surge dynamics. This would require a two to three-year repeat interval, which is – mostly due to frequent cloud cover – difficult to achieve in the Karakoram. Also the datasets we derived for 1965 had large differences for individual glaciers, also a bit too large to be useful change assessment. Our main conclusions here are:

(1) glacier terminus positions for different years should only be derived by the same person applying the same rule-set
(2) a systematic approach and careful interpretation is required

The study also revealed that length changes could not be determined for several glaciers as their outlines were connected to a larger trunk glacier. There is an urgent need to create a glacier inventory for the Karakoram where the outlines of surge-type (tributary) glaciers are separated from major trunk glaciers (that might not surge). Otherwise it will not be possible to investigate their dynamics and special behaviour. Given that many glacier extents have not been updated in the 2020s inventory by Xie et al. (2023), there is also a need for a consistent and more complete update of glacier extents using the higher resolution Sentinel-2 images. Such a dataset would also be required for calibration of glacier evolution models, most of which still predict strong future glacier shrinkage in the Karakoram. The geospatial mismatch between Sentinel-2 images and earlier Landsat scenes might, however, be a problem for a related adjustment of Landsat derived terminus positions with Sentinel-2 data (see Section 7).

3. Glacier facies

3.1. Background

The glacier surface classification provides information on the three surface types snow, clean ice and debris cover and is derived from optical satellite data. For product generation, Sentinel-2 (PB4 or higher) and Landsat 5/7/8 (collection 2, Tier 1) scenes acquired in the months April to October of selected years between 1991 and 2023 are considered. The product generation is performed per relative track from Sentinel-2 scenes and per path for Landsat scenes.

Glacierized areas are defined by the glacier outlines of the Randolph Glacier Inventory v7 (RGI 7.0). The Copernicus Global Digital Elevation Model (DEM) with 30 m cell spacing (GLO30) is used as an auxiliary layer for cast shadow estimation over the entire region of interest. Other DEMs over the Karakoram region, including SRTM v4.1, High Mountain Asia DEM and the ALOS World DEM (AW3D30) v3.1 and v3.2 were reviewed, but are all affected by major data gaps or unrealistic features over glacierized regions, hampering their use as part of an automated processing chain for glacier surface classification.

To mask clouds over glacierized areas, the cloud mask attached to each Sentinel-2 and Landsat L1C scene is used and merged per acquisition date for a track or path, respectively. For Sentinel-2 scenes, the classes cloud and cirrus in the layer QI_DATA/MSK_CLASSI_B00 .jp2 with 60 m pixel spacing are used as a cloud mask. For Landsat scenes, high confidence clouds (values 22280, 24344) and cirrus (values 54596, 54852, 55052) of the layer named [sceneID]_QA_PIXEL.TIF are used as a cloud mask. Cloud shadows are only available from Landsat data, but not for Sentinel-2 data. For Landsat based glacier surface classifications, the cloud shadow pixels from the layer named [sceneID]_QA_PIXEL.TIF (value 23888) are indicated by an additional code value in the per-pixel uncertainty estimation of the glacier surface classification product.

3.2. Methods for validation

As detailed in the PVP [RD-1], the product evaluation relies on the intercomparison of products generated from satellite data of different sensors, acquired on the same date. Three data pairs from Sentinel-2 and Landsat 8 satellites with largely cloud free conditions over glacierized regions were found. In all acquisitions, the glacier surface classes snow, clean ice, and debris cover occur. The data sets selected for the intercomparison are listed in Table 3.1.

Before the intercomparison, the Sentinel-2 based glacier surface classification maps are resampled with nearest neighbour to the grid and pixel spacing of the Landsat 8 scenes. Based on the datasets on the common grid, the intercomparison area is reduced to pixels classified as glacierized areas in both products. Pixels classified as clouds over glaciers in only one or in both products are excluded from the intercomparison of the glacier surface classification. The areas used for the intercomparison on the different dates are shown in Figure 3.1.

Table 3.1: Satellite data acquired on the same dates used for product intercomparison.

Acquisition date	Sentinel-2 relative orbits	Landsat 8 path/rows	Colour in Figure 3.1
20/07/2016	R048	150/034-035	Red + Magenta
04/08/2018	R048	149/034-035	Magenta + Blue
19/09/2020	R005	148/035-036	Black

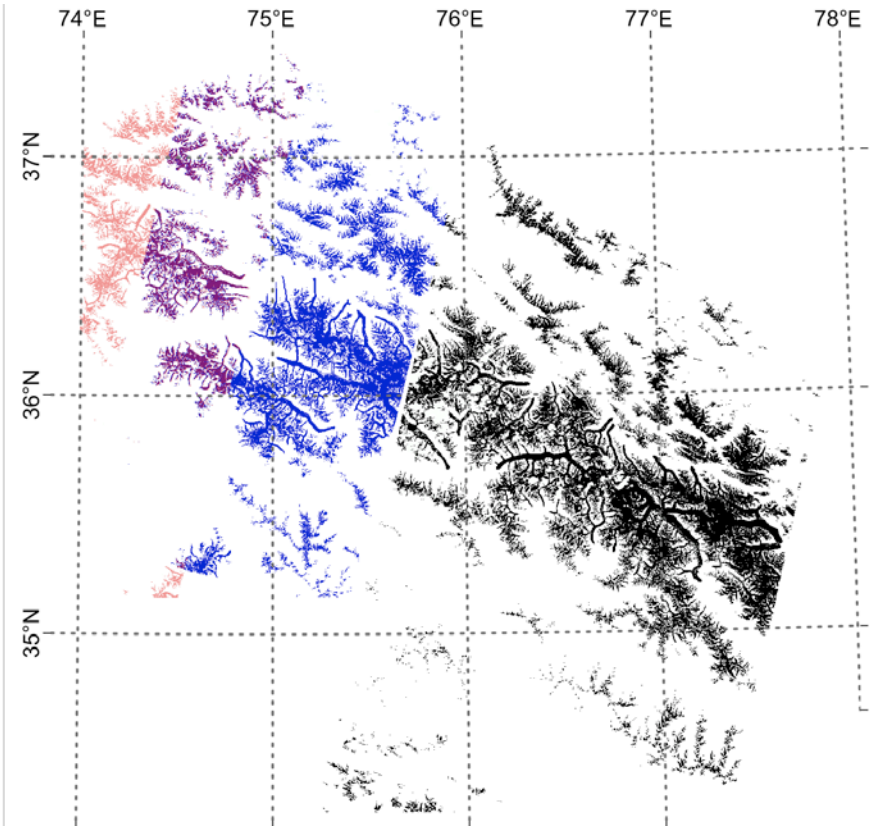

Figure 3.1: Glacier areas considered for the intercomparison of glacier surface classifications from Sentinel-2 and Landsat 8 data acquired on the same date. The different colours indicate the dates: red: 2016-07-20, magenta: 2016-07-20 and 2018-08-04, blue: 2018-08-04, black: 2020-09-19 (cf. Table 3.1). Clouds masked in either the Landsat 8 or the Sentinel-2 scene and glacier-free areas are excluded from the intercomparison (white).

For the intercomparison, the Sentinel-2 based classification is considered as the reference. The intercomparison of the surface classes is performed separately for snow, ice and debris on a pixel-by-pixel level. The intercomparison results in four classes, true positive (TP), true negative (TN), false positive (FP) and false negative (FN), which are combined in a confusion matrix per class. Based on these confusion matrices, the following statistical measures are calculated to quantify the agreement of the glacier surface classification from Landsat 8 with Sentinel-2:

- *Recall = TP / (TP + FN)*: indicates the capability of the product to classify a particular class compared to the reference;
- *Precision = TP / (TP + FP)*: provides the per cent of correctly identified class in the product compared to the reference;
- *F-Score = 2 * TP / (2 * TP + FP + FN)*: combines the recall and the precision and provides an overall measure of the product accuracy compared to the reference, even if the number of pixels for the classes is imbalanced;

- *Accuracy = (TP + TN) / (TP + TN + FP + FN)*: provides information on the per cent of all correctly classified pixels;
- *False Alarm Rate (FAR) = FP / (FP + TN)*: provides the per cent of misclassified pixels for the class compared to the reference.

To mask clouds over glacierized areas from Sentinel-2 data, the Sentinel Hub cloud detector s2cloudless (GitHub - sentinel-hub/sentinel2-cloud-detector: Sentinel Hub Cloud Detector for Sentinel-2 images in Python) was originally planned to be used. Tests were performed with the selected Sentinel-2 data sets. As there are no reference data available and the time shift between the image acquisition of Landsat 8 and Sentinel-2 result in changes in the cloud cover, the results are visually compared with false colour composites of the scenes. Related examples are illustrated.

3.3. Validation Results

3.3.1. Glacier surface classification intercomparison

For the intercomparison of the classes snow, clean ice and debris cover from Sentinel-2 and Landsat 8 data, between about 2.8M and 10.4M cloud free glacier pixels were considered. The intercomparison results show in all cases an overall accuracy of more than 90% in the classification. As an example, Figure 3.2 illustrates classification differences for a subset acquired on 19.9.2020, highlighting areas in agreement along with regions classified only by one satellite sensor as the particular class. For all surface classes, magenta indicates that the pixel was only classified from the Landsat 8 scene, while orange means that the pixel was only classified from Sentinel-2 as the particular class. A false colour composite of the same region is added for visual comparison.

On average, 93% of the snow pixels are classified by Landsat 8 compared to Sentinel-2. In 97% of the cases, the snow classification from Landsat 8 was the same as with Sentinel-2. The high average F-Score of 95% for the snow classification is also supported by an average of only 7% false classifications from Landsat 8 for snow pixels, which are classified as glacier ice or debris cover from the Sentinel-2 scene of the same dates. The detailed results for the intercomparison of snow classification from the individual data pairs and the overall average statistical values are provided in Table 3.2.

Table 3.2: Statistical measures for the intercomparison of snow-covered glacier areas classified from Sentinel-2 and Landsat-8 data of the same dates.

Date	Number of observed glacier pixels	Recall	Precision	F-Score	Accuracy	FAR
20160720	2873848	96%	98%	97%	95%	7%
20180804	6353024	88%	98%	93%	91%	3%
20200919	10422753	95%	94%	95%	93%	11%
Average		**93%**	**97%**	**95%**	**93%**	**7%**

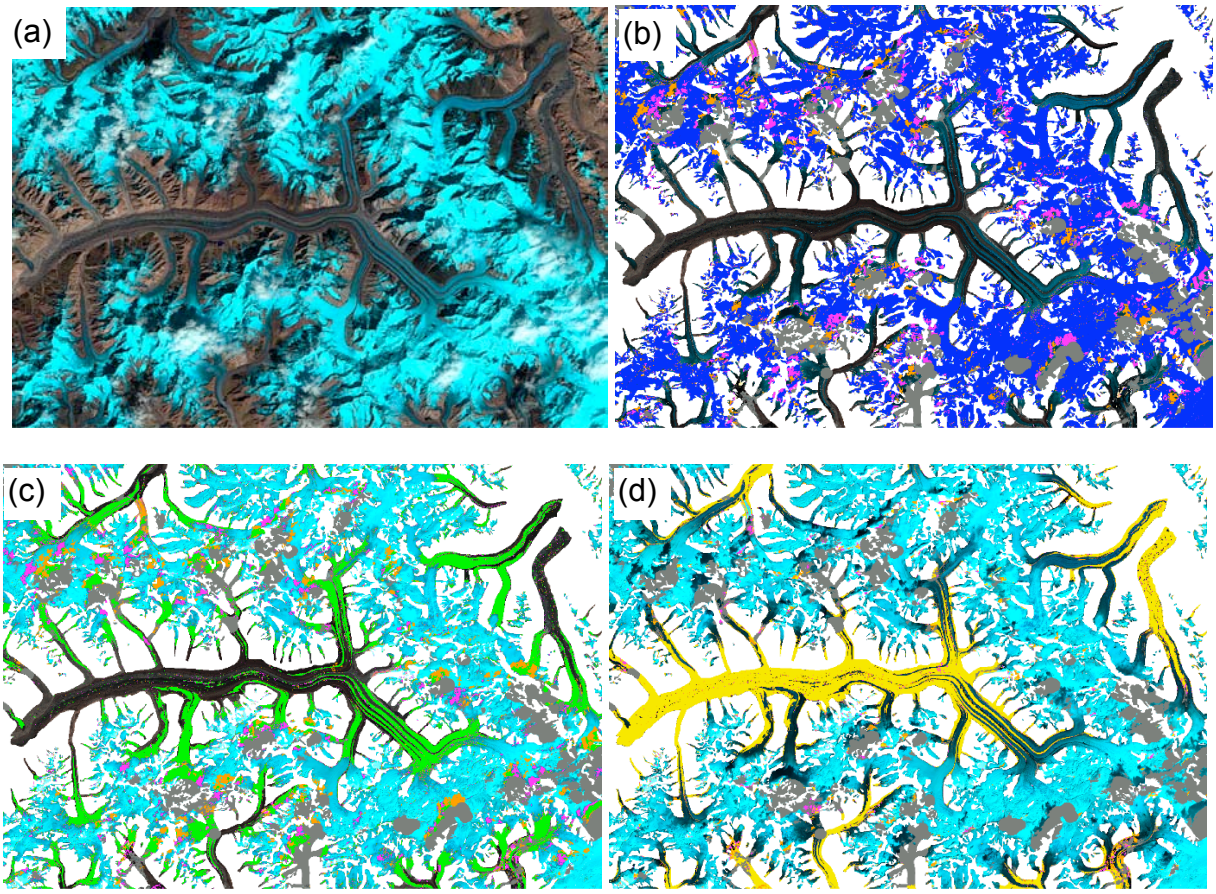

Figure 3.2: (a) False colour composite of Landsat 8 subset from 19.9.2020 (used also as a background for all other subsets). (b) Snow on glaciers classified from Landsat 8 and from Sentinel-2 of the same date (blue), only from Landsat 8 (magenta) or only from Sentinel-2 (orange). (c) As (b), but for glacier ice (green). (d) As (b), but for debris (yellow). Clouds (grey) masked in the Landsat or the Sentinel-2 scene are overlaid in the subsets (b) to (d).

In the case of cloud shadows, snow is classified as glacier ice (Figure 3.3a). As the cloud classification for Landsat data contains a flag for cloud shadows (Figure 3.3b), such areas can be marked in the associated uncertainty layer by an extra code value. For Sentinel-2, such misclassifications remain as a known issue.

For the ice classification, the products capability to correctly detect ice (recall) is on average 87%, but the average precision of 72% indicates that there are some discrepancies between the glacier ice classification from Landsat 8 and Sentinel-2 (Table 3.2). While there can be minor differences in the separation of glacier ice and debris cover, which might partly result from the resampling to a common grid, major differences in the ice classification are caused by differences in the transition zone from snow to ice, or when bright glacier ice is interpreted as snow by one sensor. Examples for such differences are shown in Figure 3.4 and Figure 3.5.

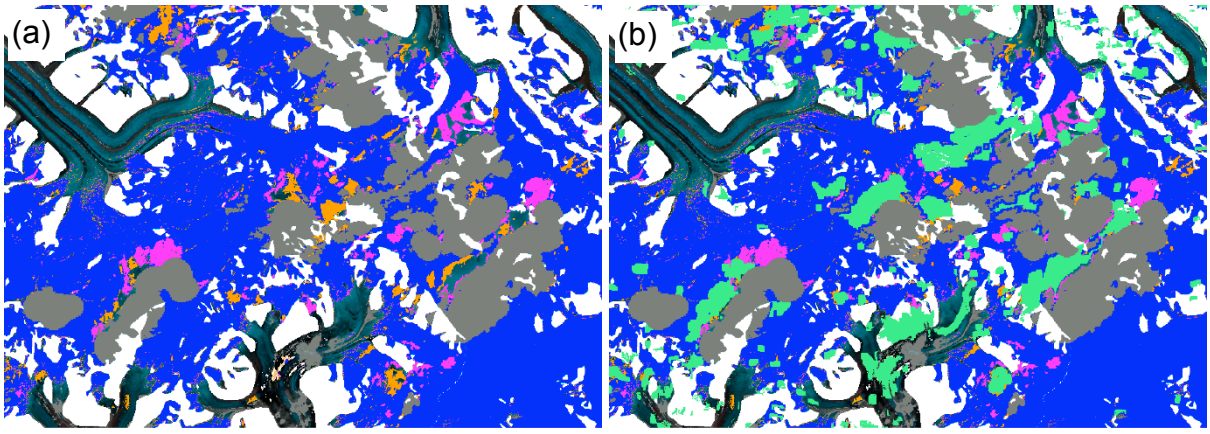

Figure 3.3: Subset of Figure 3.2b (a) without and (b) with the cloud shadow mask from the Landsat 8 scene added (shown in light green). The other colours are as in Figure 3.2.

Table 3.2: Statistical measures resulting from intercomparison of clean glacier ice as classified from Sentinel-2 and Landsat-8 data of the same dates.

Date	Number of observed glacier pixels	Recall	Precision	F-Score	Accuracy	FAR
20160720	2873848	85%	69%	76%	94%	5%
20180804	6353024	93%	65%	76%	89%	11%
20200919	10422753	83%	82%	83%	92%	6%
Average		**87%**	**72%**	**78%**	**92%**	**7%**

Figure 3.4 shows a false colour composite of a Sentinel-2 scene subset along with the related surface classification from Sentinel-2 and Landsat 8. Parts of the snow area as classified by Sentinel-2 on the major glacier tongue is not mapped by Landsat 8. However, also the visual analysis is difficult as there might be a thin snow layer on top of the ice or these areas are just bright ice. In the latter case the Landsat 8 classification would be more realistic.

Figure 3.4: Left: False colour composite of Sentinel-2 subset of 04 August 2018. Middle: Glacier surface classification from Sentinel-2. Right: Glacier surface classification from Landsat 8. Colour coding of glacier surface classification: blue = snow, cyan = ice, yellow = debris, grey = cloud.

The example in Figure 3.5 shows the combined classifications of snow, clean ice and debris-covered areas from Sentinel-2 and Landsat 8, along with a false colour composite for comparison. The areas classified as either snow or debris from Landsat, but as glacier ice from Senti-

nel-2 are marked in magenta. Also here, the major differences occur along the transition zones between snow and ice. Uncertainties in the glacier surface classifications can thus be related to (a) old, metamorphosed snow with a spectral reflectance similar to glacier ice, (b) a thin snow layer on top of the clean ice, (c) partial debris cover which is not completely obscuring the glacier surface at the scale of the pixels and thus results in a mixture of the spectral reflectance signals, or (d) bright and clean ice areas with a very similar spectral reflectance as snow areas. These are also the main errors in the glacier surface classification product, affecting mainly the discrimination of snow and ice areas on glaciers.

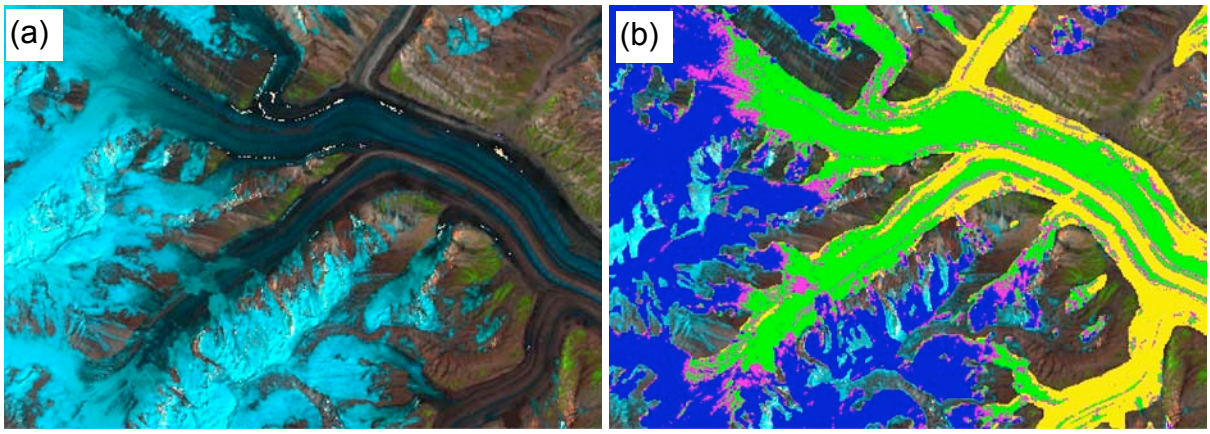

Figure 3.5: (a) RGB false colour composite of Landsat 8 subset of 4 August 2018 after atmospheric and toporaphic correction. (b) Blue, green, yellow indicating snow, glacier ice and debris cover, respectively, classified from Sentinel-2 and Landsat 8 scene of the same date. Clouds from both sensors are masked (grey). Magenta areas indicate differences in the glacier surface classification from the two satellite sensors, classified as either snow or debris from the Landsat scene, but as ice from the Sentinel-2 scene.

As shown in the examples in Figure 3.2d and Figure 3.5, the classification of debris covered glacier areas from both satellite sensors shows overall a good agreement. This is also in line with the statistical measures derived from the intercomparison of this class (Table 3.3).

Table 3.3: Statistical measures resulting from intercomparison of debris-covered glacier areas classified from Sentinel-2 and Landsat-8 data of the same dates.

Date	Number of observed glacier pixels	Recall	Precision	F-Score	Accuracy	FAR
20160720	2873848	87%	95%	90%	98%	1%
20180804	6353024	88%	95%	91%	98%	1%
20200919	10422753	85%	90%	88%	97%	1%
Average		**87%**	**93%**	**90%**	**98%**	**1%**

An average precision of 93%, meaning that this percentage is classified from both scenes, and a false alarm rate of only 1% indicate that the differentiation of debris-covered glacier areas works very well. Indeed, it has been observed that some of the debris covered cloud free glacier areas are misclassified as clouds. However, as we have not detected a systematic misclassification, this uncertainty is accepted.

3.3.2. Cloud mask checks

The s2cloudless python package was originally planned to be used to mask clouds in Sentinel-2 based glacier surface classification products. After performing the first tests and checking the results compared to false colour composites of the Sentinel-2 scene, a significant overestimation of snow-covered glacier areas as cloud cover was observed. Examples for subsets of the three Sentinel-2 scenes used for the intercomparison activities are shown in Figure 3.6, Figure 3.7 and Figure 3.8. The analysed examples cover different months and years, as well as different areas of the Karakoram Mountain range.

In each of the three images, the false colour composite of the MSI spectral bands 11 (1.6 μm), 8 (830 nm) and 4 (665 nm) for glacierized areas is shown on the left side. On the right side, the cloud mask resulting from s2cloudless (orange) and the cloud mask associated to the Landsat 8 L1C dataset in the layer QI_DATA/MSK_CLASSI_B00.jp2 (red) is overlaid.

In all cases, the s2cloudless cloud mask overestimated vast areas of snow-covered glacier areas as clouds. In contrast, the cloud mask from the L1C dataset basically identified most of the clouds over glacier areas correctly. Only the dilation of several pixels at the original 60 m pixel spacing of the cloud mask introduces an artificial overestimation. However, this is an acceptable deficit, as the main cloud-free glacier areas remain observable for the glacier surface classification.

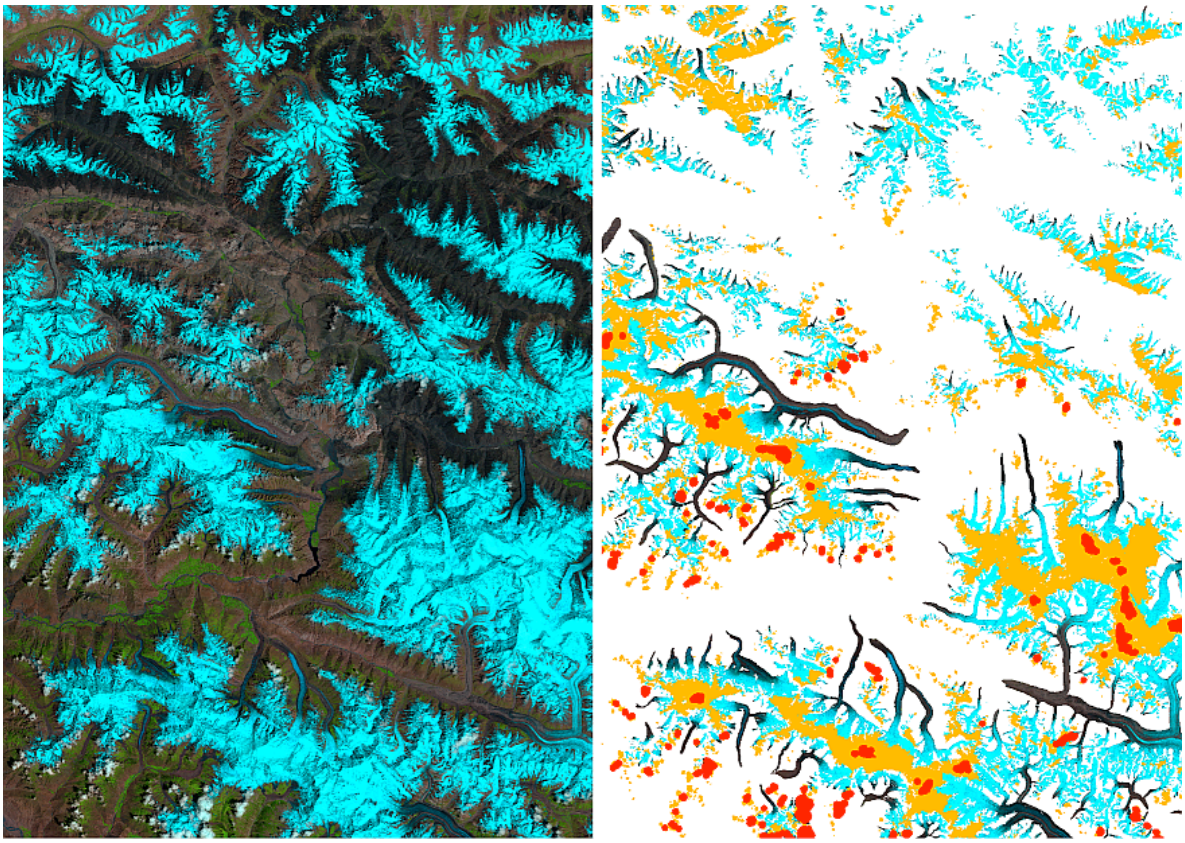

Figure 3.6: Comparison of false colour composite of the Sentinel-2 scene from 20.7.2016 (left) overlaid with cloud masks (right) from the s2cloudless (orange) and the MSK_CLASSI_B00 layer provided with the L1C product (red).

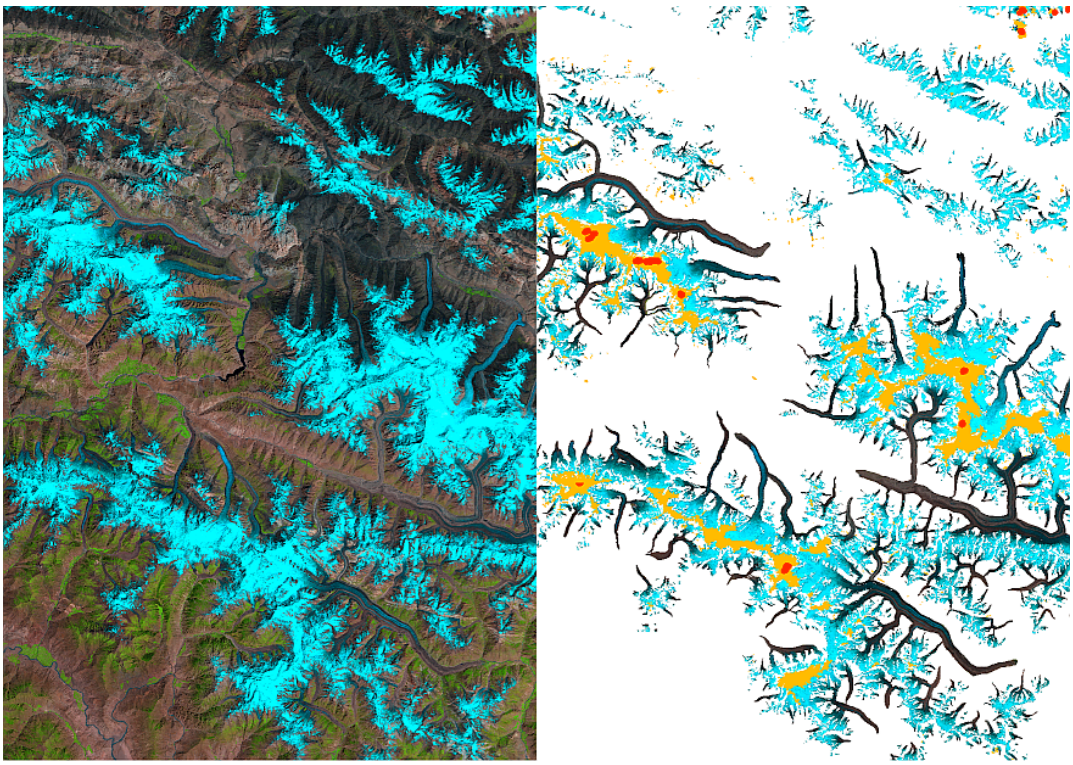

Figure 3.7: Comparison of the false colour composite of the Sentinel-2 scene from 4.8.2018 (left) overlaid with cloud masks (right) resulting from the s2cloudless (orange) and from the MSK_CLASSI_B00 layer provided with the L1C product (red).

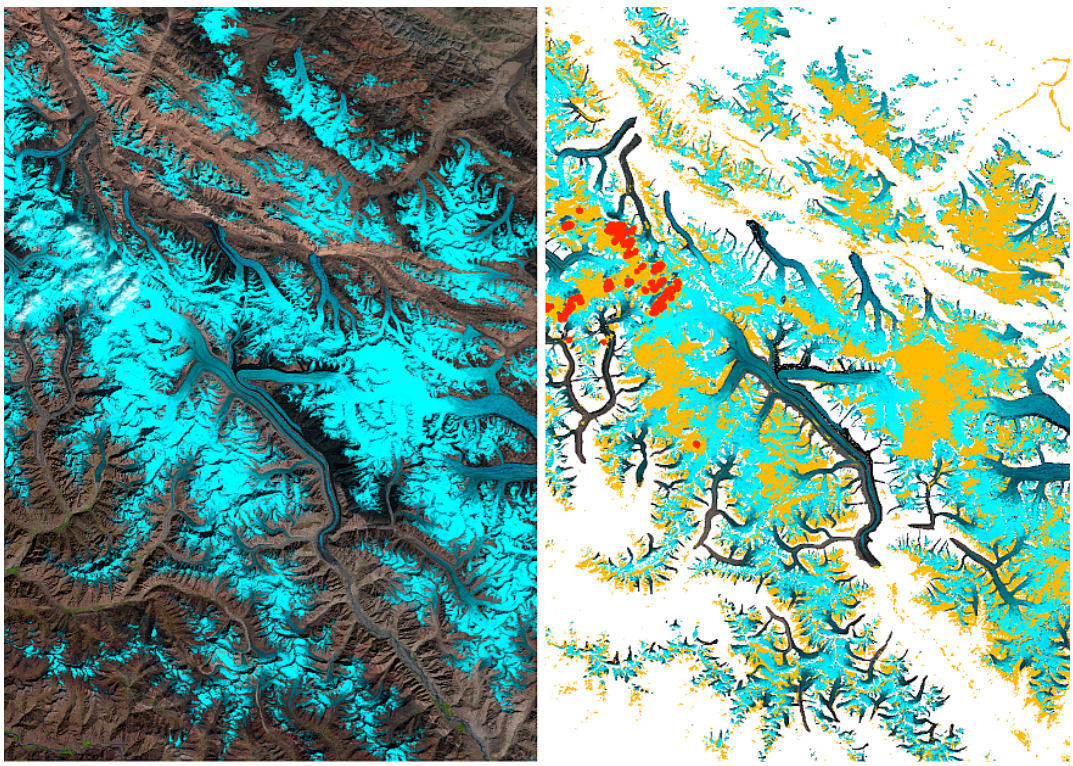

Figure 3.8: Comparison of the false colour composite of Sentinel-2 scene from 19.9.2020 (left) overlaid with cloud masks (right) resulting from the s2cloudless (orange) and from the MSK_CLASSI_B00 layer provided with the L1C product (red).

3.4. Conclusions and Recommendations

The main patterns of the glacier surface classifications from different satellite sensors acquired on the same date are very similar. The main areas of the three surface classes snow, clean ice and debris-covered ice are correctly detected from the analysed Sentinel-2 and Landsat 8 scenes. Main uncertainties are observed around the transition zone between snow and ice areas, and in some cases also between debris cover and other surface types.

As the glaciers in the Karakoram are often of surge type, the basic debris/ice layer is classified individually for each scene. But the glacier extent from the RGI 7.0 used for the classification is considered as static mask. This might introduce some inconsistencies in the debris / ice separation in the glacier surface classification product and should be kept in mind when working with the products.

Missed clouds and cloud shadows are an error source for the glacier surface classification, independent of the used satellite sensor. In the Landsat-based cloud classification, debris-covered glacier areas are sometimes masked as clouds. This is a known issue to be accepted for now. Classifications near cloudy pixels might be affected by cloud shadows, resulting in misclassified glacier surface classes. In Landsat-based glacier surface classifications, cloud shadows available from the cloud mask are available from the uncertainty layer. For Sentinel-2 based glacier surface classifications, there is no cloud shadow mask available. It is recommended to interpret surface classifications near cloud cover with caution, as such areas can be impacted by cloud shadow.

Although both, Sentinel-2 and Landsat data are prepared and available in UTM zones on the WGS84 ellipsoid, the data sets have different origins as grid references and different pixel spacings. Sentinel-2 data with 20 m pixel spacing have pixel origins ending at 00, 20, 40, 60 or 80, while Landsat data with 30 m pixel spacing have pixel origins ending at 15 or 85. This is illustrated in Figure 3.9, showing a Sentinel-2 false colour composite in its original grid in UTM43/WGS84 map projection with 20 m by 20 m pixel spacing overlaid by the Landsat 8 based snow classification in the original Landsat grid in UTM43/WGS84 map projection with 30 m by 30 m pixel spacing.

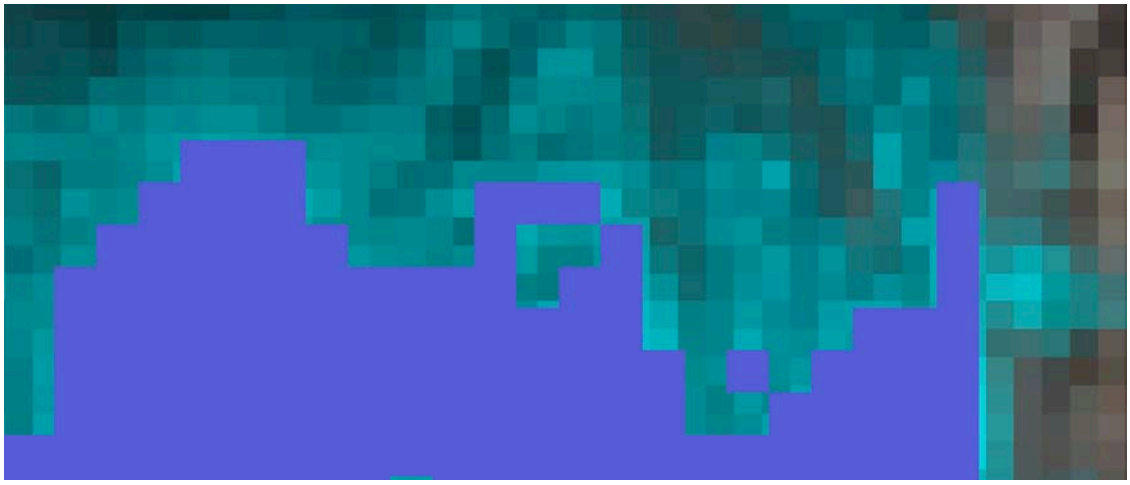

Figure 3.9: Sentinel-2 false colour composite of spectral bands 11/8/4 in the background in the original grid, overlaid with the Landsat 8 based snow classification of the same date, also in its original grid. Both data sets are in UTM43/WGS84 map projection.

Combining glacier surface classification products from both satellite systems require that at least one data set is re-gridded and resampled. As the products are classified layers, it is recommended to use nearest neighbour for any resampling approach. The user should be aware that any re-gridding and/or resampling step introduces additional uncertainties. However, compared to the other deviations the impact on product quality is probably small.

4. Glacier surge inventory

4.1. Background

The glacier surge inventory provides a global overview of glacier surges during 2017 to 2022 based on a systematic mapping of global Sentinel-1 radar backscatter data. Patterns of pronounced increases or decreases in the strongest backscatter between two winter seasons often indicate large changes in glacier crevassing, which are treated here as a sign of surge-type activity. The dataset contains 116 surge-type events globally from 2017 to 2022. The data reveal a pronounced spatial clustering in three regions (i) High Mountain Asia (HMA, ~50 surges), (ii) Svalbard (~25) and (iii) Yukon/Alaska (~9), with only a few other scattered surges elsewhere. This spatial clustering is significantly more pronounced than the overall global clustering of known surge-type glaciers. The 2017 to 2022 clustering may point to climatic forcing of surge initiation. Details on the method used to detect the surges, validation against existing/auxiliary surge data, and interpretation of the temporal/spatial surge clusters with regard to potential changes in climatic forcing are provided in Kääb et al. (2023).

The surge dataset is based on the interpretation of the above radar backscatter patterns by human operators. Both these human interpretations and the definition of a surge are not sharp and binary (surge vs. non-surge), but clearly indicate strong surge-type glacier-dynamic events. We identified the following major error sources for the detection of surges based on radar backscatter differences between subsequent years:

- **Visual interpretation**: The operator performing the visual analysis may falsely interpret a signal as a surge - or oversee / omit actual surges.
- **Weak and complex changes in glacier surge backscatter**: By nature, some surges do not produce clear changes in backscatter (i.e. there is less heavy crevassing), and the disappearance of crevasses (connected to surge termination) gives a less clear signal than the appearance of crevasses at the beginning of a surge.
- **Spurious surge-like backscatter changes**: Backscatter changes from, e.g., snow/firn melt patterns, (snow) avalanches or uplift/subsidence of glacier parts due to ice-marginal lakes, may resemble surge-like backscatter changes.
- **Topographic effects**: SAR data layover and/or foreshortening can complicate surge detection in valleys with steep flanks

These error sources are also discussed and exemplified in detail by Leclercq et al. (2021) and Kääb et al. (2023).

4.2. Methods for validation

The surge inventory was validated against observed glacier surface elevation changes, velocity time series, terminus advances and crevassing found in optical satellite images (Table 4.1). Additionally, we compared our inventory against existing inventories, which are detecting surges based on speed variations, elevation changes and visual surge signs as indicators (Table 4.1). Many existing inventories of “surge-type glaciers” are limited to a specific region, and may thus not provide a complete overview – i.e., they only record that particular glaciers are known to have surged at some time in the past. The timing of these surges may not always be known, and surges of other (more remote) glaciers in the area may pass unknown if they

didn't happen to occur in the observation period. The RGI v.6.0 indication of surge-type glaciers (RGI surge categories 3 = observed surge, 2 = probable surge, 1 = possible surge) omits the distinction of surges of different tributary glaciers within the same RGI glacier.

We also combined an extended global surge inventory from RGI v.6.0 surges and a number of other regional surge inventories, hereafter referred to as RGI-extended. Additionally, we validated regional surge clusters separately. Specifically, for Svalbard, we used the surge dataset by Farnsworth and others (2016) that contains 708 surge-type glaciers compiled from previous studies and newly mapped using land-based crevasse squeeze ridges as surge indicator. For HMA we used the dataset by Guillet and others (2022) that includes 666 glaciers that surged between 2000 and 2018 in HMA. We also considered Guo and others (2022) who, however, base their surge detection purely on elevation differences, which is a very different approach compared to the one used in the present study. All data used for validation are either publicly available (table 4.1) or shown in Kääb et al (2023).

Table 4.1: Datasets used for validation of the inventory of glacier surges.

Sensor	Period	Product	Coverage	Validation method	Source link
Auxiliary / ASTER	2000-2020	Glacier elevation changes	global	Validation against global glacier elevation changes by Hugonnet et al. (2021)	https://doi.org/10.6096/13
Sentinel-2	2017-2022	optical bands	global	Validation against terminus advances and crevassing found in optical satellite imagery	https://scihub.copernicus.eu/
Sentinel-1	2017-2022	SAR	Spits-bergen	Validation against time series of glacier velocities using standard offset tracking	https://scihub.copernicus.eu/
Auxiliary / optical sensors	2027-2022	optical bands	Scheele-breen	Validation against time series of glacier velocities from repeat Sentinel-2, Landsat 8 and ICEEYE data	Official data providers
Auxiliary	-	RGI 6.0 surge code	global	Comparison with existing surge inventory. Surge codes in RGI6 are from Sevestre and Benn (2015)	https://www.glims.org/RGI/
Auxiliary	2000-2018	HMA_STG inventory	HMA	Comparison with existing surge inventory for HMA by Guillet et al. (2022)	https://zenodo.org/records/5524861
Auxiliary	2008–2012	Svalbard surge inventory	Svalbard	Comparison with existing surge inventory for Svalbard by Farnsworth et al (2016)	https://doi.org/10.1016/j.geomorph.2016.03.025
Auxiliary	-	RGI-extended	global	Combined global surge inventory compiled for this study: RGI 6.0, Sevestre and Benn (2015), Guillet et al. (2022), Farnsworth et al. (2016), Jiskoot et al. (2003), Sevestre and Benn (2015), Bhambri et al. (2017), Goerlich et al. (2020), Guan et al. (2022), Guo et al. (2022)	Literature sources, see reference list

4.3. Validation Results

The validation confirms the robustness of this approach. All identified surges were assessed using visual comparison with alternative surge detection/validation methods listed above. In a qualitative comparison (see Leclercq et al. 2021) we found full agreement with velocity data over test regions on Svalbard. In Alaska/Yukon, seven out of eight detected surge-like events correspond to a large velocity increase (Leclercq et al., 2021). Figure 4.1 shows a visual example of a surge onset in Novaya Zemlya detected with the surge inventory backscatter method (top row) and confirmed with a glacier speed-up (bottom row). The decreasing surge activity is less clearly visible using the SAR backscatter method with continued SAR backscatter

(middle row), but much better by the decreasing glacier velocity (bottom right). Figure 4.2 shows velocity time series for Arnesenbreen, Scheelebreen, and Strongbreen, which were among several detected surges on Spitsbergen, Svalbard.

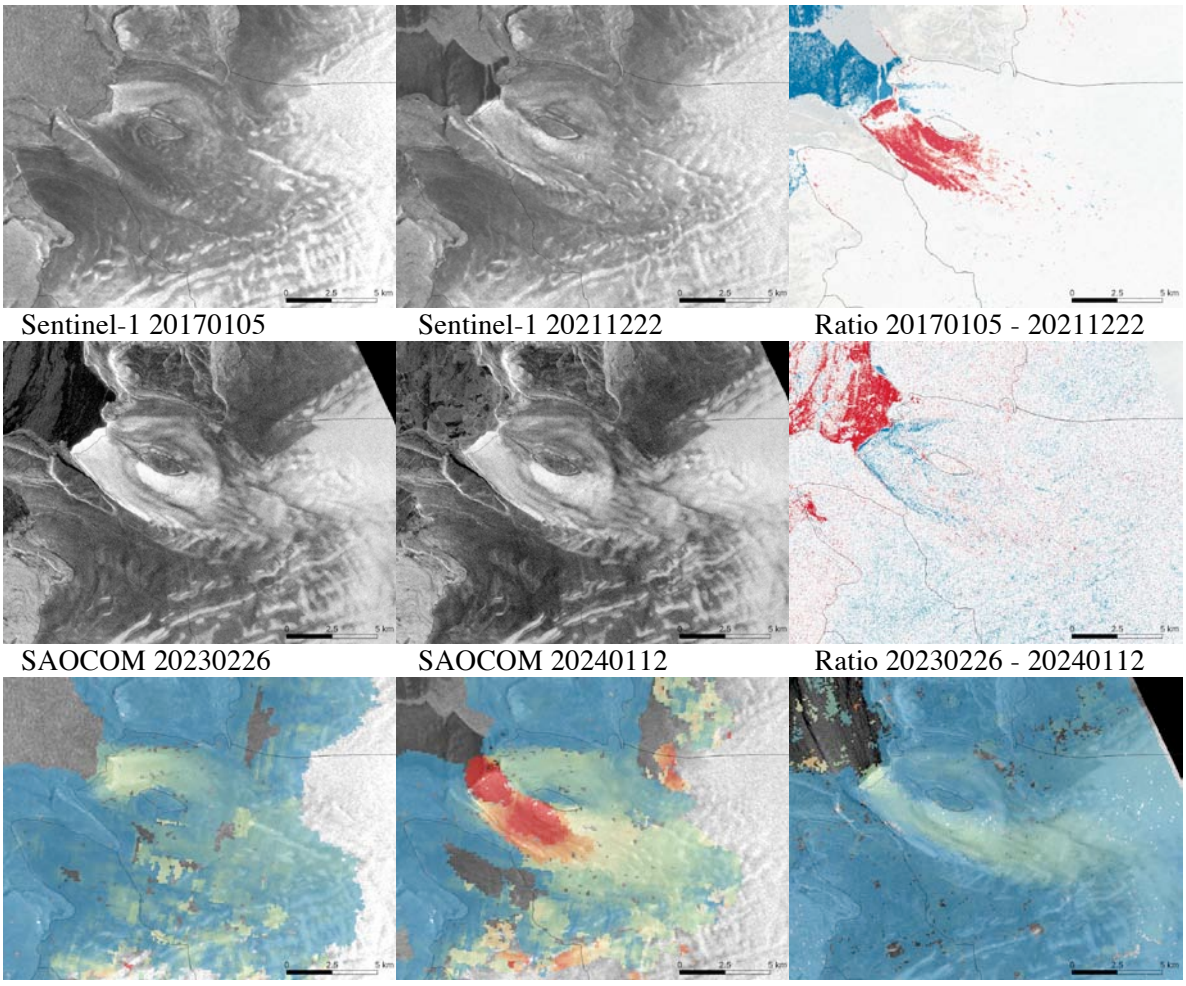

Fig. 4.1: Surge evolution of a Novaya Zemlya glacier with GLIMS ID G067156E76738N / RGI glacier ID RGI60-09.00149 (location: 66.75 N / 76.80 W which showed a surge from ca. 2018 through 2024, with a slow surge start between 2017 and 2021 seen from Sentinel-1 backscatter individual images (top row) and difference (top right, red colours show backscatter increase, i.e. surging). The SAOCOM backscatter data (middle row) indicates a decrease in surge activity between 2023 and 2024. This is confirmed by slow Sentinel-1 velocities in 2016 (bottom left), high velocities during the surge in 2021 (bottom, middle panel, red colours) and subsequent slower velocities when the surge ended (bottom right).

A comparison to the global 2000-2020 glacier volume change dataset from Hugonnet et al. (2021) shows that most surging glaciers exhibit a similar elevation change compared to surrounding non-surging glaciers, but a few surging glaciers lost up to three times more elevation than their non-surging neighbours (for details see Kääb et al., 2023). The relation between glacier mass balance and surging is not trivial as it depends on the surge phase of the individual glaciers. The comparison is thus highly sensitive to the examined time periods. By design, our study does not detect surge-type glaciers during their quiescent phase, which is however an important part of the mass-balance cycle of such glaciers.

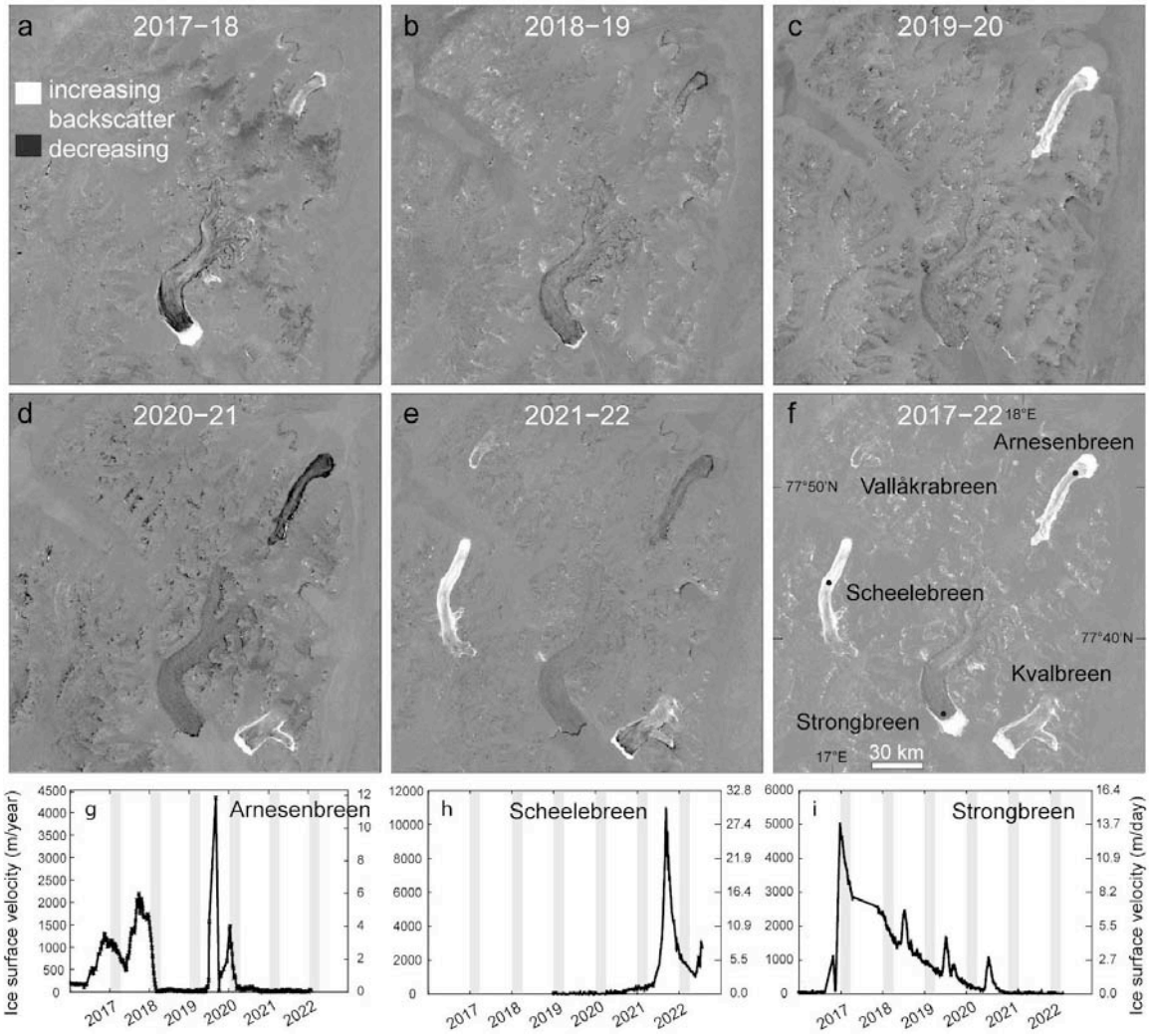

Fig. 4.2: Winter-to-winter Sentinel-1 radar backscatter changes of an area in Spitsbergen of two subsequent years (a-e) and corresponding glacier velocity series (g-i), confirming the detected surges. In the normalised backscatter change panels (a-e), brighter regions indicate increasing backscatter that is interpreted as increasing crevassing, indicating surge-type activity. In contrast, darker regions indicate decreasing backscatter/crevassing/surge-type activity. The middle right panel (f) shows the maximum backscatter normalized difference over the entire five-year period indicating five surges, with velocity profiles available for the single points denoted with black dots. The grey vertical bars in the panels of the lowest row indicate the periods over which Sentinel-1 data have been stacked for examples a-f (1 Jan – 1 Apr of each year). Source: Kääb et al. (2023).

4.3.1. Comparison with surge inventories

Hundred of the 116 detected surges are on glaciers already known as surge-type from existing surge inventories. We provide here a slightly shortened and summarised version of the extensive comparison and analysis of the surge inventory provided in Kääb et al. (2023). Compared to the RGI v6.0 inventory, 56 glaciers detected by this study were marked as surge-type, 28 were marked with the flag "observed", 11 are marked as "probable", 7 as "possible", and 9 are flagged as "no evidence" (RGI-Consortium, 2017). Notably, as this comparison is based on glacier IDs that do not distinguish individual tributaries, it is possible that our analysis and the RGI flag do not necessary mean it is the same glacier tributary that surged, in particular

for large glaciers and glacier systems with one common RGI ID. In our inventory, the same RGI ID can occur several times (up to three times) due to different tributaries of the same larger trunk glacier having surged during that time.

Over HMA, most of the 68 detected surges are also included in the inventory of Guillet et al. (2022), except for four RGI-IDs. These larger glaciers correspond to 25 smaller glacier units with individual GLIMS-IDs. However, the inventory of Guillet et al. (2022) is mostly based on earlier data and barely overlaps with our time period, and they base their inventory on RGI IDs. As for RGI surge flags, we can thus not be sure if the same tributaries surged, even if the RGI IDs agree. Nevertheless, our approach does not detect all glacier surges with 2017/2018 time stamp in the Guillet et al. (2022) inventory. During validation/comparison, we found that we overlooked four surge-type backscatter changes in Sentinel-1 data that were obscured by topographic effects or not obvious due to weak or complex backscatter signals. For completeness, these were included in the final inventory.

Sixty-five other glacier surges with 2017/2018 time stamp in the Guillet et al. (2022) inventory are missing in our inventory as they were not distinguishable from backscatter changes, or showed only a very weak signal that was purposely excluded. These include surges that do not lead to significant crevassing or slow and long-lasting surge-like processes that our backscatter approach does not pick up, as it mainly detects changes between subsequent years/winters. A future version of the method that includes multi-annual stacks of the Sentinel-1 archive may be able to detect some of the missed surge-like events.

Compared to the 708 surge-type glaciers on Svalbard listed by Farnsworth and others (2016), which is also based on an earlier time period/earlier evidence, our dataset contains three new surge-type glaciers. In addition to the 116 detected surges, we excluded 9 unclear cases that are likely surge-type events but were not considered robust enough when validating with auxiliary data (see also Kääb et al. 2023, Table S1 and Fig. S2). An example for an excluded surge-like event of parts of a glacier tongue is shown in Fig. 4.3

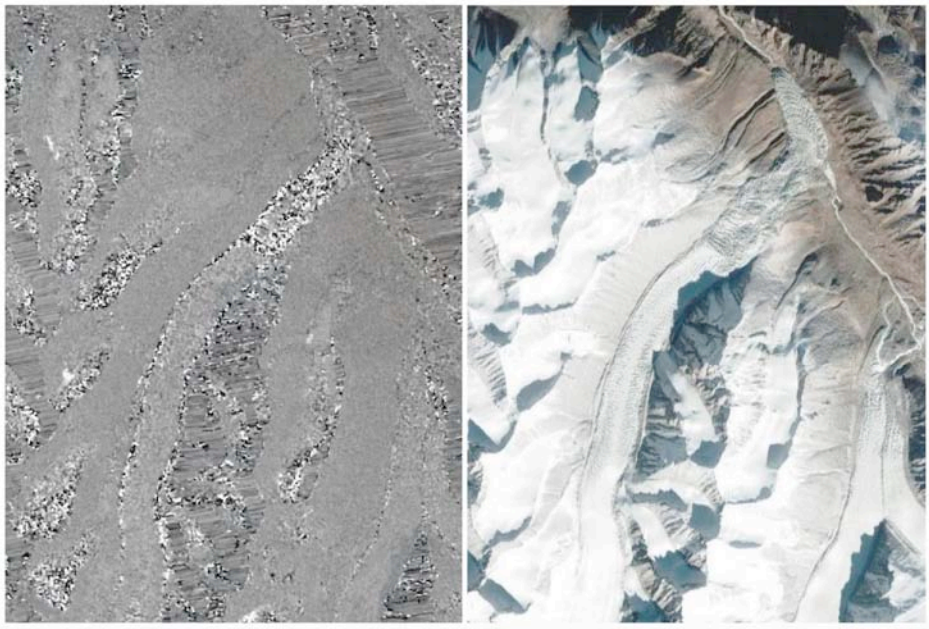

Fig. 4.3: Unclear surge-like event of the glacier with ID G081220E35458N that was excluded from the surge inventory. Left: Sentinel-1 backscatter change, right: satellite image from 10.1. 2020 showing heavy crevassing of the bottom part of the tongue. Source: Kääb et al. (2023), supplementary material.

4.3.2. Comparison of surge clustering

For all detected surging glaciers in this inventory (116 glaciers), we computed the spatial distances to all other surging glaciers (example for one glacier: Fig. 4.4, red dots). This results in a distance-dependent empirical cumulative distribution function (eCDF) of distances between the different glacier samples. It shows the percentage of the sample of glaciers for which this average mutual distance is below a given distance (Fig. 4.4). The same procedure is applied to surges detected by other surge inventories (Fig. 4.4), namely the 1343 RGI 6.0 surges and the RGI-extended surge inventory compiled for this study (Table 4.1), as well as regional surge inventories for Alaska/Yukon, Svalbard and HMA, respectively (Fig. 4.5). Our surge inventory shows higher spatial clustering and has the most pronounced steps, i.e. clusters.

For our surge inventory, on average 21% of all surging glaciers are within <1000 km of one surging glacier (red curve in Fig. 4.4) compared to 16% for RGI "sure" surging glaciers, or 6% of all glaciers (surging and non-surging). We estimated the probability that the clustering we observe could be obtained just by random sampling (see Kääb et al., 2023) and found that relative to the RGI surge data the global surge clustering observed over 2017–2022 is statistically significant and hardly a result of random sampling. On a regional scale, our 2017-2022 glacier surge inventory for Alaska/Yukon is more clustered than surge reference data, but this is not the case for Svalbard and HMA (Fig. 4.5). Glacier surges and underlying processes are likely influenced by climatic factors that are currently changing rapidly. Reference inventories from past time periods may thus represent surge activity under different climatic conditions.

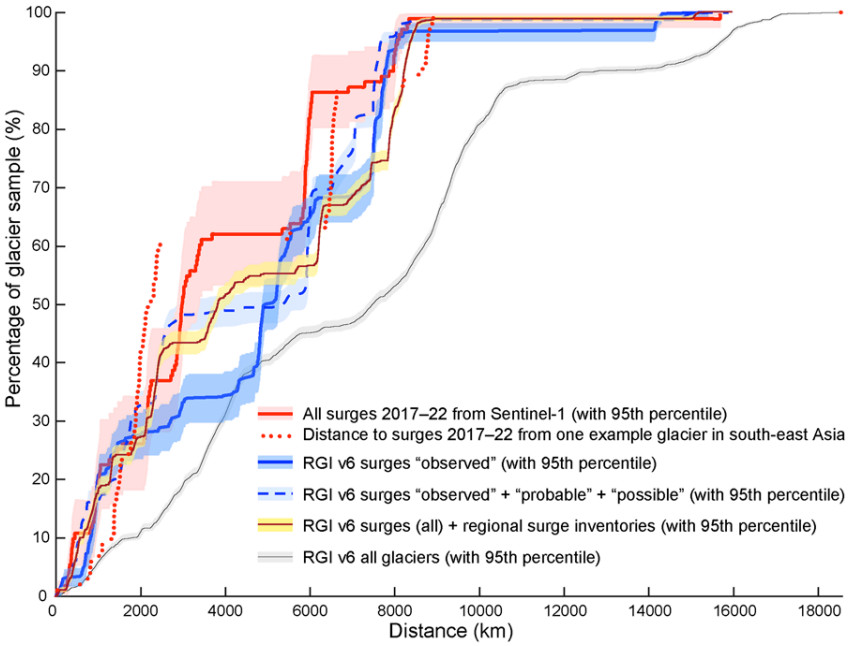

Fig. 4.4: Average distance of any 2017–2022 glacier surge to all other 2017–2022 surging glaciers, given as percentage of the entire sample (empirical cumulative distribution function, eCDF, of distances between 116 glaciers, red curve). Similar distributions for different subsamples of RGI v6.0 glaciers with surge indication (blue curves) and the compiled RGI-Extended inventory (brown curve). The higher on the y-axis and the more stepped a curve appears, the more spatial clustering of the sample it indicates. For illustration, the distance distribution of one surging glacier (Sedongpu Glacier; 94.9° E, 29.8° N, south-east Asia) to all other 2017–2022 surging glaciers is also given (each glacier one red dot). For reference, the eCDF of distances is also given for all RGI v6.0 glaciers (216'502 glaciers including all non-surging glaciers, black curve). Source: Kääb et al. (2023)

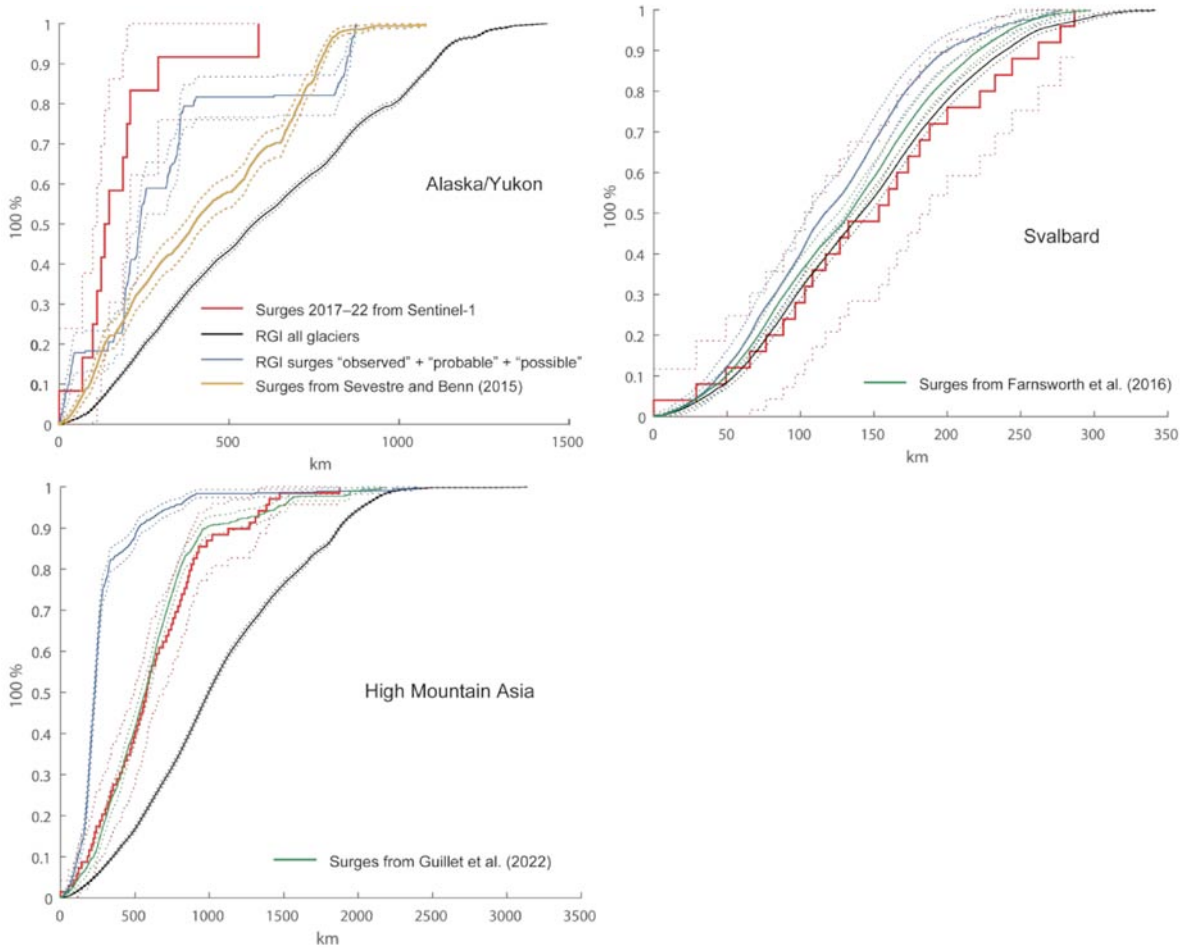

Fig. 4.5: Regional empirical cumulative distribution functions for surge clustering, as in Fig. 4.4. Source: Kääb et al. (2023), supplementary material.

4.4. Conclusions and Recommendations

Our 2017–2022 inventory of surge-type events is to our best knowledge the first globally consistent time-stamped inventory of active surges. It appears robust and well-suited for this purpose, and our extensive validation efforts suggest that we have identified a large percentage of surge-like events with significant backscatter changes. Compared to other surge inventories, our approach is sensitive to changes in crevassing rather than traditionally used surge indications such as glacier terminus advances, elevation changes/surge bulges, glacier speed-up, or indications of past surge activities like looped moraines. Different definitions of surge-type events can cause inconsistencies between surge inventories, and we find that our approach misses out on surge-like events included in other inventories that did not lead to significant backscatter changes such as slow, multi-year events. Other sources of error are related to visual interpretation of the signal, especially for weak or spurious surge-like backscatter changes that are difficult to interpret, or in case of topographic effects that obscure glaciers in steep valleys. On the other hand, the probability for false positives is low, i.e. it is unlikely that our approach would classify a random signal as a surge. Our approach is thus complementary to other (semi-) automatic approaches of surge characterisation and mapping.

5. Elevation change time series

5.1. Background

Five products were made for the Glaciers_cci+ project, of which four are new radar altimetry datasets for the Greenland peripheral glaciers region and one is a laser altimetry dataset derived for use in validation. As discussed in the PVP [RD1], few datasets specifically target the Greenland peripheral glaciers, especially over the full CryoSat-2 mission period (2010 to present), but the datasets made by Khan et al (2022a) provide surface elevation change (SEC) from both ICESat and ICESat-2 individually, and from crossovers between them to fill the time gap, 2008 to 2019, between their missions.

The error budget for radar altimetry measurements includes imprecision of the retracking algorithm used in determining elevation, radar surface penetration, slope-induced error due to surface topography, uncertainty in geophysical and atmospheric correction, radar speckle and sensitivity to anisotropic scattering in the near-surface snowpack (ESA, 2019, Scagliola et al., 2016 and Remy et al., 2006). Their effects are inter-combined and not enough is known to formally propagate these errors through the processing chain from radar echo to elevation measurement. Validation with laser altimetry ensures an independent basis for comparison, as laser altimeters measure from the snow/air interface. In regions that are flat within the radar echo footprint, penetration effects can be seen as a bias between the mean or median elevations from each instrument. As topography within the footprints becomes more complex, penetration and slope effects combine, and can be characterised by the bias between elevations in a set of slope ranges. For discussion of these effects, see Schröder et al. (2019). Seasonal effects may arise, e.g. laser altimeters should be more sensitive to winter snowfall. As an examples of this, see Horton et al. (2022).

5.2. Methods for validation

The validation methodology consists of bulk statistical comparison of elevation and SEC over comparable time-scales and regions. The datasets used for validation (see Table 5.1), are derived from the GLAS and ATLAS laser altimeters on-board NASA's ICESat and ICESat-2 missions respectively. Comparisons are made to the ICESat-2 gridded DEM and SEC time-series products for Greenland, and to SEC datasets specific to the Greenland peripheral glaciers derived from ICESat and ICESat-2 products prepared by Khan et al (2022b). The latter include a dataset covering a similar time period to the CryoSat-2 mission, made by crossover analysis between ICESat and ICESat-2 data.

The ICESat-2 DEM was chosen in preference to other available DEMs as it is the only independent DEM available for a reference time within all three missions used in the radar altimetry products. The Arctic DEM (Polar Geospatial Center, 2023) is not independent as it uses CryoSat-2 data in its vertical registration, and the Copernicus DEM (ESA, 2022) uses earlier data from 2011 to 2015, which is not ideal in such a rapidly changing environment. Both ICESat-2 ATLAS datasets have been updated since the start of this project, so the actual versions used are linked below.

Table 5.1: Summary of alidation dataset parameters.

Dataset	Description	Time range	Spatial resolution	Temporal resolution	Source
ICESat-2 ATL14	High resolution DEM	2000.0	100 m	One instance only	https://nsidc.org/data/atl14/versions/3
ICESat-2 ATL15	Change maps with res-pect to the ATL14 DEM	October 2018 to March 2023	1 km used, coarser available	3 months	https://nsidc.org/data/atl15/versions/3
Khan SEC from ICEsat-ICESat-2 crossovers	Gridded elevation change rates	October 2008 to April 2019	500 m	One instance covering whole range	https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL098915 See 'Open Research' section for a link to the dataset
Khan SEC from ICESat-2	Gridded elevation change rates	October 2018 to December 2021	500 m	One instance covering whole range	As above

There are five quality tests, labelled QA-ECTS-1 to 5. The most important is test 4, which compares SEC directly. Tests 1, 2, 3 and 5 compare elevations. Since the products being tested are primarily time-series of elevation change derived from a modelling process, these are converted to time-series of absolute elevations by recalculation from the results and the model parameters. The joint product is not included in this process as it was created from the elevation change time-series of its individual components, which will be tested.

5.2.1. Test QA-ECTS-1: Elevation

The elevation time-series in each test cell is interpolated to the DEM reference time, 2000.0. This is differenced to the average DEM elevation from all DEM cells within the test cell. The mean and median of the difference distribution over all test cells indicate accuracy, and standard deviation and mean absolute deviation indicate uncertainty.

5.2.2. Test QA-ECTS-2: Penetration

Using the difference dataset made for test 1, select cells in flat regions. As these areas slope effects are not expected, only penetration and bias are indicated by the statistics used in test 1.

5.2.3. Test QA-ECTS-3: Slope effects

Using the difference dataset made for test 1, select cells in regions of increasing slope range. Progressive changes in the statistics used in test 1 to indicate slope effects.

5.2.4. Test QA-ECTS-4: Elevation change rates

The elevation change time-series in each test cell is used to calculate an elevation change rate over the same period as the nearest validation data cell – there are three possibilities for validation, the ICESat and ICESat-2 crossovers and ICESat-2 data used by Khan, and the longer ICESat-2 product made for this project. As for elevation in test 1, the mean and median of the difference distribution indicate accuracy, and standard deviation and mean absolute deviation indicate uncertainty.

5.2.5. Test QA-ECTS-5: Seasonal effects

The elevation time-series in each test cell is averaged by season and differenced to the same season in the nearest validation data cell. Change in the mean and median of the difference distribution per season indicates differing responses to seasonal surface change.

An overview of the quality tests performed and the datasets involved in each is given in Table 5.2 below.

Table 5.2: Overview of the quality tests performed.

Test	Test datasets	Validation dataset	Measure	Calculation	Statistics	Quality measure
QA-ECTS-1	CryoSat-2 point and gridded swath, S3	ICESat-2 ATL14	Elevation parameters	Elevation differencing at near-coincident times and locations	Mean, median, standard deviation, mean absolute deviation	Accuracy and uncertainty
QA-ECTS-2	CryoSat-2 point and gridded swath, S3	ICESat-2 ATL14	Penetration	As above, flat regions only	Mean, median	Bias
QA-ECTS-3	CryoSat-2 point and gridded swath, S3	ICESat-2 ATL14	Slope effects	As above, binned by slope range	Mean, median, count	Bias
QA-ECTS-4	CryoSat-2 point and gridded swath, joint product	Khan SEC from ICESat and ICESat-2 crossovers	Elevation change rate parameters	Elevation change rate differencing over near-coincident time periods and locations	Mean, median, standard deviation, mean absolute deviation	Accuracy and uncertainty
	All	Khan SEC from ICESat-2				
	All	ICESat-2				
QA-ECTS-5	CryoSat-2 point and gridded swath, S3	ICESat-2 ATL14 and ATL15	Seasonal effects	Elevation differencing along time series, binned by season	Mean, median	Time-varying bias

5.3. Validation/Intercomparison Results

The four products tested were:

- CryoSat-2 point swath product: ESACCI-GLACIERS-L3-SEC-GREENPERI_CS2_POINT_SWATH-2010_2023-v001.nc
- CryoSat-2 gridded swath product: ESACCI-GLACIERS-L3-SEC-GREENPERI_CS2_GRID_SWATH-2010_2023-v001.nc
- Sentinel-3 product, containing data from both satellites, A and B: ESACCI-GLACIERS-L3-SEC-GREENPERI_S3-2016_2023-v001.nc
- Joint product, containing CryoSat-2 point swath and Sentinel-3 A and B data: ESACCI-GLACIERS-L3-SEC-GREENPERI_JOINT-2010_2023-v001.nc

The product made as a validation dataset, by stacking ICESat-2 datasets, was: ESACCI-GLACIERS-L3-SEC-GREENPERI_IS2-2018_2023-v001.nc

5.3.1. Coverage

The Greenland peripheral glaciers defined by the Randolph Glacier Inventory (RGI Consortium, 2023) have an overall area of 90482 km^2, and consist of 19994 individual glaciers, the majority of which are smaller than 1 km^2. Figure 5.1 shows their distribution in the periphery of the Greenland Ice Sheet.

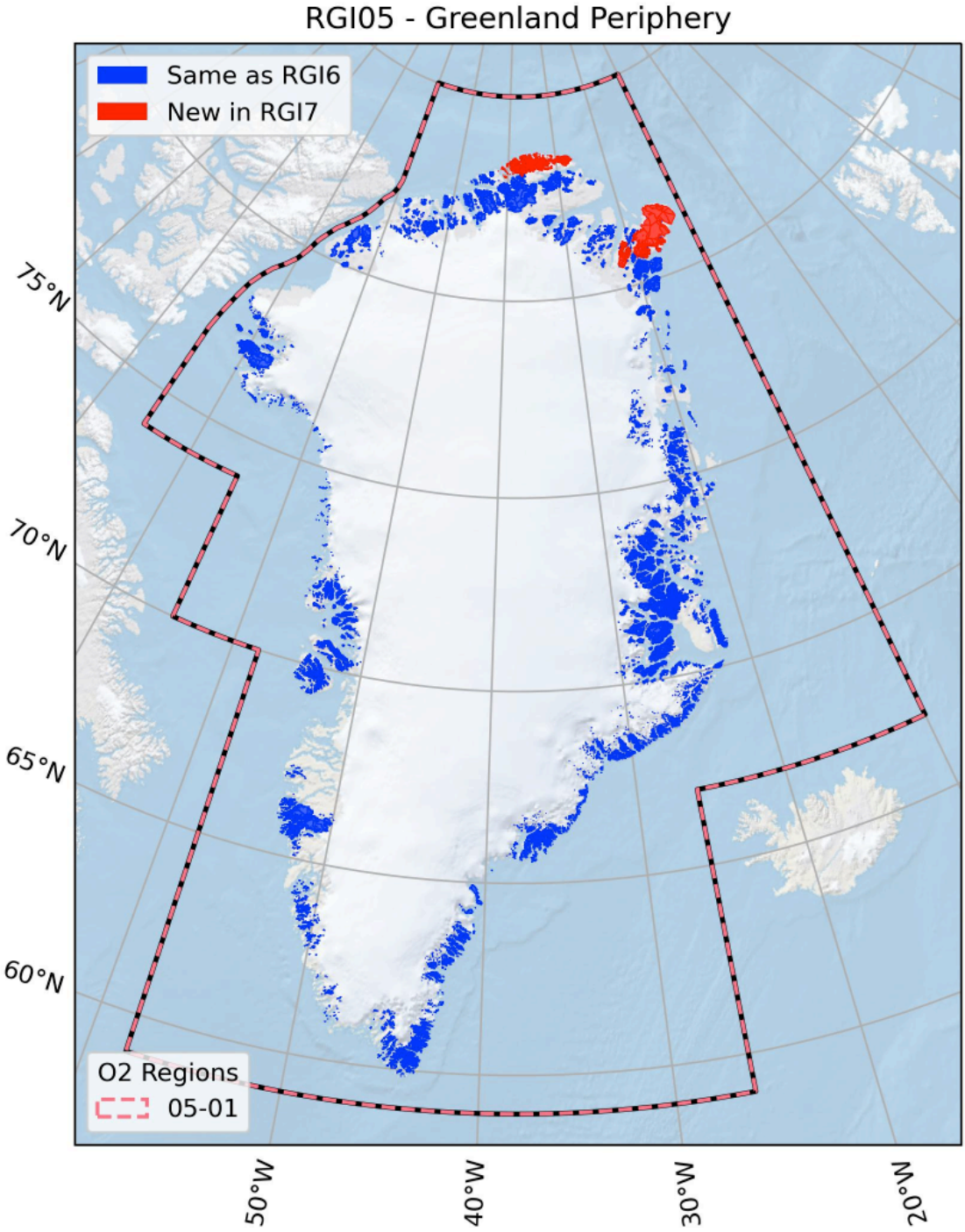

Fig. 5.1: Greenland Peripheral Region, i.e. Region 05, from RGI 7.0. Credit: RGI Consortium (2023).

The processing masks for the gridded products were derived so that any cell used should be at least 50% glacierized. Coverage is calculated as the number and percentage of cells in the processing mask in which a full-mission time-series was calculated, and the number and percentage of glaciers for which a time-series was derived. The results are given in Table 5.3 and cells containing a full-mission time-series are plotted in Fig. 5.2.

Table 5.3: Product coverage

Product	Cell reso-lution	Count of cells containing a time-series	Cells contain-ing a time-series (%)	Number of glaciers with a time-series	Percentage of glaciers with a time-series
CryoSat-2 point swath	500 m	114118	32.20	1299	6.50
CryoSat-2 gridded swath	2 km	3345	15.47	N/A	N/A
Sentinel-3, satellite A	500 m	3621	1.02	19	0.10
Sentinel-3, satellite B	500 m	2882	0.81	19	0.10
Joint	500 m	4129	1.16	46	0.23
IceSat-2	1 km	67203	76.57	1954	9.77

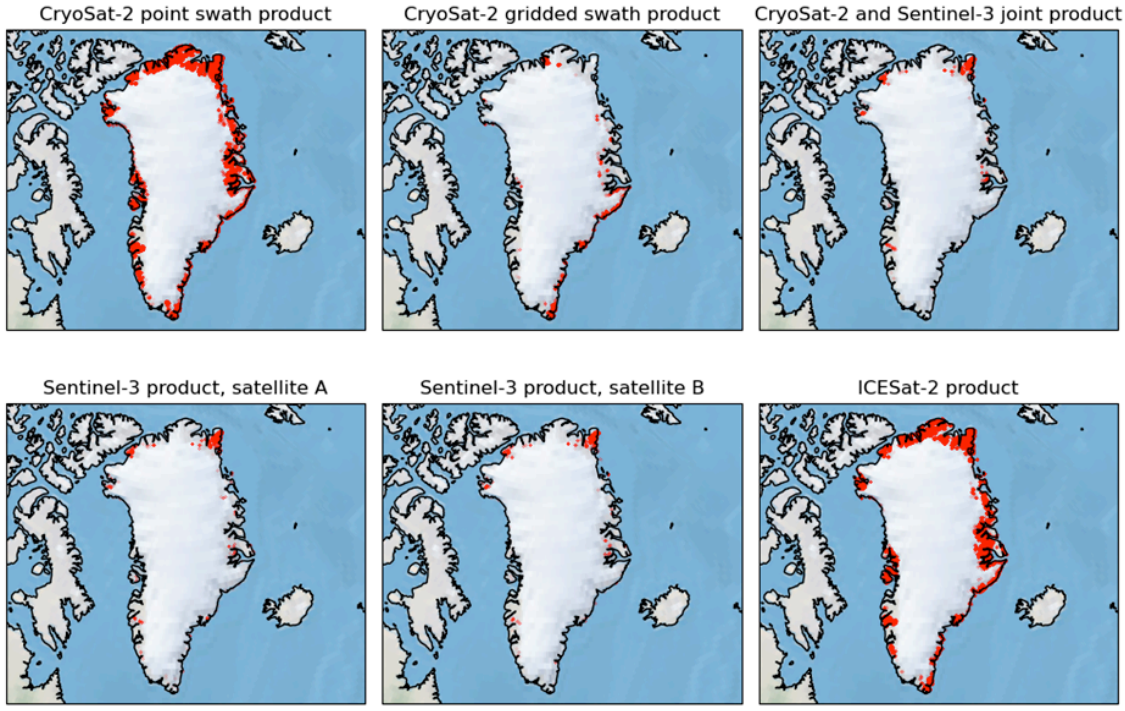

Fig 5.2: Product coverage.

5.3.2. Tests QA-ECTS-1 to 3: Elevation, Penetration and Slope effects

The first three tests all use the same datasets, where the single-sensor products are differenced with the ATL14 DEM. The difference is taken cell by cell. In each cell the elevation time-series is interpolated to give an elevation at the reference time of the DEM, 2020.0. As these cells are all lower resolution than the DEM, the average elevation of all DEM cells within each product cell is taken. The difference in each cell is the averaged DEM elevation minus the interpolated product elevation. Slopes are taken from the Arctic DEM for the centre of each product cell. The distribution is clipped at 3 sigma to remove outliers, and statistics – mean, median, standard deviation and mean absolute deviation – calculated for the full distribution and in half-degree slope ranges. Only the CryoSat-2 point swath product has a substantial number of cells in each slope range, the others are very sparse. Figure 5.3 shows the difference distributions at all slopes, Fig. 5.4 shows the changing distributions by slope range in the CryoSat-2 point swath product, and Fig. 5.5 shows the statistics by slope range for each product. Tables 5.4 to 5.7 give the statistics shown in Fig. 5.5, and the number of samples used to calculate each value.

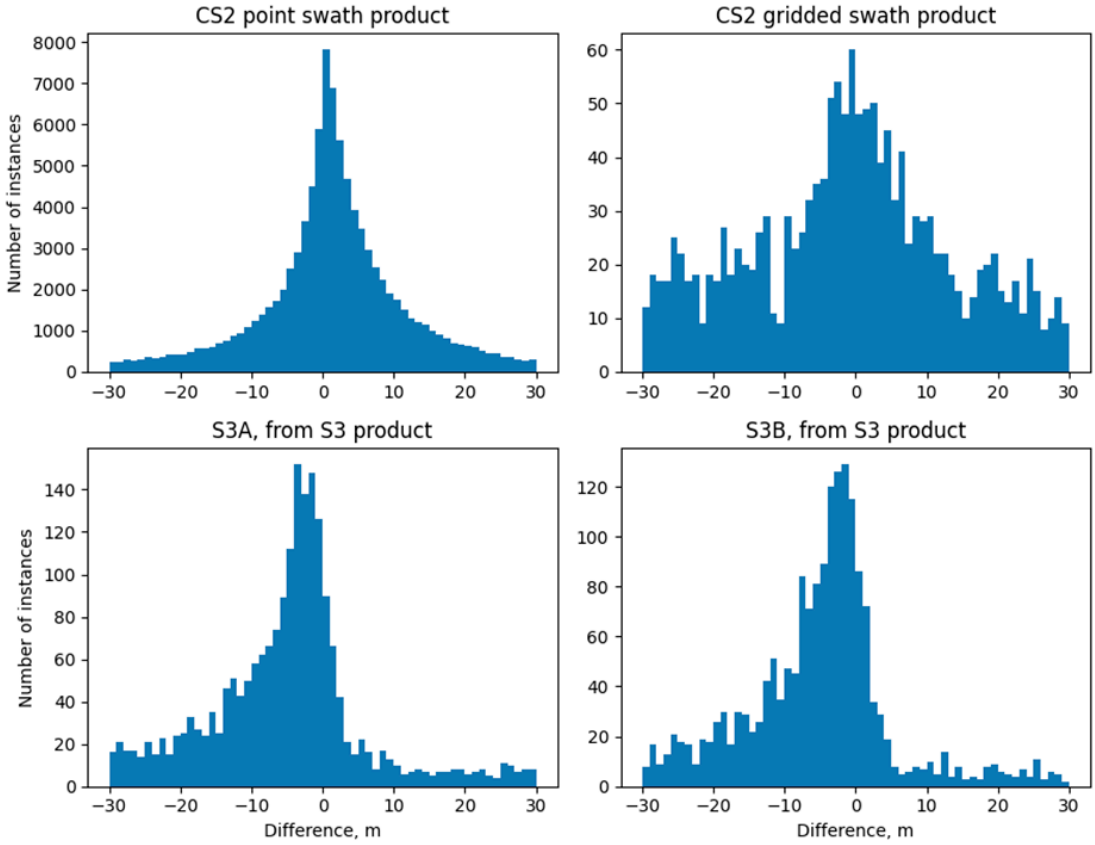

Fig. 5.3: Distributions of per-cell differences, ATL14 DEM minus product. The wings of the distributions have been truncated for clarity.

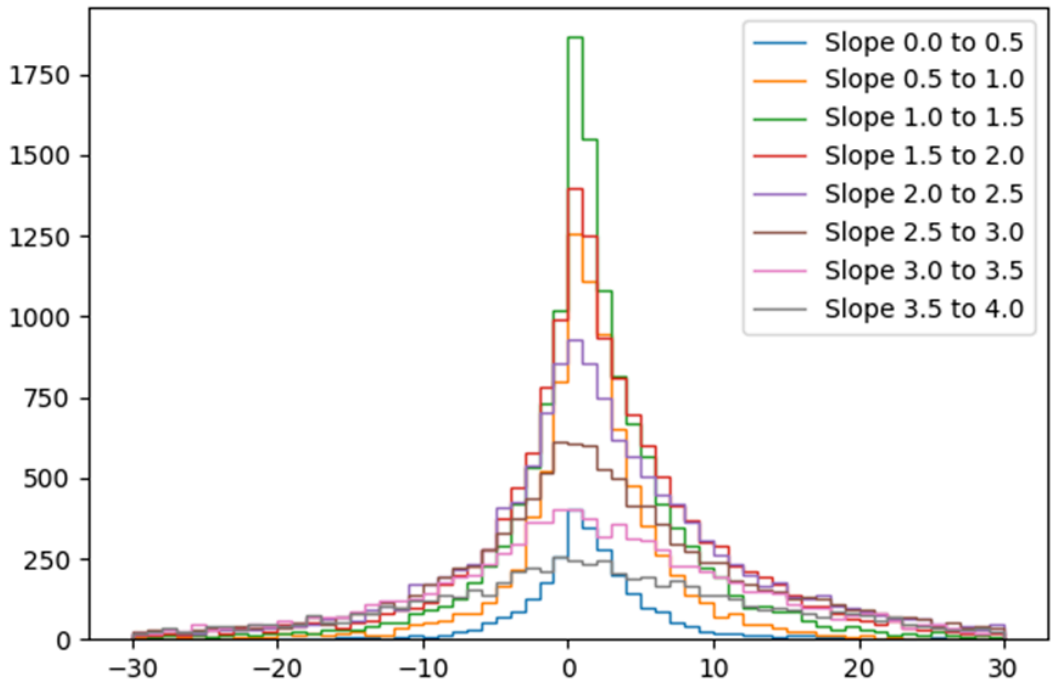

Fig. 5.4: Distributions of per-cell differences, by cell surface slope range, ATL14 DEM minus CryoSat-2 point swath product.

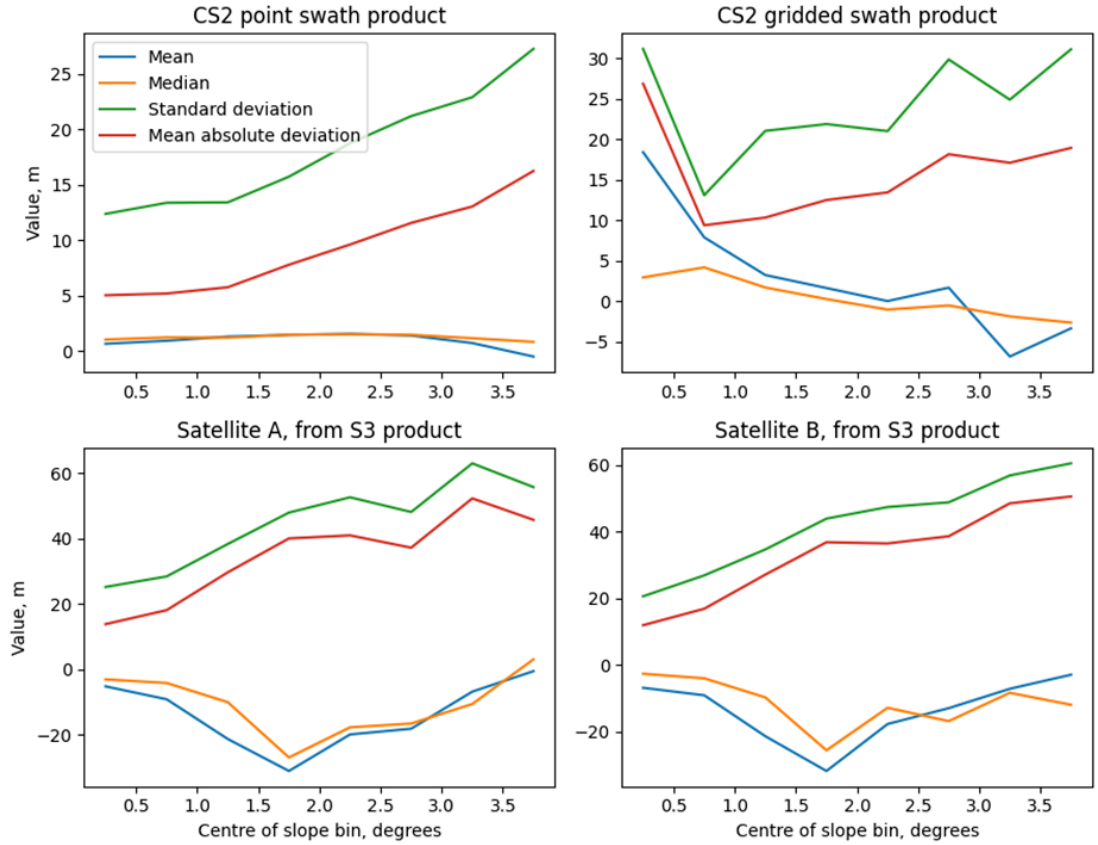

Fig. 5.5: Statistics of per-cell difference distributions, by cell surface slope range, ATL14 DEM minus product.

Table 5.4: CryoSat-2 point swath product, elevation difference to ATL14, by surface slope

Slope range	Mean	Median	Standard deviation	Mean absolute deviation	Number of samples
All slopes	0.38	1.21	26.54	13.59	81869
0 to 0.5°	0.68	1.06	12.39	5.06	2823
0.5 to 1°	0.96	1.26	13.39	5.21	8885
1 to 1.5°	1.33	1.23	13.43	5.78	13411
1.5 to 2°	1.48	1.49	15.75	7.81	14801
2 to 2.5°	1.57	1.55	18.76	9.63	13699
2.5 to 3°	1.43	1.49	21.21	11.58	11633
3 to 3.5°	0.74	1.18	22.91	13.06	9450
3.5 to 4°	-0.47	0.85	27.27	16.26	7167

Table 5.5: CryoSat-2 gridded swath product, elevation difference to ATL14, by surface slope

Slope range	Mean	Median	Standard deviation	Mean absolute deviation	Number of samples
All slopes	-26.24	-11.2	94.37	65.17	684
0 to 0.5°	18.41	2.97	31.15	26.85	4
0.5 to 1°	7.9	4.2	13.1	9.4	28
1 to 1.5°	3.25	1.72	21.04	10.34	77
1.5 to 2°	1.65	0.3	21.88	12.51	110
2 to 2.5°	0.05	-1.0	21.01	13.45	128
2.5 to 3°	1.7	-0.49	29.84	18.16	116
3 to 3.5°	-6.79	-1.85	24.88	17.1	112
3.5 to 4°	-3.32	-2.61	31.1	18.93	109

Table 5.6: S3 product, satellite A, elevation difference to ATL14, by surface slope

Slope range	Mean	Median	Standard deviation	Mean absolute deviation	Number of samples
All slopes	-14.5	-7.45	44.6	32.31	3109
0 to 0.5°	-5.3	-3.16	25.1	13.73	201
0.5 to 1°	-9.24	-4.26	28.32	18.01	1000
1 to 1.5°	-21.39	-10.12	38.24	29.57	943
1.5 to 2°	-31.17	-27.03	47.84	39.94	406
2 to 2.5°	-20.0	-17.81	52.52	40.87	230
2.5 to 3°	-18.24	-16.66	48.01	37.1	127
3 to 3.5°	-6.9	-10.69	62.9	52.18	101
3.5 to 4°	-0.65	2.96	55.63	45.63	101

Table 5.7: S3 product, satellite B, elevation difference to ATL14, by surface slope

Slope range	Mean	Median	Standard deviation	Mean absolute deviation	Number of samples
All slopes	-14.55	-6.85	39.8	28.59	2548
0 to 0.5°	-6.82	-2.61	20.62	11.96	187
0.5 to 1°	-9.06	-3.97	26.9	16.89	882
1 to 1.5°	-21.35	-9.69	34.69	27.13	763
1.5 to 2°	-31.77	-25.57	43.92	36.84	327
2 to 2.5°	-17.68	-12.81	47.4	36.48	180
2.5 to 3°	-12.93	-16.81	48.82	38.63	87
3 to 3.5°	-7.12	-8.32	56.87	48.51	72
3.5 to 4°	-2.9	-11.9	60.5	50.57	50

5.3.3. Test QA-ECTS-4: Elevation change rates

The fourth test uses the products' time-series to calculate surface elevation change rates over the same periods as the three comparison datasets, Khan's ICESa – ICESat-2 crossovers, Khan's ICESat-2 and this project's ICESat-2 product. The Sentinel-3 missions have not been operating for long enough to compare to Khan's crossovers, so they are only compared to the two ICESat-2 datasets.

In each case the difference is taken cell by cell, and comes from the comparator dataset minus the product dataset. Pairs of values are chosen so that they come from the comparator cell centre nearest to the product cell centre, allowing a maximum separation distance of 1 km. The distribution is clipped at 3 sigma to remove outliers, and statistics – mean, median, standard deviation and mean absolute deviation – calculated. Figure 5.6 shows the difference distributions per product against each comparison dataset and Table 5.8 gives the statistics in each comparison, and the number of samples used to calculate each value.

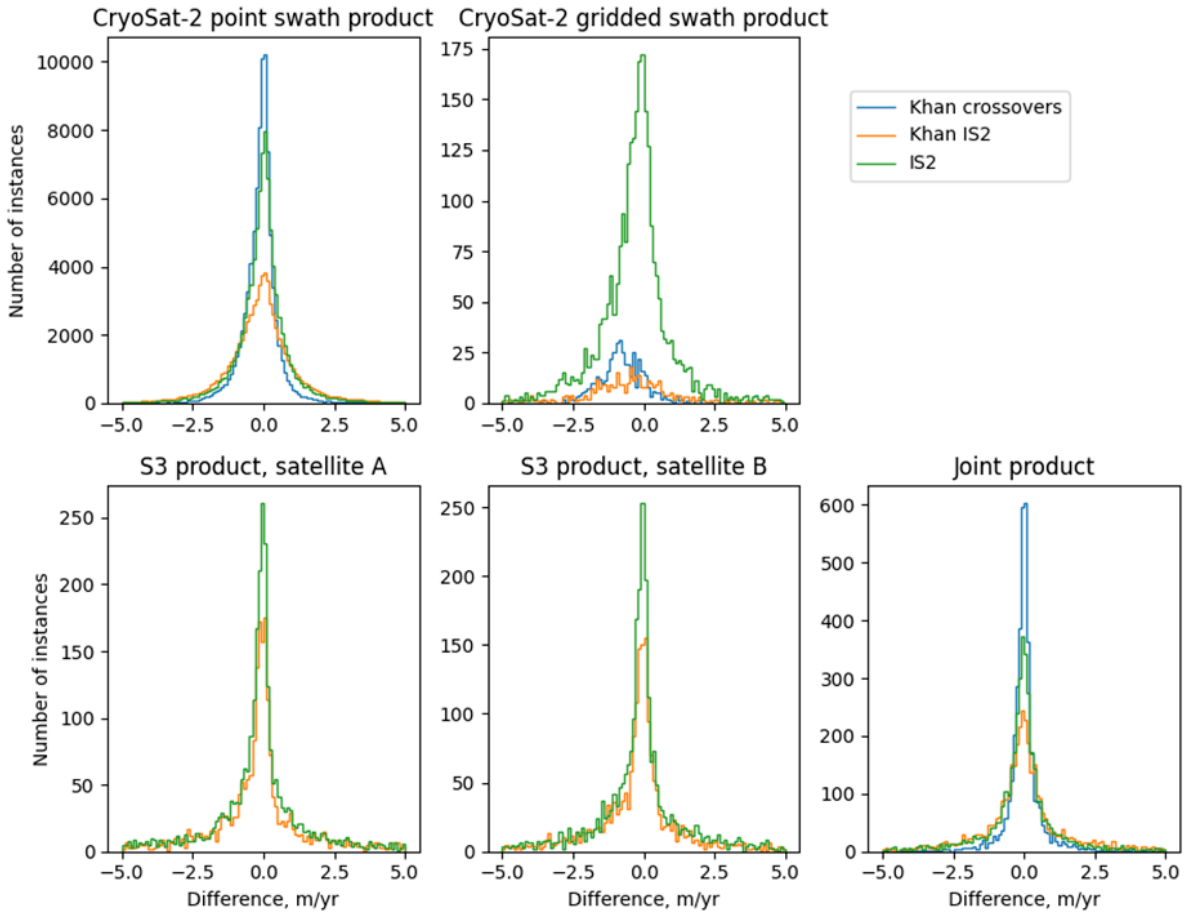

Fig. 5.6: Distributions of per-cell differences, SEC comparison dataset minus product. The wings of the distributions have been truncated for clarity.

Table 5.8: All products, surface elevation change rate difference to comparison datasets

Product	Comparison dataset	Mean	Median	Standard deviation	Mean absolute deviation	Number of samples
CryoSat-2 point swath	Khan crossovers	-0.13	-0.07	0.57	0.4	86220
	Khan ICESat-2	-0.1	-0.08	1.36	0.93	70744
	Product ICESat-2	-0.07	-0.03	0.94	0.63	90491
CryoSat-2 gridded swath	Khan crossovers	-0.8	-0.83	0.83	0.63	433
	Khan ICESat-2	-0.6	-0.46	1.95	1.38	371
	Product ICESat-2	-0.34	-0.26	1.43	0.92	2707
Sentinel3, satellite A	Khan ICESat-2	-0.02	-0.11	4.17	2.23	2406
	Product ICESat-2	-0.31	-0.11	3.27	1.81	2941
Sentinel3, satellite B	Khan ICESat-2	-0.1	-0.12	4.1	2.19	2187
	Product ICESat-2	-0.19	-0.11	3.01	1.65	2715
Joint	Khan crossovers	-0.03	-0.02	0.59	0.35	3674
	Khan ICESat-2	-0.05	-0.07	2.68	1.5	3846
	Product ICESat-2	-0.14	-0.06	1.76	0.98	3958

5.3.4. Test QA-ECTS-5: Seasonal effects

The last test compares the ICESat-2 product elevations by season to the other single-sensor products. The ICESat-2 time-series are produced by stacking the ATL15 elevation changes onto the ATL14 DEM, which is averaged to match ATL15's resolution. Product cell elevation time-series are averaged by season and a comparison made. Sentinel3-B was launched later than ICESat-2, and does not have comparison values for the first two seasons. Figure 5.7 shows the mean and median differences for each season, both in succession and superimposed by year. Table 5.9 gives the mean and median differences and number of samples for the average of each season over the whole time period.

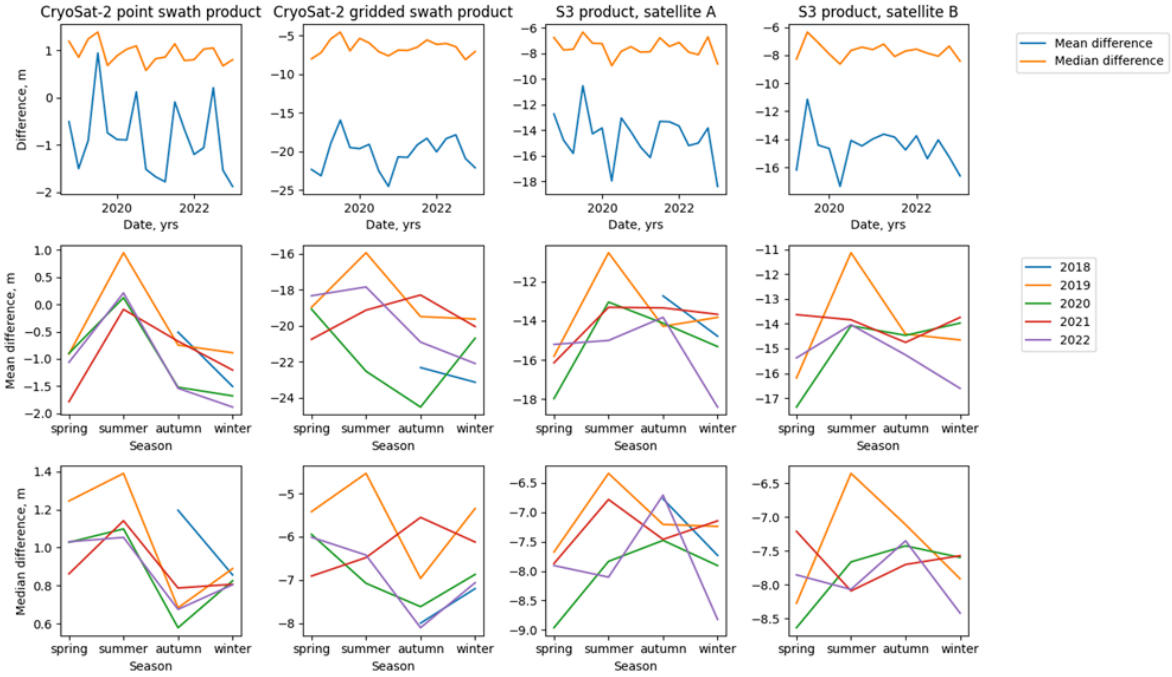

Fig. 5.7: All products, seasonal mean and median differences to the ICESat-2 product, in each year. Top row – means and medians in time order, middle row – means per season, years superimposed. Bottom row – median values per season, years superimposed.

Table 5.9: All products, seasonal mean, median and number of samples over the whole comparison period.

Product	Measure	Spring	Summer	Autumn	Winter
CryoSat-2 point swath	Mean	-1.17	0.3	-1.0	-1.43
	Median	1.04	1.17	0.78	0.84
	Sample count	307970	310603	395274	390680
CryoSat-2 gridded swath	Mean	-19.28	-18.94	-21.19	-21.18
	Median	-6.11	-6.1	-7.28	-6.57
	Sample count	8356	8193	10400	10330
Sentinel-3, satellite A	Mean	-16.3	-13.01	-13.68	-15.28
	Median	-8.12	-7.3	-7.15	-7.77
	Sample count	8132	7580	10112	10331
Sentinel-3, satellite B	Mean	-15.61	-13.31	-14.73	-14.74
	Median	-7.99	-7.54	-7.41	-7.84
	Sample count	7891	7696	7420	7739

5.3.5. Selected example intercomparison

As the bulk statistics above show, the products are only available in certain regions and should be used with care. However, there are situations where the CryoSat-2 point swath product may be of use, and an example is given here. Flade Isblink is a large and very flat ice cap to the north-east of the Greenland ice sheet, partially covered by product data (Fig. 5.8).

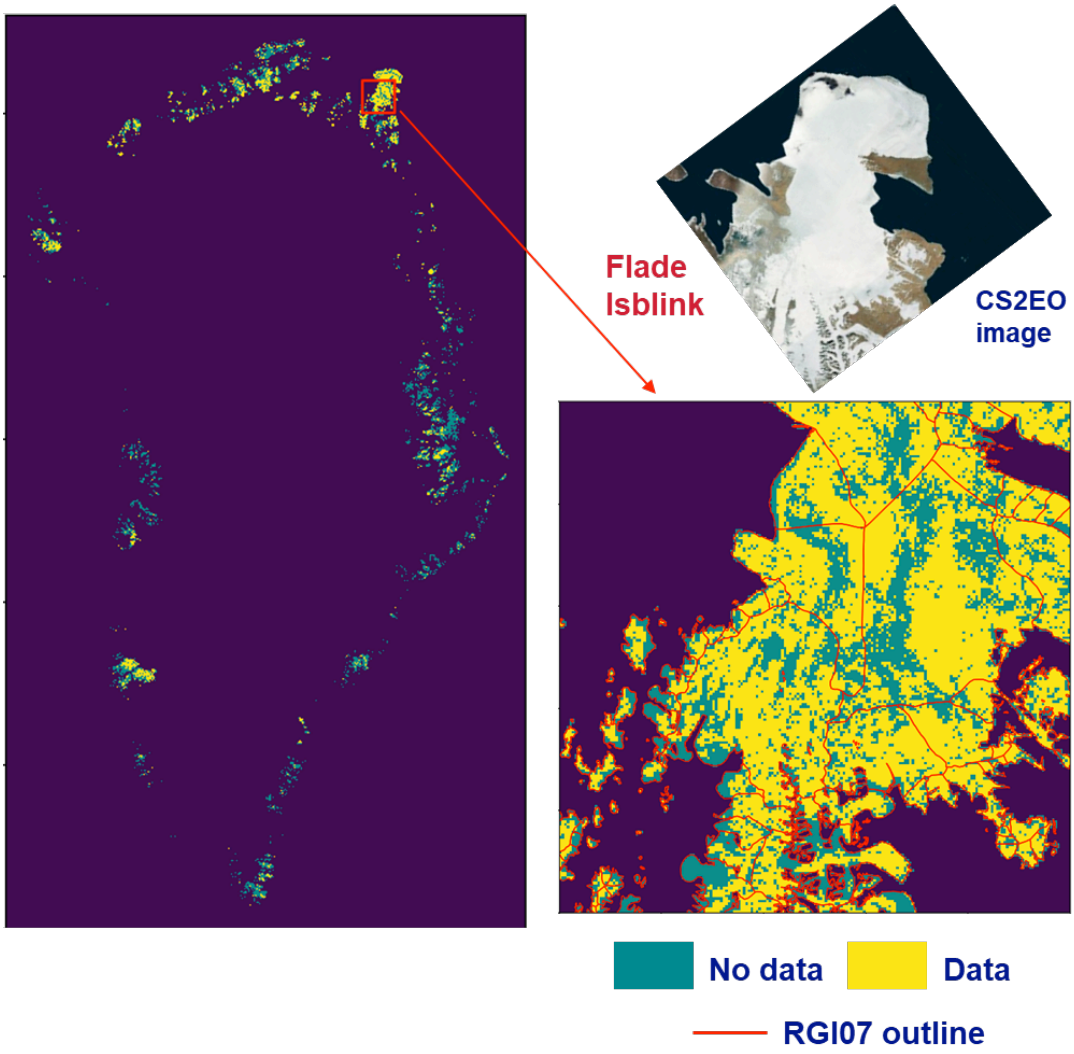

Fig. 5.8: Part of Flade Isblink, showing coverage by CryoSat-2 point swath product.

The glacier to centre left of Fig. 5.8, pointed out in the lower row of Fig. 5.9, below, is shown by Khan's crossover dataset to be gradually gaining ice, while in the ICESat-2 product it is shown as losing ice. The CryoSat-2 point swath product time-series for an example cell in the centre of the sinking region agrees with both comparison datasets - over the full CryoSat-2 mission period the surface change is positive, but over a time window of the last five years the change is negative. The geographic variation in sinking is shown by an example use case in the Product User Guide [RD2], repeated below in Fig. 5.10.

However, as an example of the caution needed, the glacier to the centre top of Fig. 5.8, pointed out in the upper row of Fig. 5.9, does not agree with Khan's crossover dataset, although it does agree with the ICESat-2 product.

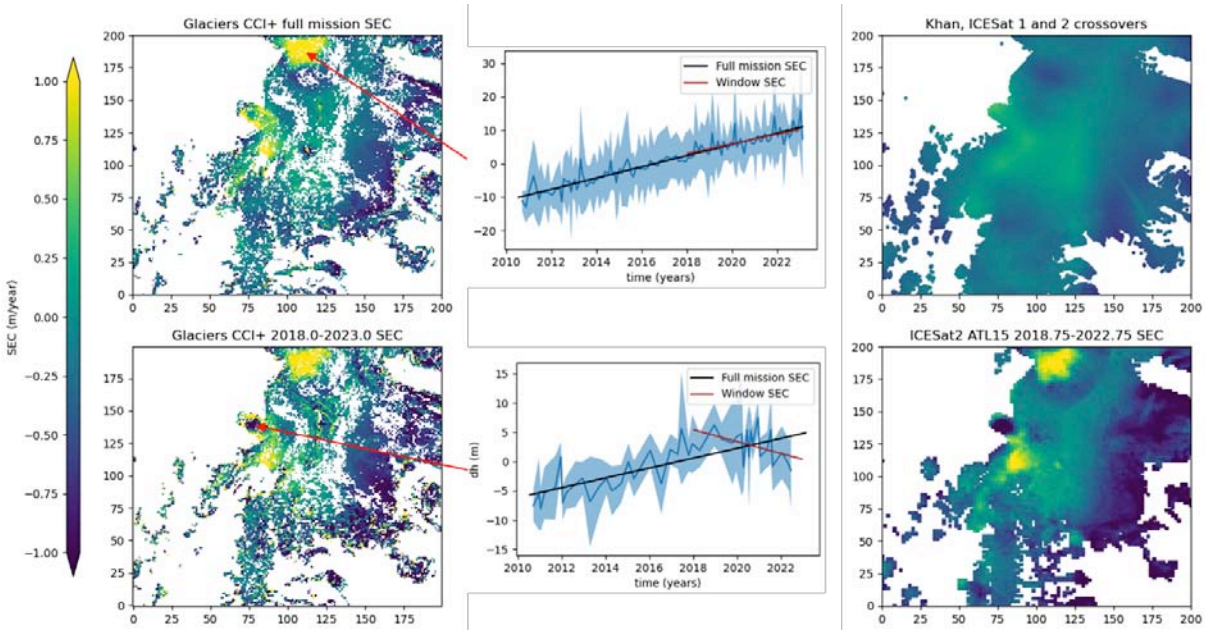

Fig 5.9: SEC rates over Flade Isblink from the CryoSat-2 point swath product (left-hand plots and time-series, top plot full mission duration, bottom plot time-windowed to the last 5 years) and comparison datasets, Khan's crossovers (top right) and the ICESat-2 product (bottom right).

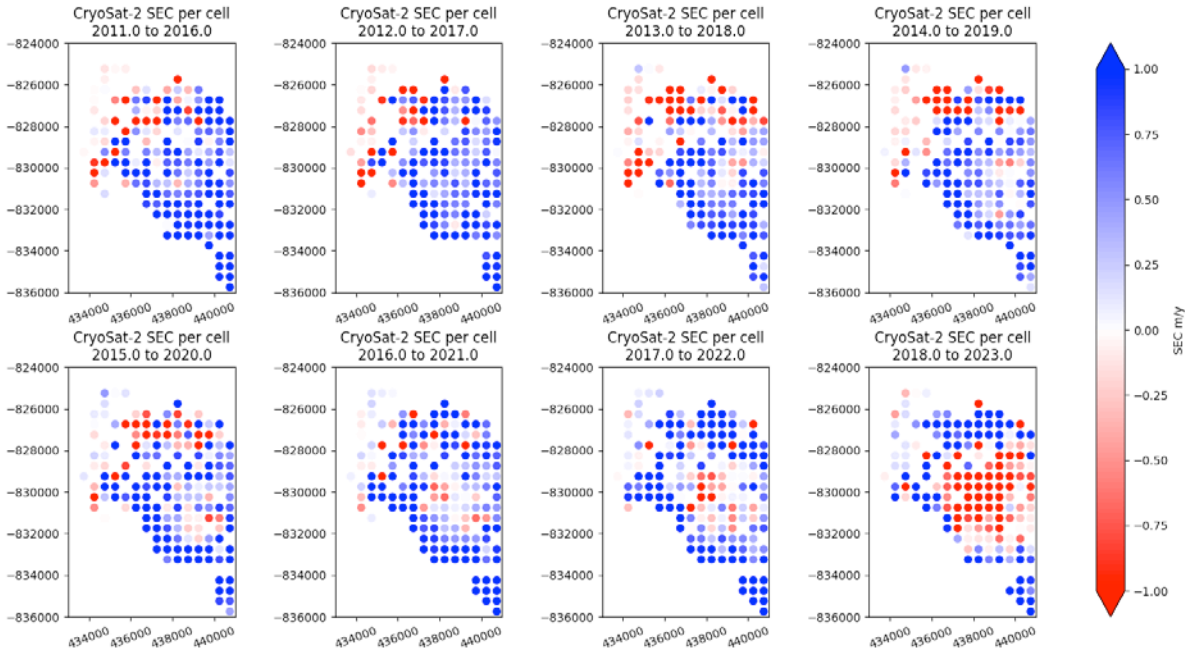

Fig. 5.10: Geographic variation in SEC from the CryoSat-2 point swath product for example sinking Flade Isblink, given by 5-year overlapping time windows.

5.4. Conclusions and Recommendations

5.4.1. Coverage

Based on coverage alone, of the radar altimetry products only the CryoSat-2 point swath product provides a large dataset, and even that only covers a third of the peripheral glaciers region (Fig. 5.2). Because of their orbital limits both Sentinel-3 satellites, and therefore the joint product, cannot view the northern glaciers, which are important because they are losing increasing amounts of ice over time (Khan, 2022a). Further, 68% of the glaciers have an area of less than 1 km^2, which means they are likely to fall between the orbit tracks of the Sentinels, and are smaller than the CryoSat-2 gridded swath resolution. The gridded swath, at time of writing, also misses many glaciers that lie at a distance from the main ice sheet.

5.4.2. SEC quality tests

The SEC statistics from test QA-ECTS-4 show all products have some correlation with the three comparison datasets, centred near to zero but with a wide distribution (Fig. 5.6). The breakthrough target from GCOS is represented by an uncertainty (standard deviation or mean absolute deviation) of the differences of 2 m / decade, which is 0.2 m/yr (GCOS 2022). There is no threshold target given. The CryoSat-2 point swath product, with long-term standard deviation of 0.57 m/yr, is the closest to target, and the joint product value is similar because all of its pre-2016 data comes from CryoSat-2.

Surface elevation changes result from an analysis of change, so biases to absolute elevation are not seen. Hence, the CryoSat-2 point swath product may be of some use where it exists, especially as it covers a long time period and overlaps with earlier missions such as Envisat.

5.4.3. Elevation quality tests

There are no GCOS targets for absolute elevation, but again, although the statistics are centred close to zero, the difference distributions are wide, biased and skewed (Fig. 5.3). Partially this can be accounted for by differences in DEMs, especially at lower elevations, as detailed in the 'Known issues' section of the Product User Guide [RD2]. For Sentinel-3, relocation of data points using slopes from the Alfred Wegener Institute DEM (Helm et al. 2014) moves the data to very different locations than would be indicated by the Arctic DEM. Because of this effect, data points closer to the edge of the periphery were preferentially rejected.

An analysis of differences by slope range was attempted in tests QA-ECTS-1 to 3. Penetration effects should be visible in the lowest slope bin, and combined with slope effects in the higher bins, leading to an increasing difference between datasets as slope increases. Unlike ice sheets, the peripheral glaciers are generally highly sloped – see for example the centreline angle distribution in Liu et al (2022). An analysis of the ATL14 DEM cells that were used to calculate the central elevation in each 500 m grid cell showed a mean range of 20 m in height change across each cell, which, if taken to be along the diagonal, corresponds to a 2° slope. Because of this there were lower numbers of samples in the lowest slopes, with most observations concentrated in the 1 to 3° slope range. Results from the lowest slope range for the gridded swath should probably be discarded as it only held four data points.

The difference datasets have large variance. The expected behaviour is seen in the deviation measures, but mean and median show their biggest difference in the mid-level slopes. Possibly the whole distribution is skewed by other factors, such as satellite track viewing angle with respect to the glacier centrelines, or location of the glacierized region within each cell

area – slope is taken from the centre of the cell, but the glacier may lie in part of the cell away from the centre.

5.4.4. Seasonality

The seasonality test, QA-ECTS-5, looks for an elevation difference between laser and radar measurements by season. It was expected that differences would be larger in winter due to surface snowfall only affecting the laser. This comparison had much larger datasets than for the previous elevation tests as all slopes were used, but even so only the CryoSat-2 point swath product dataset used a substantial number of data points. This did show a seasonal signal with its smallest mean difference in summer and largest in winter, but when comparing individual years the differences vary erratically. The per-year results for mean and median (Fig. 5.7, lower two rows) show a different shape, as the effects of the wings of the distribution are more pronounced in the mean. The total mean/median over all years is affected by the higher autumn/winter values of 2018, which can be seen as much higher than in all other years.

5.4.5. Summary

- Only the CryoSat-2 point swath product is usable, and only with care, but it may be worthwhile where a long time-series is required.
- Radar altimetry is not suited to the very difficult peripheral glacier terrain, but point swath processing provides the best dataset.
- CryoSat-2's 'drifting' orbit, coupled with its high latitude range that encompasses the northern glacial fringe, is the only configuration capable of seeing the majority of the peripheral glaciers.
- Recent DEMs should be used in processing whenever possible.

6. Flow velocities across sensors

6.1. Background

In the Glaciers_cci+ Phase 2 project several new satellite sensors were tested to study the capabilities and limitations for glacier monitoring. For ice velocity, these were the SAOCOM-1 and ICEYE SAR satellite constellations.

The SAOCOM-1 SAR mission is composed of two satellites (SAOCOM-1A and 1B) with L-band polarimetric SAR launched in 2018 and 2020, respectively. The mission is managed and operated by the Comisión Nacional de Actividades Espaciales (CONAE). SAOCOM-1 provides L-Band SAR data over glaciers and ice sheets as a background mission, without a systematic acquisition plan, and has a revisit time of 16 days (with 1 satellite) and 8 days (with the constellation).

The ICEYE SAR satellite constellation is capable of providing 1-day repeat coverage observing at X-Band co-polarisation. ICEYE operates at Stripmap mode or the SPOTLIGHT Extended Area Mode, providing data with a resolution in the order of 1 m, which is suitable to retrieve ice velocity for alpine type glaciers. In order to access the data, the ICEYE SAR satellites must be tasked.

We generated demonstration products over Patagonia, Svalbard, Novaya Zemlya (see Section 4) and Franz-Josef Land using both SAOCOM-1 and ICEYE sensors. These products were not included in the CRDP, but are here included for a preliminary quality assessment.

6.2. Methods for validation

The ice velocity products that are generated within the project are inter-compared with independent reference data sets from different sensors following the quality assessment tests described in the Product Validation Plan (PVP; [RD1]). The selection of suitable reference datasets aims to take into account the resolution of the datasets and time-lag between product and reference dataset. The time-lag is kept at a minimum to avoid contamination of the statistical results due to glacier dynamic changes. However, larger time lags are not always avoidable. The intercomparisons are carried out qualitatively as a visual representation and quantitatively on a pixel-by-pixel level. The resulting statistics are provided together with graphical visualisations, e.g. scatter diagrams and histograms. The quality of the demonstration products are also assessed against velocity products from other sources and sensors such as ITS LIVE, GOLIVE or ENVEO's Sentinel-1 and TerraSAR-X ice velocity products. Figure 6.1 shows for instance the ice velocity maps of the Southern Patagonian Ice Field.

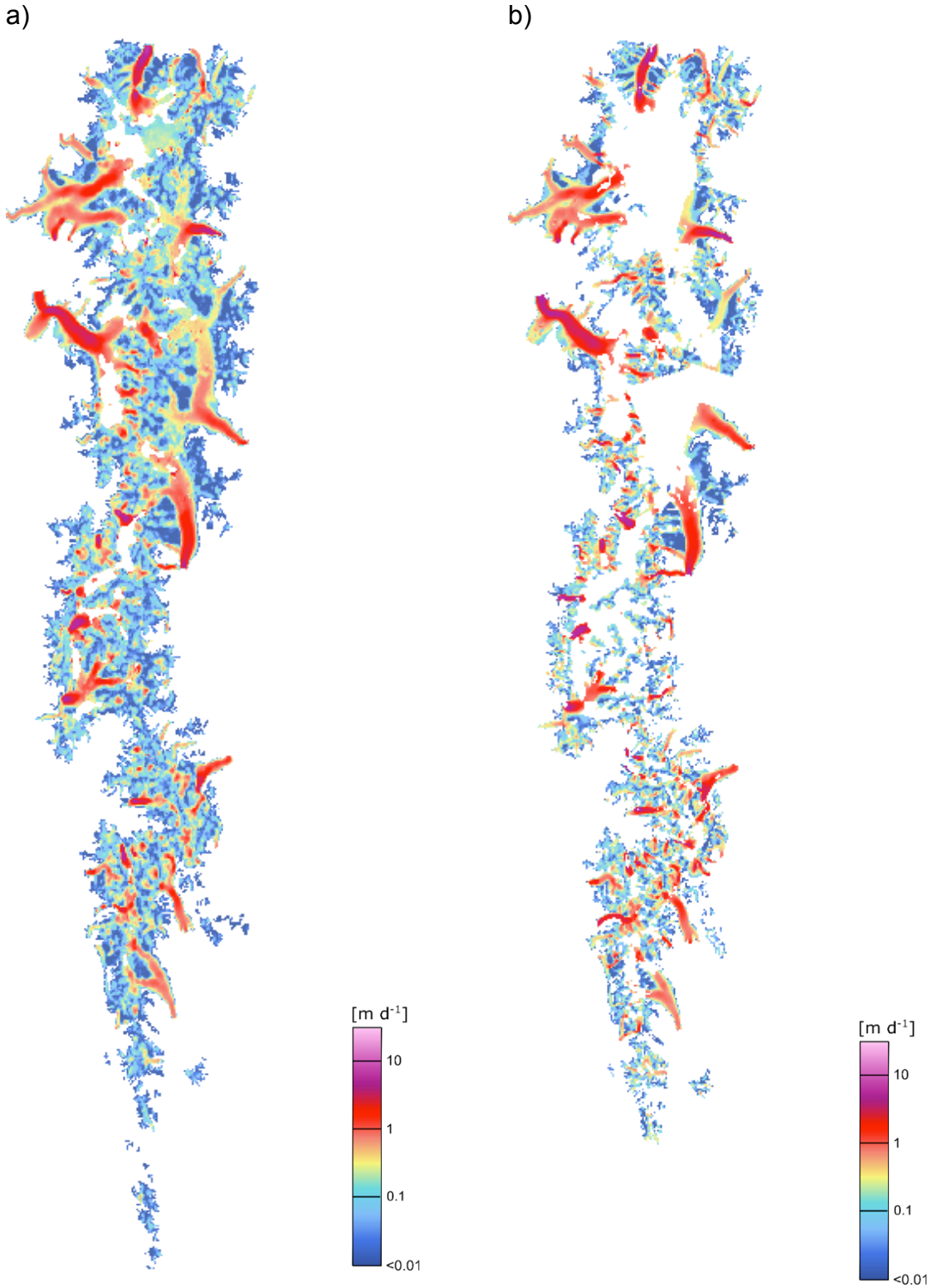

Fig. 6.1: Ice velocity of the Southern Patagonian Ice Field derived from a) Sentinel-1 (2015-2018), b) TerraSAR-X (2016).

6.3. Validation Results

6.3.1. SAOCOM demonstration product for Patagonia

The SAOCOM demonstration product for Patagonia covers the Southern Patagonian Ice Field (SPI) and consists of a merged mosaic based on ice velocity maps covering the period June-October 2022 (Fig. 6.2a). The ice velocity map is derived by applying offset tracking using repeat pass SAOCOM StripMap data. The map is provided at 100 m grid resolution and covers approximately 91% of the icecap based on RGI v7 outlines (Fig. 6.2b).

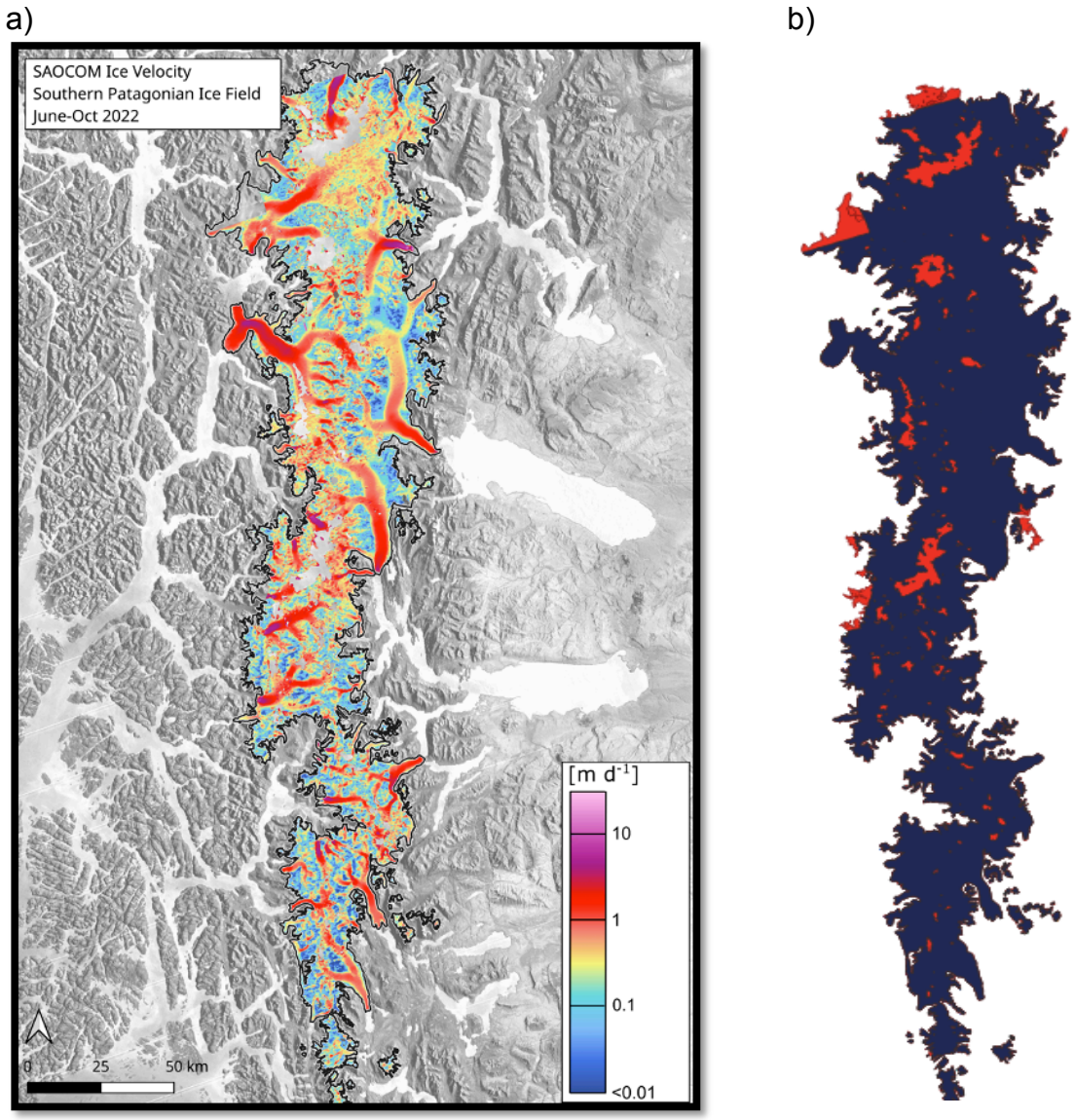

Fig. 6.2: a) Ice flow velocity of the Southern Patagonian Ice Field (SPI) derived from offset tracking using SAOCOM StripMAP SAR data acquired in the period June-October 2022. b) Outline of SPI (RGI v7.0) indicating gaps in the ice velocity coverage (red).

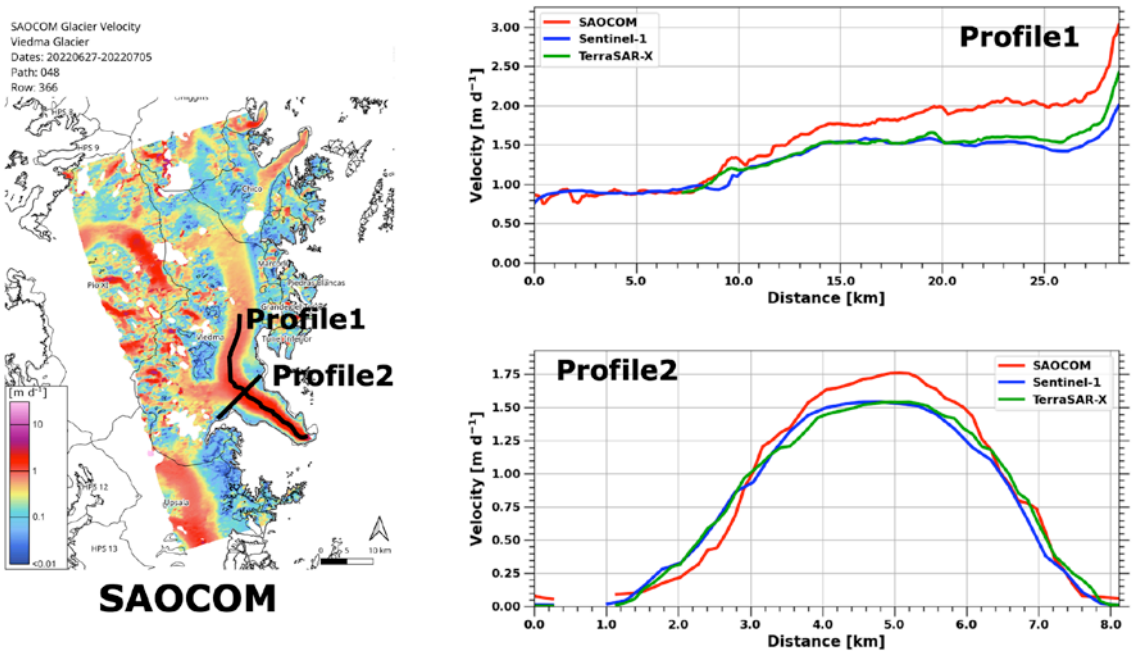

Fig. 6.3: SAOCOM ice velocity map covering Viedma Glacier in July-Aug 2022 (left panel) and intercomparison with Sentinel-1 long term averaged map (2015-2018) and TerraSAR-X derived ice velocity (Sept 2010) along two profiles (right panels).

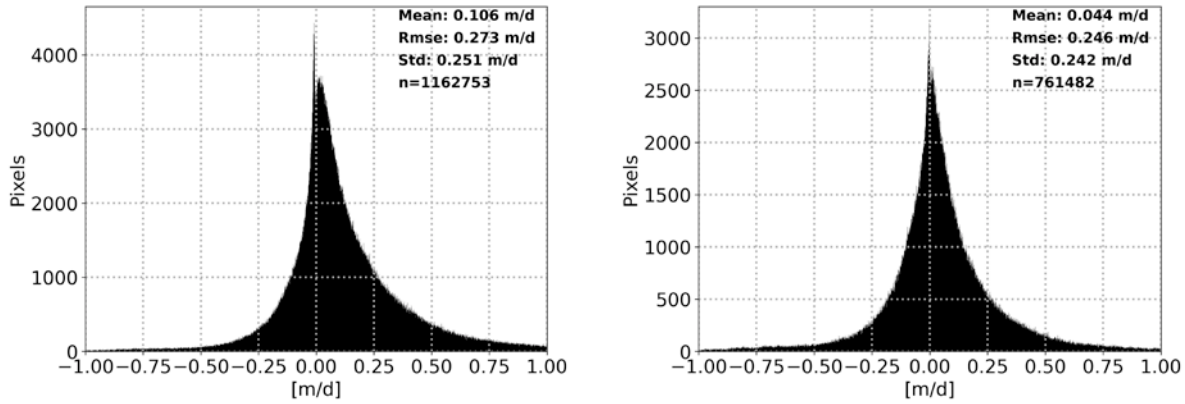

Fig. 6.4: Histograms of the pixel by pixel differences between the SAOCOM and Sentinel-1 (left) and TerraSAR-X (right) ice velocity products.

Figure 6.3 provides an intercomparison of SAOCOM derived ice velocity covering Viedma Glacier with Sentinel-1 and TerraSAR-X derived ice velocity along two profiles in along and across flow direction. Note that the Sentinel-1 and TerraSAR-X derived velocity maps cover different periods in time and therefore an exact agreement is not expected. The difference in velocity includes real dynamic changes over the time period. Figure 6.4 provides histograms of the pixel by pixel differences between the SAOCOM and Sentinel-1/TerraSAR-X ice velocity magnitude of the mosaics. The calculated statistics are mean difference, root mean square error (RMSE) and standard deviation (std). Based on a sample size >1M pixels the mean difference between the SAOCOM and Sentinel-1 dataset is in the order of 0.1 m/d with an RMSE of 0.27 m/d (std 0.25 m/d). For the comparison between SAOCOM and TerraSAR-X the mean difference is in the order of 0.04 m/d with an RMSE of 0.25 m/d (std 0.24 m/d) based on a sample size of >750k pixels.

6.3.2. SAOCOM and ICEYE demonstration products for Svalbard

The SAOCOM and ICEYE demonstration products for Svalbard cover the central part of Spitsbergen, where we find a cluster of on-going glacier surges (Table 6.1, from N to S).

Table 6.1: Characteristics of selected recent glacier surges in Svalbard.

GLIMS_ID	RGI_ID	Lon	Lat	Surge_start	Surge_end	Name
G017158E77876N	RGI60-07.00266	17.114	77.865	2021	3000	Vallåkrabreen
G016964E77694N	RGI60-07.00283	16.974	77.728	2020	3000	Scheelebreen
G017366E77696N	RGI60-07.01470	17.366	77.696	2022	3000	Paulabreen
G016796E77570N	RGI60-07.01474	16.796	77.570	2023	3000	Doktorbreen
G016915E77433N	RGI60-07.01472	16.700	77.430	2022	3000	Listoelbreen

The SAOCOM ice velocity map (Fig. 6.5a) is derived by applying offset tracking using repeat pass StripMap data acquired on 22 February 2022 and 26 March 2022 with a time interval of 32 days. The map is provided at 100 m grid resolution and covers approximately 97% of the glacierized area based on RGI v6 outlines. Over the ice-free region, the mean and standard deviation of the SACOM velocity map are 9 m/a and 24 m/a, respectively. The ICEYE ice velocity map (Fig. 6.5b) is derived by applying offset tracking using repeat pass StripMap data acquired on 4 and 5 January 2022 with a time interval of 1 day. The map is provided at 100 m grid resolution and covers approximately 97% of the glacierized area based on RGI v6 outlines. Over the ice-free region, the mean and standard deviation of the SACOM velocity map are 148 m/a and 285 m/a, respectively. Figure 6.5c provides an intercomparison of the SAOCOM and ICEYE derived ice velocity maps with C-Band Sentinel-1. Note that the SAOCOM, ICEYE and Sentinel-1 derived velocity maps cover different periods in time and therefore an exact agreement is not expected. The difference in velocity includes real dynamic changes over the time period. The SAOCOM velocity map clearly surpasses the quality of the Sentinel-1 product in terms of spatial coverage and lower noise level.

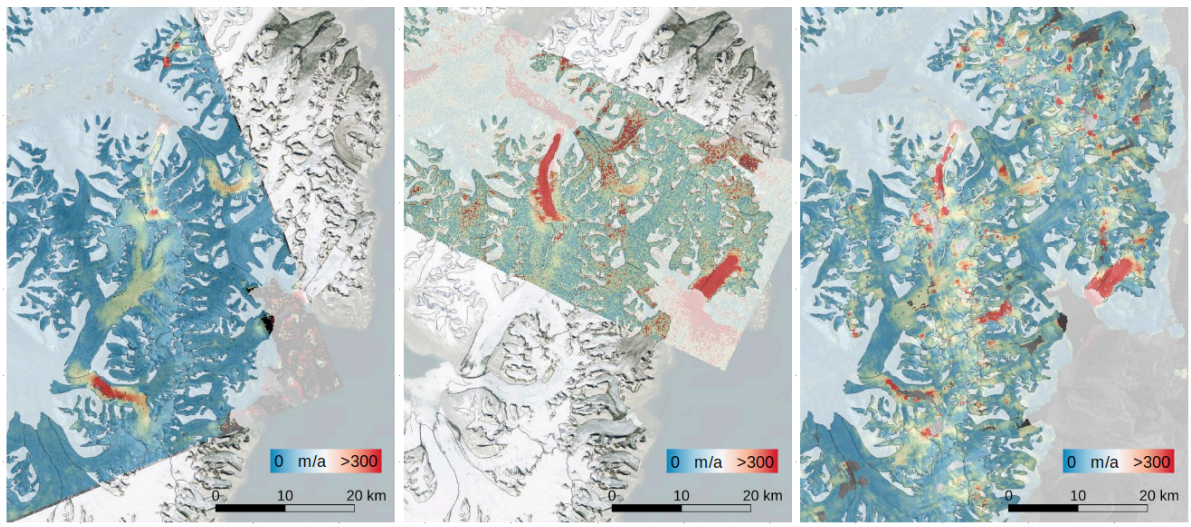

Fig. 6.5: Ice flow velocity of Central Spitsbergen derived from a) offset tracking using SAOCOM StripMAP SAR data of the period 22.2. to 26.3.2022 (32 days), b) ICEYE-1 StripMAP SAR data of the period 4. to 5.1.2022 (1 day) and c) Sentinel-1 TOPSAR data of the period 14. to 26.2.2022 (12 days, right). Glacier outlines in black are from RGI 7.0.

This can be seen particularly well in the profile along the centre line of Paulabreen (Fig. 6.6). However, the SAOCOM ice velocity is largely underestimated over the very fast-moving and narrow Scheelebreen (Fig. 6.7). Here, the best results are obtained with ICEYE data over one

day, which indicate maximum velocities of around 2500 m/a or 7 m/d. On the other hand, with a very short time interval of one day the quality of the ICEYe velocity map for slow-moving glaciers is quite low (see Fig 6.6).

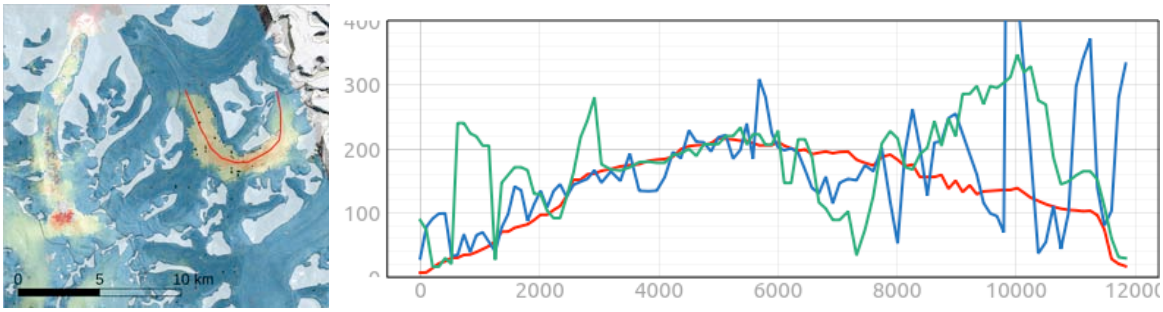

Fig. 6.6: SAOCOM ice velocity map covering Paulabreen in the period 22.2.2022 to 26.3.2022 (left panel, red is >300 m/a) and intercomparison between SAOCOM (22.2.2022 to 26.3.2022, red line), ICEYE (4.1.2022 to 5.1.2022, blue line) and Sentinel-1 (14.2.2022 to 26.2.2022, green line) derived ice velocity along a profile (right panel).

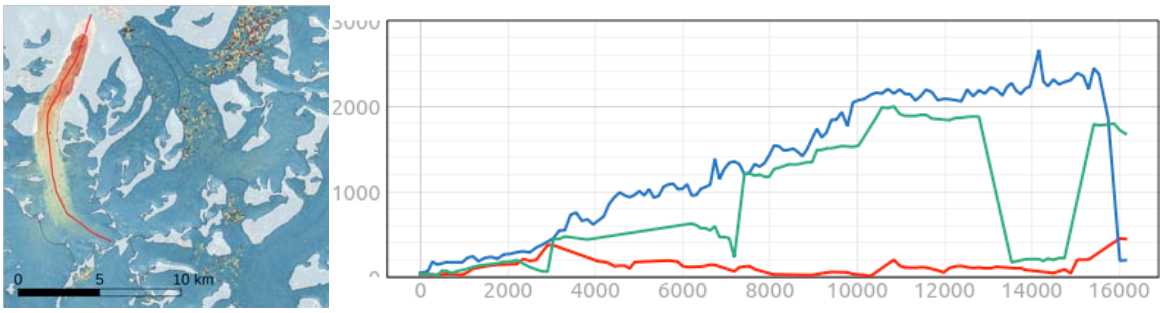

Fig. 6.7: ICEYE ice velocity map covering Scheelebreen in the period 4.1.2022 to 5.1.2022 (left panel, red is >2500 m/a) and intercomparison between SAOCOM (22.2.2022 to 26.3.2022, red line), ICEYE (4.1.2022 to 5.1.2022, blue line) and Sentinel-1 (14.2.2022 to 26.2.2022, green line) derived ice velocity along a profile (right panel).

6.3.3. SAOCOM demonstration product for Franz Josef Land

The SAOCOM demonstration product for Franz Josef Land covers McClintock Island (Fig. 6.8a). The ice velocity map is derived by applying offset tracking using repeat pass StripMap data acquired on 20 November 2023 and 23 January 2024 with a time interval of 64 days. The map is provided at 100 m grid resolution and covers approximately 96% of the icecap based on RGI v6 outlines. Over the ice-free region, the mean and standard deviation of the SACOM velocity map are 17 m/a and 25 m/a, respectively. Figures 6.8b and 6.8c provide an intercomparison of the SAOCOM derived ice velocity map with L-Band JERS-1 and C-Band Sentinel-1. Note that the JERS-1 and Sentinel-1 derived velocity maps cover different periods in time and therefore an exact agreement is not expected. The difference in velocity includes real dynamic changes over the time period. Despite the comparatively large time interval of 64 days, the SAOCOM velocity map surpasses the quality of the JERS-1 and Sentinel-1 products in terms of spatial coverage and lower noise level.

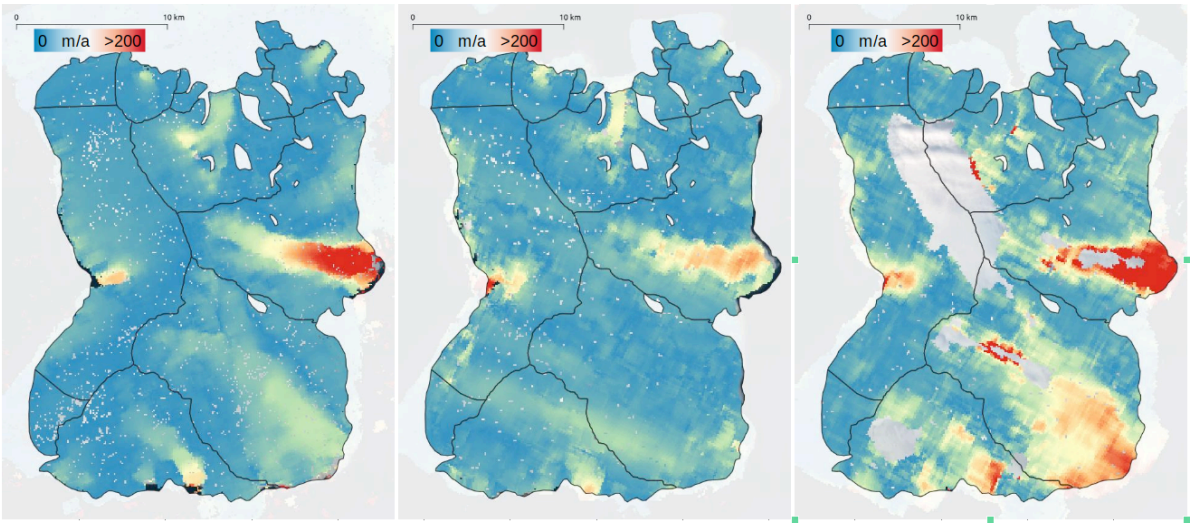

Fig. 6.8: Ice flow velocity of McClintock Island in Franz Josef Land derived from offset tracking using SAOCOM StripMAP SAR data acquired in the period 20.11.2023 to 23.1.2024 (64 days, left), JERS-1 StripMAP SAR data of the period 6.1.1998 to 19.2.1998 (33 days, middle) and Sentinel-1 TOPSAR data of the period 30.11.2021 to 12.12.2021 (12 days, right). Glacier outlines in black are from RGI 6.0.

6.4. Conclusions and Recommendations

We tested the use of SAOCOM-1 L-band and ICEYE X-band SAR data for ice velocity retrieval over the Southern Patagonian Ice Field, Svalbard and Franz Josef Land. For this purpose, the SAOCOM-1 catalogue was explored for suitable repeat pass acquisitions with short temporal baselines and ICEYE image pairs were ordered from the satellite operator.

The available SAOCOM image pairs for the Southern Patagonian Ice Field were processed using offset tracking and then merged to generate a velocity mosaic. The resulting velocity map provides nearly complete coverage of the ice cap. The velocity fields do not appear to be affected by ionospheric stripes as observed on the Greenland Ice Sheet. On the other hand, the available SAOCOM image pairs for Svalbard and Franz Josef Land are slightly affected by ionospheric stripes. However, they could be corrected very well with the available algorithms. A main limitation for regional ice velocity mapping is that currently SAOCOM StripMap data are not acquired on a systematic basis resulting in a limited number of suitable repeat pass pairs precluding the generation of velocity time series or InSAR applications. A revision of the acquisition strategy of SAOCOM for this purpose is therefore recommended.

The available ICEYE image pairs for Svalbard show the excellent performance of this very high-resolution X-band satellite data for recording the very rapid ice motion of comparatively narrow surging glaciers. This was well demonstrated in the case of Scheelebreen. However, tasking of the satellite images by the satellite operator is required. This satellite can therefore contribute to studies at a local level, but its use for large-scale, systematic monitoring is currently ruled out.

7. Geolocation intercomparison

7.1. Background

The mis-registration between Landsat 8 and Sentinel-2 images was obvious since autumn 2015, when the first Sentinel-2 images arrived. The geolocation shift could reach more than 30 m, i.e. one Landsat pixel. This is far more than the given nominal geolocation uncertainty of about 12 m for both sensors (Storey et al. 2016). It basically excluded the joint use of both datasets and was well visible when overlaying glacier outlines derived from historic Landsat scenes with those derived from Sentinel-2 (e.g. Fig. 10 in Paul et al. 2022). A study that used glacier outlines derived from Sentinel-2 as a base to derive a 'new' glacier inventory from Landsat 7 data acquired in 2001 had thus to use manual delineation and virtual adjustment of the shift (Paul et al. 2023). This shift was recognized early and it was decided to reprocess the Landsat archive with the much better ground control of Sentinel-2 using a new Ground Reference Image (Storey et al. 2016) and space-based bundle adjustment (Rengarajan et al. 2020).

Here we investigate the effects of this adjustment by calculating the shifts between the images resulting from the old and new processing. We focus on the old and new versions of the same Landsat scenes but have also determined the residual shifts between Sentinel-2 and Landsat 8 with the new processing for comparison. We investigated the shifts for scenes from the Alps, New Zealand and Patagonia acquired in 2015, 2000 and 2016, respectively.

7.2. Methods for validation

The displacement magnitude and azimuth was calculated for the image pairs using standard normalized cross-correlation that we usually apply in Glaciers_cci to determine flow velocities of glaciers from optical images. When used with the same but differently processed images, changes due to the different processing should become well visible. The normalized cross-correlation processing chain is described in the Glaciers_cci ATBD and, for instance, by Kääb and Vollmer (2000). We applied matching template sizes of 20-30 pixels (300 m) and grid distances of 200-300 m. The results provide the x- and y-components of the two-dimensional shifts, and their polar coordinates, i.e. local shift magnitude and azimuth. We display shift magnitude and azimuth to visualize deformations and rotations between the images.

7.3. Validation Results

The results of the intercomparison of images before and after reprocessing are summarized in Table 7.1 and visualized in Figures 7.1 to 7.3. For the Swiss Alps, mean displacement magnitudes of about 12 m are found between the old and new processing of Landsat 8 as well as between Sentinel-2 and Landsat 8 with the new processing. The previous shift of >30 m could thus be reduced to about 1/3. Also the Sentinel-2 reprocessing (with a better DEM) resulted in a shift, but only by 6 m. When analysing the related colour-coded displacement maps (Fig. 7.1), one can see for the Landsat intercomparison a trend surface and additional local shifts, often along mountain ridges with steep slopes and hence likely resulting from a changed DEM (Fig. 7.1c). The much noisier region towards the southeast is likely due to a reduced quality of the previous DEM covering Italy. There is also a trend in displacement azimuth (Fig. 7.1e), resulting from a slight deformation (combined rotation + shift) between the two scenes.

Table 7.1: Results of the scene comparison.

#	Scene 1	Scene 2	Displacement magnitude [m]				Displacement azimuth [°]			
			Mean	Stddev.	Median	NMAD	Mean	Stddev.	Median	NMAD
	Swiss Alps									
1	Landsat 8 195/28 30.8.2015 Old processing	Landsat 8: 195/028 30.8.2015 New processing	12.2	3.2	12.6	3.6	268	26	270	8
2	Sentinel-2 29.8.2015 2015 processing	Sentinel-2 29.8.2015 2023 processing	6.1	1.0	6.0	0.7	317	39	321	7
3	Sentinel-2 29.8.2015 2023 processing	Landsat 8 30.8.2015 New processing	11.4	3.2	10.9	1.7	303	92	331	16
	New Zealand									
4	Landsat 7 75/90 13.4.2000 Old processing	Landsat 7 75/90 13.4.2000 New processing	19.8	2.8	20.3	2.6	231	6.3	231	6.8
	Patagonia									
5	Landsat 8 232/93 12.3.2016 Old processing	Landsat 8 232/93 12.3.2016 New processing	21.4	4.6	21.2	4.0	262	16	267	5

The right panel in Fig. 7.1 for Sentinel-2 does not show any trend surfaces or rotations, but rather noisy changes over steep terrain and pronounced shifts over selected glacier tongues, likely due to the fact that their surfaces have substantially lowered and a different (more recent) DEM has been used for the orthorectification (cf. Kääb et al. 2016). One can also see thin stripes that likely result from the misalignment/overlap of individual detector panels.

Example scenes 4 and 5 from New Zealand and Patagonia are depicted in Fig. 7.2 (left and right panel, respectively). With about 20 m (Table 7.1) the shifts are about twice as large as for the Alps but the direction of the shift (west) is about the same. Interestingly, the standard deviation is much smaller than for the Alps, indicating a much more directional shift. Both scenes have a trend surface of the displacement magnitude and azimuth, indicating a deformation (combined shift + rotation). The trend surface in displacement magnitude for New Zealand looks a bit spherical, similar to a correction resulting from geoid to ellipsoid transformation. However, his scenes contains a significant portion of open sea, enabling control points or co-registration during the processing only for parts of the scene. This uneven distribution of control can also produce displacement patterns like the ones observed. Moreover, regions in shadow have systematically larger shifts (brighter orange areas in the purple to pink region). These are most likely not resulting from the reprocessing of the scenes, but by the difficulties in image matching over low-contrast parts in images with (only) 8-bit radiometric resolution.

Finally, we compare scene pair 3 (Sentinel-2 with Landsat 8) in Fig. 7.3, revealing no systematic shifts, but considerable noise. This is also expressed in the high standard deviation of displacement azimuth. These are likely a result of problems with the image matching procedure, as the two scenes are from different sensors / spatial resolutions / spectral bands and acquired at one day apart , for instance including different cover of (small) clouds. Glacier outlines derived from reprocessed Landsat data should thus now have a much better spatial agreement.

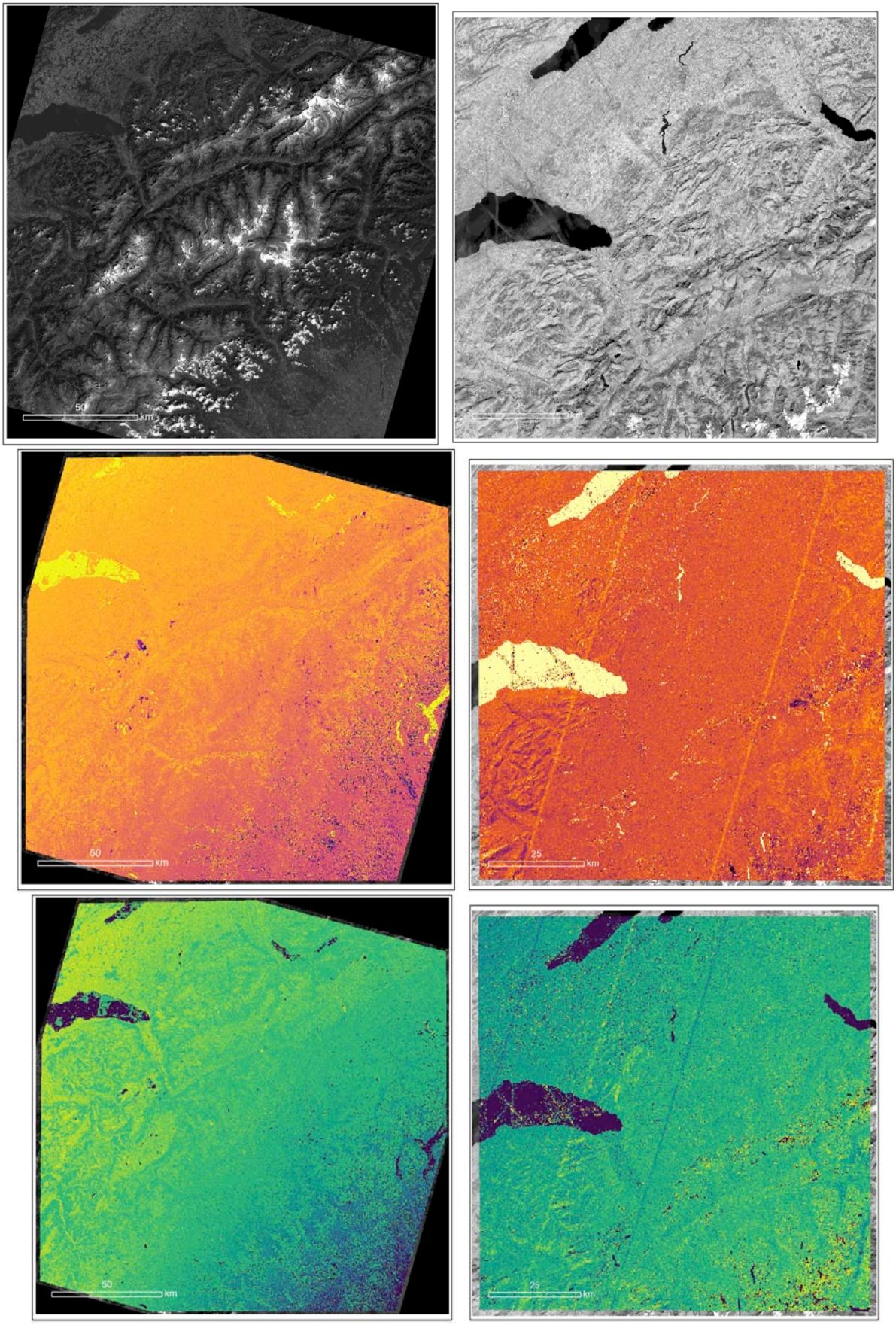

Fig. 7.1: Scene 1 from Landsat 8 (left panels) and scene 2 from Sentinel-2 (right). Upper row: pan band and NIR band, middle: displacement magnitude, bottom: displacement azimuth.

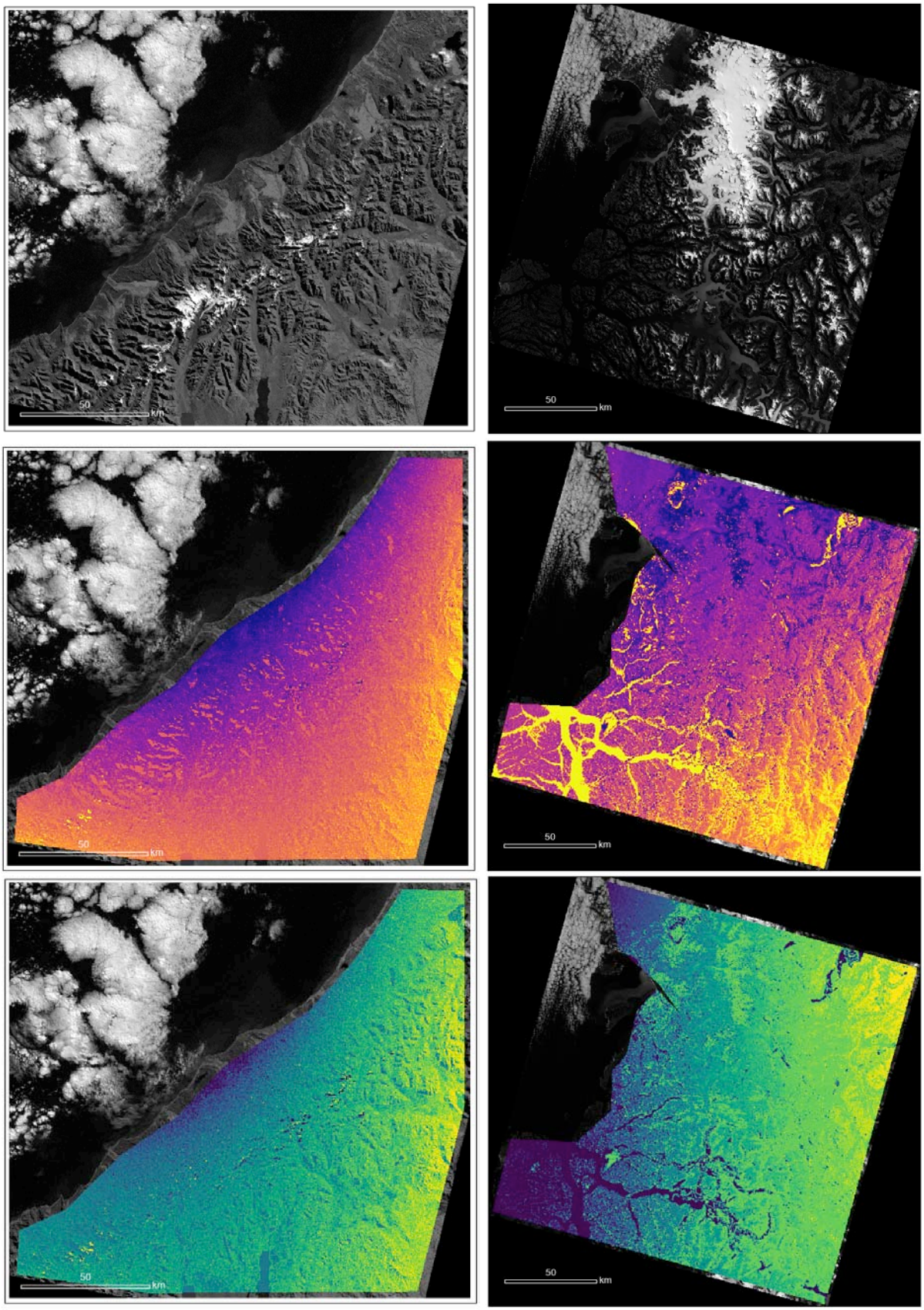

Fig. 7.2: Scene 4 from Landsat 7 (left panels) and scene 5 from Landsat 8 (right). Upper row: pan band and NIR band, middle: displacement magnitude, bottom: displacement azimuth.

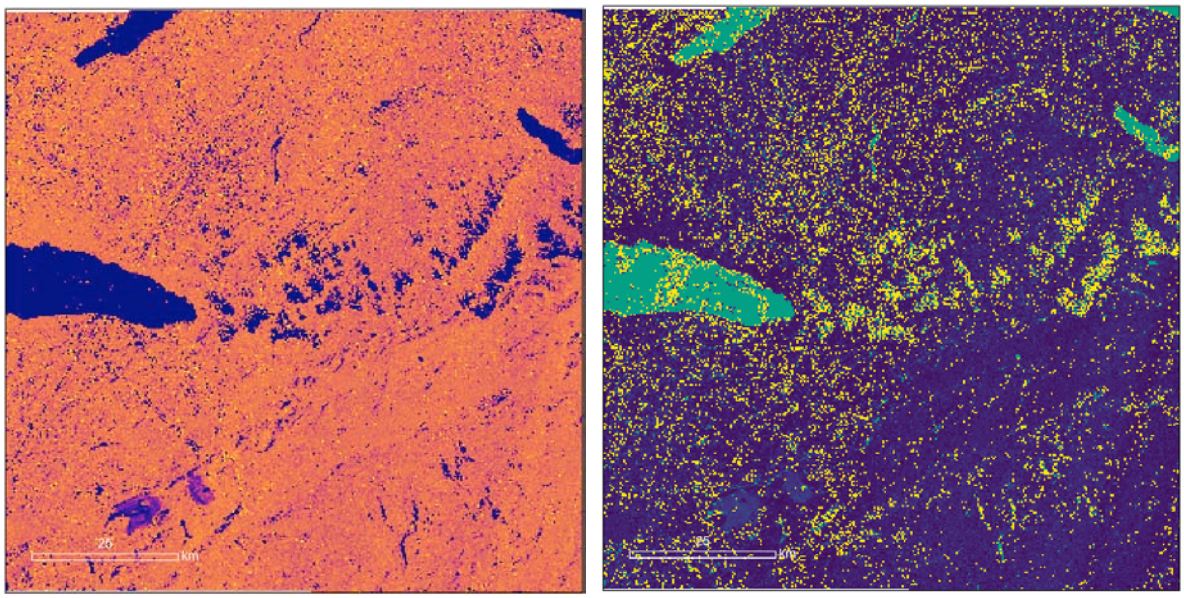

Fig. 7.3: Scene 3 comparing Sentinel-2 with Landsat 8, both as new versions. The spatial subset is as in Fig. 7.1 (right panel). a) Displacement magnitude, b) displacement azimuth.

7.4. Conclusions and Recommendations

The reprocessing of the Landsat data has obviously largely reduced the systematic shift between Landsat and Sentinel-2 scenes. This helps in their synergetic use, i.e. to reduce the time period between scene acquisitions in regions with frequent clouds (e.g. to better follow the evolution of snow cover on a glacier). Direct merging of the two datasets, e.g. to derive glacier flow velocities) has to be investigated closer, though.

As the manual correction of glacier outlines is a very time consuming task, it is not expected that already existing glacier outlines will be created again from the now shifted historic Landsat scenes. For the time being, we have thus to live with the now available slightly wrong outline locations. However, the here-applied image matching method allows the determination of a mean shift vector that can be applied to the region of interest also in the vector domain. At least, this will give a visually better appearance of outline overlays and might even be applied before corrected outlines are digitally intersected with drainage divides from a recent DEM. It might be worth testing if the shift is an issue for a study focussing on multi-temporal analysis.

Whereas the geolocation shift seems solved, it is yet unclear whether USGS/NASA and CNRS/ESA are using the same (freely available) DEM for orthorectification. On a global scale, only 30 m resolution DEMs are available and these are not sufficient for orthorectification of 10 m Sentinel data in steep mountain terrain. Moreover, rapidly lowering glacier surfaces demand to use a DEM that has been acquired close to (a few years) the acquisition date of the satellite image. Such frequently updated global DEMs are currently not publicly available and – despite its urgency – related satellite missions are not planned. Unfortunately, geolocation shifts due to inappropriate DEMs will remain an issue in mountain regions.

Finally, it has to be noted that above results, conclusions and recommendations are based on a few test scenes. Our findings are thus indications of errors to be expected, but our sample is by far not large enough to provide statistically sound and general findings that would be representative for many/all glacier regions on Earth.

References

Bhambri, R., Hewitt, K., Kawishwar, P. and Pratap, B. (2017). Surge-type and surge-modified glaciers in the Karakoram. Scientific Reports 7(1), 15391; doi.org/10.1038/s41598-017-15473-8

ESA (2019). CryoSat-2 Product handbook, Baseline D1.1, https://earth.esa.int/documents/10174/125272/CryoSat-Baseline-D-Product-Handbook

ESA (2022). https://spacedata.copernicus.eu/collections/copernicus-digital-elevation-model

Farnsworth, W.R., Ingolfsson, O., Retelle, M. and Schomacker, A. (2016). Over 400 previously undocumented Svalbard surge-type glaciers identified. Geomorphology 264, 52–60; doi.org/ 10.1016/j.geomorph.2016.03.025

GCOS, 2022. The 2022 GCOS Implementation Plan. Geneva: World Meteorological Organization, 85. https://gcos.wmo.int/en/publications/gcos-implementation-plan2022

Goerlich, F., Bolch, T. and Paul, F. (2020). More dynamic than expected: an updated survey of surging glaciers in the Pamir. Earth System Science Data 12(4), 3161–3176; doi.org/10.5194/essd-12-3161-2020

Guan, W.J. and 7 others (2022). Updated surge-type glacier inventory in the West Kunlun mountains, Tibetan Plateau, and implications for glacier change. Journal of Geophysical Research-Earth Surface 127(1), Artn e2021JF006369; doi.org/10.1029/2021JF006369

Guillet, G. and 6 others (2022). A regionally resolved inventory of High Mountain Asia surge-type glaciers, derived from a multi-factor remote sensing approach. Cryosphere 16(2), 603–623; doi.org/10.5194/tc-16-603-2022

Guo, L. and 5 others (2022). A new inventory of High Mountain Asia surging glaciers derived from multiple elevation datasets since the 1970s, ESSD, 15, 2841–2861; doi.org/10.5194/essd-15-2841-2023

Horton, A., Ewart, M., Gourmelen, N., Fettweis, X. and Storkey, A. (2022). Using deep learning to model elevation differences between radar and laser altimetry. Remote Sensing, 14(24), 6210; doi.org/10.3390/rs14246210

Hugonnet, R. and 10 others (2021). Accelerated global glacier mass loss in the early twenty-first century. Nature 592(7856), 726-731; doi.org/10.1038/s41586-021-03436-z

Jiskoot, H., Murray, T. and Luckman, A. (2003). Surge potential and drainage-basin characteristics in East Greenland. Annals of Glaciology 36, 142–148; doi.org/10.3189/172756403781816220

Kääb A. and Vollmer M. (2000). Surface geometry, thickness changes and flow fields on creeping mountain permafrost: automatic extraction by digital image analysis. Permafrost and Periglacial Processes, 11, 315-326.

Kääb, A., Winsvold, S.H., Altena, B., Nuth, C., Nagler, T. and Wuite, J. (2016). Glacier Remote Sensing Using Sentinel-2. Part I: Radiometric and geometric performance, and application to ice velocity. Remote Sensing, 8(7), 598; doi.org/10.3390/rs8070598

Kääb, A., Bazilova, V., Leclercq, P.W., Mannerfelt, E.S., Strozzi, T. (2023). Global clustering of recent glacier surges from radar backscatter data, 2017–2022. Journal of Glaciology. 69 (277), 1515-1523; doi.org/10.1017/jog.2023.35

Khan, S.A., Colgan, W., Neumann, T.A., Van Den Broeke, M.R., Brunt, K.M., Noël, B., Bamber, J.L., Hassan, J. and Bjørk, A.A. (2022a). Accelerating ice loss from peripheral glaciers in North Greenland. Geophysical Research Letters, 49(12). doi.org/10.1029/2022GL098915

Khan, S. A. et al. (2022b). Accelerating ice loss from peripheral glaciers in North Greenland, Dryad, Dataset; doi.org/10.5061/dryad.zpc866tb8

Leclercq, P.W., Kääb, A. and Altena, B. (2021). Brief communication: Detection of glacier surge activity using cloud computing of Sentinel-1 radar data. The Cryosphere, 15, 4901–4907.

Paul, F. (2015). Revealing glacier flow and surge dynamics from animated satellite image sequences: examples from the Karakoram. The Cryosphere, 9, 2201-2214.

Paul, F., Rastner, P., Azzoni, R.S., Diolaiuti, G., Fugazza, D., Le Bris, R., Nemec, J., Rabatel, A., Ramusovic, M., Schwaizer, G., and Smiraglia, C. (2020). Glacier shrinkage in the Alps continues unabated as revealed by a new glacier inventory from Sentinel-2. Earth Systems Science Data, 12(3), 1805-1821.

Paul, F., Baumann, S., Anderson, B. and Rastner, P. (2023). Deriving a year 2000 glacier inventory for New Zealand from the existing 2016 inventory. Annals of Glaciology, 90; doi.org/10.1017/aog.2023.20Polar Geospatial Center, University of Minnesota (2023). https://www.pgc.umn.edu/data/arcticdem/

Remy, F., Legresy, B. and Benveniste, J. (2006). On the azimuthally anisotropy effects of polarization for altimetric measurements. IEEE Transactions on Geoscience and Remote Sensing, 44(11), 3289–3296; doi.org/10.1109/TGRS.2006.878444

Rengarajan, R., J.C. Storey and M.J. Choate (2020). Harmonizing the Landsat Ground Reference with the Sentinel-2 Global Reference Image Using Space-Based Bundle Adjustment. Remote Sensing, 12, 19: 3132; doi.org/10.3390/rs12193132

RGI Consortium (2023). The Randolph Glacier Inventory version 7.0 User guide v1.0, 2023. doi.org/10.5281/zenodo.8362857. Online: doi.org/10.5281/zenodo.8362857

Scagliola, M., Fornari, M. and Bouffard, J. (2016). Theoretical and experimental characterization of the speckle in cryosat l1b waveforms. 740, 320. https://ui.adsabs.harvard.edu/abs/2016ESASP.740E.320S

Schröder, L., Horwath, M., Dietrich, R., Helm, V., van den Broeke, M.R. and Ligtenberg, S. R.M. (2019.Four decades of Antarctic surface elevation changes from multi-mission satellite altimetry. The Cryosphere, 13(2), 427–449; doi.org/10.5194/tc-13-427-2019

Sevestre, H. and Benn, D.I. (2015). Climatic and geometric controls on the global distribution of surge-type glaciers: implications for a unifying model of surging, Journal of Glaciology, 61(226), 646-662.

Storey, J., Roy, D.P., Masek, J., Gascon, F., Dwyer, J., Choate, M. (2016). A note on the temporary misregistration of Landsat-8 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) imagery. Remote Sens. Environ., 186, 121-122; doi.org/10.1016/j.rse.2016.08.025

Xie, F. and 27 others (2023). (2022). Interdecadal glacier inventories in the Karakoram since the 1990s. Earth Syst. Sci. Data, 15, 847–867; doi.org/10.5194/essd-15-847-202

Acronyms

ATLAS	Advanced Topographic Laser Altimetry System
DEM	Digital Elevation Model
eCDF	empirical cumulative distribution function
GLAS	Geoscience Laser Altimeter System
GLIMS	Global Land Ice Measurements from Space
GoLIVE	Global Land Ice Velocity Extraction from Landsat 8
HMA	High Mountain Asia
ICESat	Ice, Cloud, and land Elevation Satellite
ID	Identifier
InSAR	Interferometric SAR
ITS LIVE	Inter-Mission Time Series of Land Ice Velocity and Elevation
NSIDC	National Snow and Ice Data Center
QA	Quality Assessment
RADAR	Radio Detection and Ranging
RGI	Randolph Glacier Inventory
RMSE	Root Mean Square Error
SAR	Synthetic Aperture Radar
SEC	Surface elevation change
STD	Standard Deviation
USGS	United States Geological Survey