Product Validation and Algorithm Selection Report (PVASR)

Prepared by: Glaciers_cci consortium
Contract: 4000101778/10/I-AM
Name: Glaciers_cci-D2.5-PVASR
Version: 0.8
Date: 20.07. 2012

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

Technical Officer:
Stephen Plummer
ESA ESRIN

Document status sheet

Version	Date	Changes	Approval
0.1	18.12. 2012	Draft ToC	
0.2	08.02. 2012	Extended ToC for area	
0.3	22.02. 2012	Extended ToC for EC (DEM) / VEL (OPT)	
0.4	12.04. 2012	Area section written, first input from other products	
0.5	21.05. 2012	Further sections added (altimetry, velocity from microwave)	
0.6	12.06. 2012	Feedback from technical officer included	
0.7	14.07. 2012	Extension from SEEL included	
0.8	20.07. 2012	Final edits included	

The work described in this report was done under ESA contract 4000101778/10/I-AM. Responsibility for the contents resides in the authors that prepared it.

Author team:
Frank Paul (GIUZ), Christopher Nuth, Torbog Heid, Andreas Kääb (GUIO), Thomas Nagler, Helmut Rott, Killian Scharrer (Enveo), Andrew Shepherd, Francesca Ticconi (SEEL), Tazio Strozzi, Andreas Wiesmann (Gamma)

Glaciers_cci Technical Officer at ESA:
Stephen Plummer

Table of Contents

1. Overview 4

2. Glacier area 5
2.1 Assessment of algorithm performance 5
2.2 Methods for accuracy determination 7
2.3 Quantitative measures for accuracy assessment 8
2.4 Analysis of the multiple digitizations 10
2.5 Comparison of the applied algorithms 20
2.6 Glacier mapping accuracy and algorithm selection 25

3. Elevation changes from altimetry 26
3.1 Overview of the round robin experiments 26
3.2 Assessment of algorithm performance 28
3.3 Analysis and comparison of the round robin results 28
3.4 Selection criteria 48

4. Elevation changes from DEM differencing 51
4.1 Assessment of algorithm performance 51
4.2 Methods for accuracy determination 52
4.3 Comparison of the applied algorithms 53
4.4 Statistical analysis of the results 56
4.5 Discussion of the investigated algorithms 63
4.6 Algorithm selection 64

5. Glacier velocity - optical 67
5.1 Context of the algorithm development and product validation 67
5.2 Methods for accuracy determination 70
5.3 Comparison of the applied algorithms 72
5.4 Discussion of the investigated algorithms 82
5.5 Algorithm selection 86

6. Glacier velocity from microwave sensors 89
6.1 Context of the algorithms and accuracy determination 89
6.2 Overview of the round robin experiments 89
6.3 Intercomparison of displacement fields from SAR data 90
6.4 Statistical analysis of the results 100
6.5 Discussion of the investigated algorithms and algorithm selection 106

References 108

Abbreviations 112

1. Overview

This is the Product Validation and Algorithm Selection Report (PVASR) of the Glaciers_cci project. It describes the analysis done in the round robin inter-comparison, discusses the results achieved and the algorithm selections made. In particular it contains:

- a description of the protocols, methods, reference datasets and tools used to scientifically assess and inter-compare the performance of the prototype algorithms and to validate the resulting products
- a comparison of the results from the various algorithmic approaches in terms of geophysical consistency and accuracy,
- a statistical analysis of the retrieval errors,
- a discussion of the similarities, differences, problems and limitations of the tested algorithms
- identification of all the algorithm(s) to be implemented for all relevant steps and contributing missions.

As the Glaciers_cci project is concerned with four largely different products (glacier area, elevation change from DEM differencing and repeat altimetry, and velocity), it was required to set-up and perform four round robins for each of the products. Due to available knowledge from previous round robins on glacier mapping, we started with the glacier area product and used this to guide the related steps for the other products. Due to the differences in the products, computational methods, and user communities, the results reported here have a slightly different set-up for each of the products. However, all round robins were successful in providing numerous lessons learned and criteria for selecting a most appropriate method for product generation.

For each of the products we provide in the beginning a background on data generation and accuracy assessment, including measures for accuracy quantification. These are required to better understand the set-up of the individual round robins, the results obtained and the statistical analysis performed. The experiments confirmed that determination of a ‘best’ algorithm is not always straight forward, either due to inappropriate reference data or due to a purpose-dependent interpretation of what is accurate and what not. In particular qualitative measures (e.g. overlay of glacier outlines from multiple digitizations) provided important information on methodological issues that need to be addressed in future work.

2. Glacier area

2.1 Assessment of algorithm performance

2.1.1 Impact of the algorithm used for glacier classification

A wide range of methods was and still is applied to map glaciers from optical images (i.e. classify snow and ice). They mostly differ in complexity, pre-processing demands, required input bands and degree of automation, but not so much in the classification result. A review of the most often applied methods is given in section 3.3 of the ATBDv0 (Glaciers_cci, 2012b). With a focus on the most suitable optical sensors, the methods are largely independent of the sensor used, as the spectral bands cover very similar spectral ranges (see 3.1 in the DARD, Glaciers_cci, 2011). We thus refer in the following to different spectral bands rather than sensors. From the proposed algorithms we exclude here manual delineation as this is an important step during post-processing and also required for generating a reference dataset. We also exclude algorithms that were already considered as being less suitable or less accurate in previous studies such as all (scene-dependent) supervised (e.g. Maximum-Likelihood and principal component analysis) and unsupervised (e.g. ISODATA clustering) classification methods, as well as those which require atmospheric and topographic correction (Albert, 2002; Paul et al., 2003). The focus is thus here on the two most often applied methods, simple band ratios (e.g. Paul et al., 2002) and the Normalized Difference Snow Index (NDSI) (e.g. Dozier et al. 1989; Racoviteanu et al., 2008). Past studies have already shown that both methods differ only at the level of individual pixels, with errors occurring in different regions of a glacier, but at about the same quantity (Paul and Kääb, 2005).

What remains to be tested is the influence of the manually selected threshold for creating the binary glacier map from the two ratios (e.g. Andreassen et al., 2008; Svoboda and Paul, 2009). A recent study by Gjermundsen et al. (2011) and earlier studies (e.g. Paul, 2002; Paul and Hendriks, 2010) have shown that the changes in glacier size are indeed small for changes in the threshold or the application of a noise filter. We thus assume that the use of an automatically calculated or fixed threshold might help to further automate the processing line for the glacier area product without degrading product accuracy too much. This is investigated during the round robin by the Glaciers_cci EO team. However, for the round robin with global participation (e.g. by GLIMS contributors and the CRG), the selection of the algorithm to be applied is free of choice to consider latest developments. It has to be stressed that this part of the assessment of the algorithm accuracy must not include any manual corrections by the analyst. Results need to be compared 'as is' to exclude any operator-related decisions.

2.1.2 External conditions influencing product accuracy

Apart from the applied algorithm for the initial glacier mapping, a wide range of external factors influence product accuracy (e.g. Racoviteanu et al., 2009). This includes adverse snow conditions with seasonal snow hiding a part of the glacier perimeter, local clouds doing the same, regions with haze requiring a different threshold than the clear part of the image (LeBris et al., 2011), or glacier parts in shadow that cannot be mapped due to missing contrast in the respective spectral bands (Paul et al., 2011a). The errors for the final product that can be introduced by these factors are about one to two orders of magnitude larger than those resulting from using a different threshold for the band ratio. Hence, only cloud-free images from the end of the ablation period in a year without snow outside of glaciers should be used

to map glaciers. Though this restriction reduces the number of suitable scenes for glacier mapping in some regions of the world considerably, it is a mandatory requirement for product accuracy. In part, these restrictions can be avoided by including additional scenes in the analysis, but this can be a challenge for the database when an individual glacier has to be mosaiced from two or more satellite scenes. Hence, exchanging a scene when a better one becomes available is always preferable (Le Bris et al., 2011).

The mapping of glaciers in cast shadow is part of the round robin as this is a common feature in mountain terrain and has to be adequately solved by the initial (bare-ice) mapping algorithm. Actually, the recommendation is to select the threshold value for the band ratios in a way to optimally map glacier ice in cast shadow and hence reduce the workload for post-processing (e.g. Paul et al., 2002; Bishop et al., 2004; Raup et al. 2007). However, as manual editing in shadow regions is frequently required, we will also include glaciers in shadow in the multiple digitizing experiments (see 2.4.1). The other two factors (snow/haze) will not be investigated in the round robin to keep the effort manageable.

2.1.3 Post-classification issues

After a raw glacier map has been created (GL0a product), post-processing is required to remove gross errors (e.g. wrongly classified lakes, missing debris cover, local clouds) and edit other misclassification (e.g. ice bergs, shadow) to generate a GL0b product. In general, this is done by visual comparison with a contrast-enhanced version of the satellite image used. From this 'glacier cover only' product a higher-level product (GL1) can be derived, the individual glacier entities. This step requires a co-registered DEM to derive drainage divides and digitally intersect them with the GL0b outlines. While this is in general straightforward for alpine glaciers surrounded by steep valley walls, it can be challenging for ice fields or ice caps (Racoviteanu et al., 2009). In particular the division of ice caps into hydrologic catchments does not make much sense in a glaciological context. In this regard the drainage divide issue has to be discussed from a methodological point of view rather than a technical one and will not be investigated in the round robin.

Manually removing wrongly classified water bodies is easy, as often a strong spectral contrast is found between water and ice. However, when the water surface is frozen or a largely dissected glacier calves into water with lots of icebergs close to the front, the issue is more challenging and requires some experience. Whereas clear water can be mapped automatically and removed (e.g. Huggel et al., 2002), turbid water often remains and needs manual editing (Paul and Kääb, 2005; Gjermundsen et al., 2011). We will thus include water surfaces in contact to glaciers in the round robin.

The correct automated mapping of debris-covered glaciers is still not possible (e.g. Shukla et al., 2011) and the available semi-automated methods (e.g. Paul et al., 2004; Bolch et al. 2007) also require careful manual editing. As debris can cover more than 50% of a glacier tongue and is often difficult to identify in low-contrast (i.e. high elevation of the sun) optical images, wrongly mapped debris cover is actually the single most important factor influencing product accuracy. This step has thus to be done with great care to meet the accuracy specifications for the glacier area product (better than 5%). A comparison of the interpretation of debris cover by different analysts is thus part of the round robin. Overlay of glacier outlines will reveal the critical regions of interpretation.

2.1.4 Multi-temporal considerations

Further important aspects of product accuracy have to be considered when multi-temporal analysis is performed or when different datasets are combined. The most important one is the accuracy of the geolocation. As previous studies have shown (e.g. GlobGlacier) only orthorectified satellite images can be used for product generation. Such a product is meanwhile provided by USGS for all Landsat scenes (called 'L1T' for terrain corrected), with a geolocation uncertainty of about 1 image pixel or less (RMSE). Though this is acceptable for the global glacier area product, a more detailed analysis of the geolocation error (available from the metadata of the respective satellite scene) reveals much higher values in steep high-mountain topography or in regions where the used DEM has artifacts (Frey et al., in revision). For the regions with voids in the SRTM DEM, shifts of about 5 pixels (150 m) or more were found in that study compared to an independent dataset (the ASTER GDEM). Such a shift causes also problems for deriving drainage divides, topographic parameters and digital overlay with other orthorectified satellite images when their correction is based on a different DEM. As the L1T orthorectification of the Landsat scenes by USGS is an operational process, there is not much Glaciers_cci can do about it. On the other hand, the processing at USGS is continuously improved and hence also better DEMs (e.g. GDEM2) might be considered in the future for orthorectification.

When all scenes used for change assessment are orthorectified with the same DEM, a potential error in the geolocation does not matter. This becomes only an issue when the multi-temporal analysis combines data from different sources (e.g. from different sensors or glacier outlines digitized from maps). In this regard the proper transformation of coordinates from one projection to another is an important issue to consider. When details of the used ellipsoid/datum are only poorly known or implemented in the software used, unsystematic shifts between two datasets can occur that make a direct comparison challenging. However, differences in the interpretation of glaciers by cartographers might be even more severe and have to be considered with the respective care (Bolch et al., 2010; Paul and Andreassen, 2009). For this reason we will not use vector data as available from independent sources (e.g. national mapping agencies) for product validation. Of course, changes in glacier extent can be due to a differing interpretation of details (debris cover, snow) by different analysts and we are mostly interested in these. They will thus form an important part of the round robin.

2.2 Methods for accuracy determination

2.2.1 Product validation using reference data

There are basically two different measures to assess product accuracy, one is validation with so-called 'ground-truth' or better 'reference' data and the other one is a relative comparison of results from different algorithms, analysts etc. (see 2.2.2). In regard to reference data, the major problem is that they do seldom exist (depending on the criteria defining 'reference') and that the final product includes in most cases a manual correction (e.g. for debris-cover) that is obtained by correction against a 'ground-truth' (the satellite image itself). To circumvent these problems, there are two options:
(1) using data that have been independently acquired at the same date (week), for example from GPS ground surveys or from high-resolution (1 m or better) aerial photography or satellite imagery, and

(2) a full manual digitization of the glacier extent without considering the result of the automated methods (band ratio, NDSI).

When (1) is available for an entire glacier, two kinds of validation are possible:
(i) comparison of the total area and
(ii) analysis of the omission and commission errors (cf. Gjermundsen et al., 2011).

When only parts of a glacier are covered, the digital overlay of the respective vector outlines can still be used for a qualitative statement about the agreement, but little can be said in absolute terms. In most cases differences in the interpretation of details (e.g. debris cover at the terminus) will drive the differences rather than shortcomings in the automated mapping.

The latter is the reason to use the same satellite image for a full manual digitization. Such a vector line is at least independent of resolution and interpretation differences (Paul et al., 2003). Of course, for such a comparison only debris-free glaciers can be used. When this is done for a couple of glaciers with different sizes, the differences between (i) and (ii) can be calculated and statistically analyzed. For the round robin we will utilize both ways of assessing product accuracy, the comparison with outlines derived from a higher resolution dataset (done by the EO team) and the full manual digitization of glacier extents (done by the CRG and the invited global community).

2.2.2 Relative comparisons

The second way to determine product accuracy is a relative one without considering a reference dataset. This includes points (i) and (ii) from 2.2.1 for the glacier extents resulting from (a) different algorithms as well as (b) multiple digitizations of the same glacier. Whereas for (a) the overlay of grids is most suitable to visualize the differences of algorithms, the overlay of vector outlines is more suitable for (b). A third kind of comparison (c) results from the round robin: different analysts map the same glacier (type many-to-one). This will be more suitable to reveal differences in the interpretation rather than for calculating absolute differences. The last comparison is also important to improve the consistency of the glacier outlines as available from the GLIMS database.

2.3 Quantitative measures for accuracy assessment

The measures to assess the accuracy of the glacier outline product can be distinguished into qualitative and quantitative ones. The former describe the differences observed for an overlay of outlines from different sources, analysts or multiple digitizings. They help to learn where methodological differences in image interpretation occur, for example in regard to the interpretation of glacier forefields, tributaries, debris-cover, ice in shadow, disintegrating and calving glaciers, position of the glacier terminus, etc. Once these issues are resolved and documented in illustrated guidelines, quantitative measures can be applied to assess product accuracy. They focus on the direct calculation of differences in glacier area to a reference dataset and can be appended by mean values and standard deviation for larger samples (scalar metrics). When the results of glacier mapping differ only locally, the comparison of omission and commission errors (visually and quantitatively) is a valuable measure to quantify product accuracy (raster metrics). This is required as the same area of a glacier can be obtained by two digitizations (indicating perfect agreement), but the regions considered for the total glacier area are completely different (e.g. missing debris cover is compensated by an additional

tributary). In such a case the area difference alone has little meaning. A further quantitative assessment of the error can be applied when multiple outlines are available for the same glacier by calculating the mean distances of the respective segments (vector metrics). These can be illustrated in box plots showing mean, median, standard deviation and percentiles in comparison to a reference dataset. In all cases it is required to also illustrate the outlines or raster maps with overlays to allow a meaningful interpretation. In Table 2.1 we provide an overview on the accuracy assessments that can be performed and the details to assess them. Apart from the distance of outlines and impact of noise filters, all comparisons have been performed in the round robin.

The calculation of the quantitative accuracy measures is based on calculations of the glacier area as implemented in the GIS with subsequent statistical anlysis of the derived values (e.g. mean, standard deviation). Omission and commission errors can be calculated after subtraction of a grid based representation of the selected glaciers, but this has not been done here as the location of the differences were too irregular. The in-depth analysis of the results compares mean values and standard deviations from the different datasets.

Comparison	Calculation	Statistics	Measure	Metrics
Satellite vs. 'reference data'	relative difference	mean, std. deviation	absolute	scalar
Area from multiple digitizings	variability	mean, std. deviation	relative	scalar
Overlay of outlines	visual interpretation	none	qualitative	-
Distance of outlines	variability	overlay	qualitative	scalar
Comparison of algorithms	relative difference	omission/commission	relative	raster
Area change by threshold	relative difference	none	absolute	scalar
Impact of noise filter	pixel count	omission/commission	relative	raster

Table 2.1: Overview of the different possibilities to assess product accuracy for glacier area.

Based on these possibilities, four test regions were selected for the glacier area round robin as described in Chapter 5 of the product validation plan (Glaciers_cci, 2012a). Three of them (GA-1, GA-2, GA3) were selected for manual digitizing experiments and one (GA-4) for algorithm comparison. The selection considers remote sensing data with different spatial resolution and the typical challenges like debris, shadow, and seasonal snow. In Table 2.2 the main characteristics of each test region is listed, further details are provided in section 2.4.

Region	Source data	Resolution	Task	Comment
GA1- Alaska	Quickbird	1.4 m	1x manual digitizing	8 glaciers
GA2- Alps (Otztal)	Landsat TM Ikonos	30 m 1.4 m	3x manual digitizing 1x manual digitizing	10 glaciers only for validation
GA3 - Alps (Switzerland)	Aerial Photogr.	0.6 m	1x manual digitizing	3 glaciers
GA4 - Himalaya	Landsat ETM+	30 m	automated mapping and manual editing of a subregion	raw image bands provided

Table 2.2: Overview of the test regions analysed in the glacier area round robin.

2.4 Analysis of the multiple digitizations

2.4.1 Overlay of outlines

In three of the four test regions glacier outlines were obtained from multiple digitizations, either by the same or by different analysts. The related variability is a relative measure of the accuracy and provides mean values and standard deviation for each glacier in a scalar format (see Table 2.1). A major point here is that we use the mean area value from all digtizations of a specific glacier as a reference value for later validation of other datasets (e.g. derived by automated methods). The resulting difference can be seen as an absolute measure of the accuracy, whereas the overlay of outlines illustrates regions with a difficult interpretation and thus helps to create guidelines supporting a more consistent digitizing. In the following we present the results (area differences and outline overlay) for the three test regions and the individual glaciers. Statistical data are summarized for all glaciers in Table 2.3.

In Fig. 2.1 an overview of the results for test region **GA1** (Alaska) is shown. On this high-resolution (2.3 m) image from the Quickbird satellite the participants were asked to digitize each of the marked 8 glaciers once. Challenges were regions in shadow, debris-covered glacier tongues and seasonal snow. In general, the participants had little problem with the bare ice and only few had problems with the region in shadow for glaciers 2 and 3. Either the regions in shadow were disregarded, or bare rock in shadow was included. These are typical problems also for much lower resolution images and are thus not surprising. When the contrast is simply too low for a clear identification, the accuracy of the product is reduced. The debris-cover was more tricky. The interpretation of the debris-cover on the tongues of the larger glaciers 6 and 8 differed widely and also bare rock at the terminus of glaciers 2 and 7 was interpreted as being part of the glacier. On the other hand, the debris on the surface between glaciers 2 and 3 (see close-up in Fig. 2.2) or on glacier 5 was not detected. This supports the thesis that higher spatial resolution alone must not necessarily result in a more accurate delineation of the outline. However, also the interpretation plays a role.

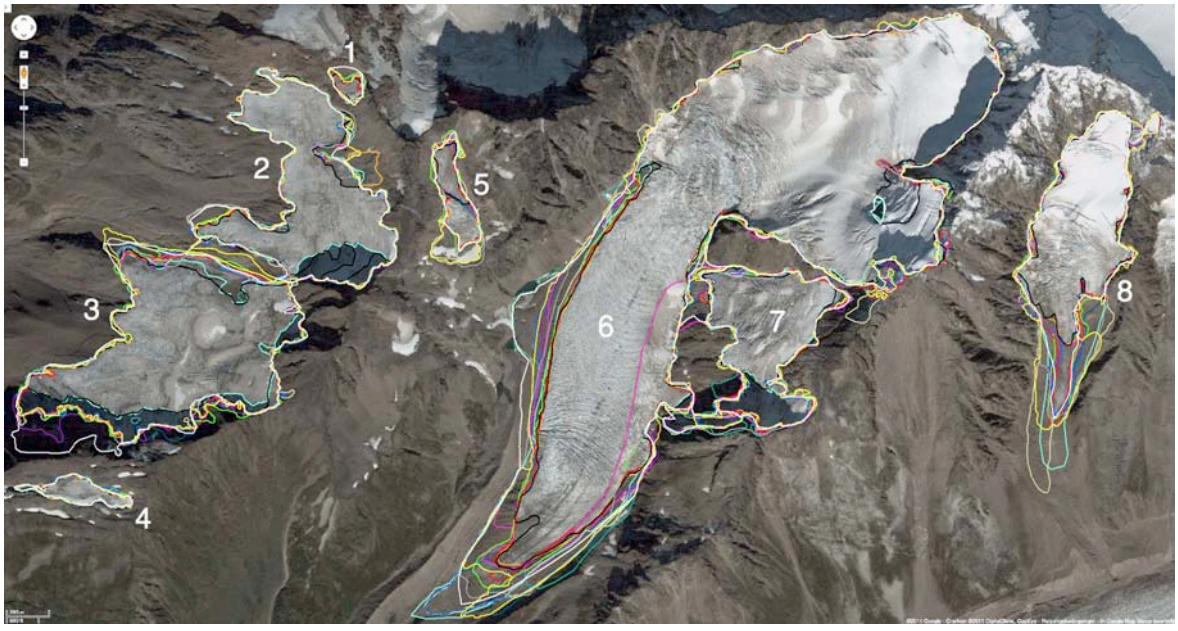

Fig. 2.1: Results for the eight glaciers from GA1. Image: screenshot from Google maps.

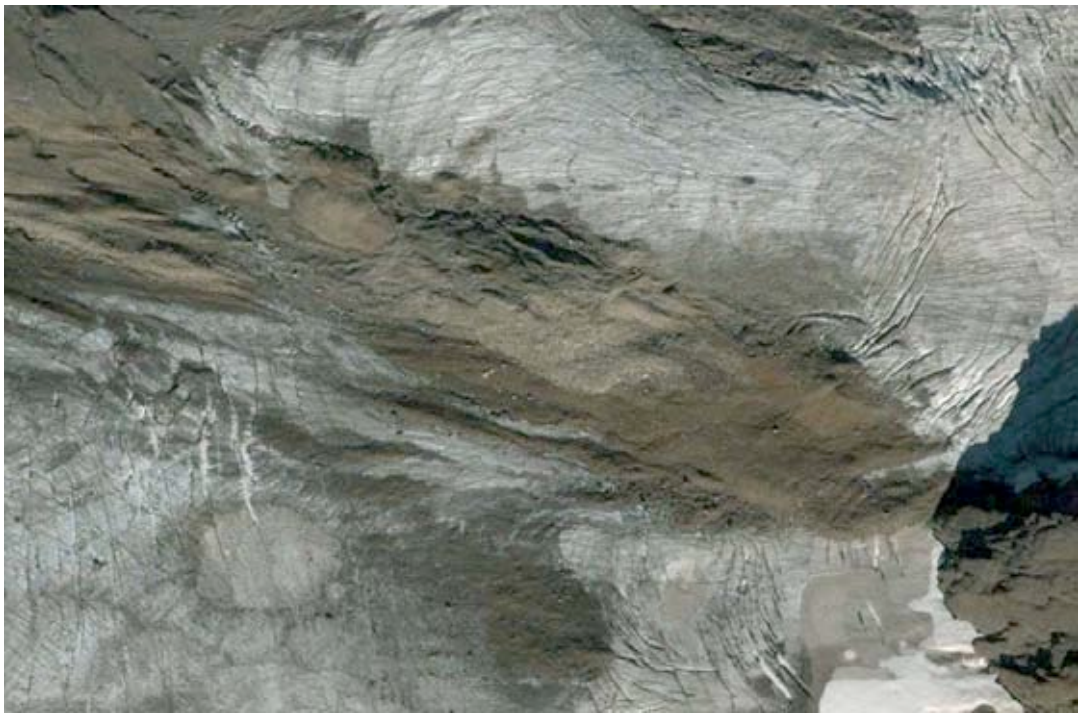

Fig. 2.2: Debris cover on the glacier surface between glaciers 2 and 3. This is only visible at maximum resolution. Image: screenshot from Google Maps, North is at top.

The debris cover between the marked glaciers 2 and 3 was only identifiable as such with the full resolution image from Google Earth (Fig. 2.2). It was not visible at the provided image resolution. Accordingly, the participants interpreted it as a natural separation (rock outcrop) between the two glaciers. As the comparison of the outlines on Fig. 2.1 with the close-up in Fig. 2.3 illustrates, the medial moraine of glacier 5 was sometimes interpreted as the glacier terminus and the region completely under debris and covered by snow patches) was partly included and partly not. Overall, these differences in interpretation resulted in a standard deviation of the derived areas larger than 30% (see Table 2.3). It was hence more a methodological issue whether to include or exclude these regions in the outline.

The length differences resulting from the interpretation of the terminus positions are 500 m for glacier 7 and 600-700 m for glacier 8. Of course, it was very difficult to spot the exact position as debris-covered dead ice and ice-cored moraines were wide-spread. As glacier 7 is much larger than glacier 5, the variability in interpretation resulted only in a standard deviation of the total area of 13%. As the second close-up in Fig. 2.3 for glacier 8 illustrates, many tongues ended at the big boulder of the bare ice strip, but vary in interpretation of the ice under debris north of it. Hence, determination of length changes is not possible for such glaciers when different analysts provide the outlines from different points in time. A useful analysis is only possible when performed by the same analyst and when supported by the digital outlines and the images used for the previous assessment.

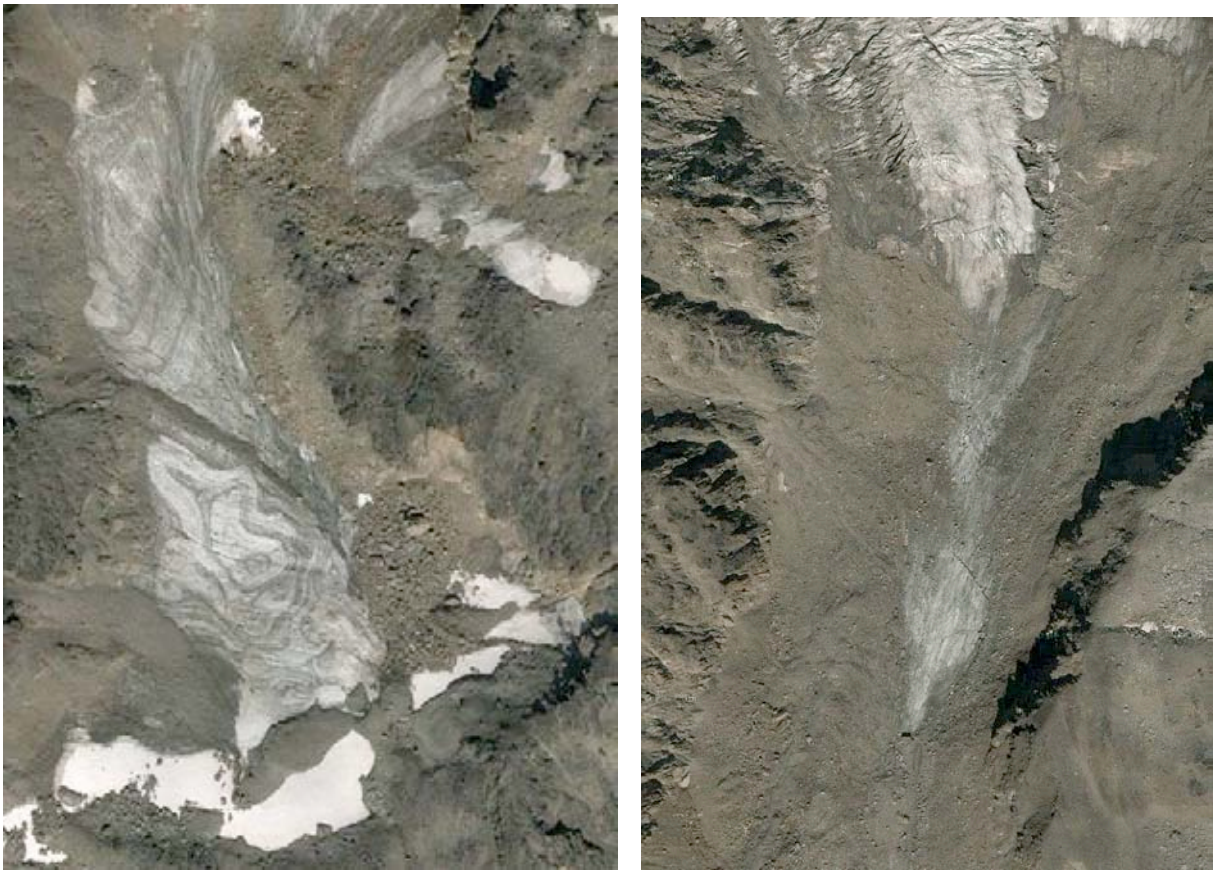

Fig. 2.3: Two close-ups showing glacier 5 (left) and the terminus of glacier 8 (right) from test region GA1. The boundary of the ice can - despite the high spatial resolution - only roughly be estimated. Images: screenshots from Google Maps.

For test region **GA2** (Alps) colour composites of a Landsat TM scene from 2003 were used as a base for manual digitization. The automatically derived outlines were taken from the new glacier inventory for the entire European Alps (Paul et al., 2009). According to the feedback forms, all participants used the band 543 composite as a base for the digitization. This already illustrates that the improved contrast provided by a composite with the SWIR band is highly preferable over the natural colour composite, which had to be used for GA1. Basically, the delineation is conducted by tracing the light to dark bluish colour that marks ice and snow while considering the glaciers geometry. There were not many challenges for most of the glaciers, maybe apart from one with somewhat more debris cover. This guarantees that differences result from the variability in digitization rather than from interpretation challenges. The composite in Fig. 2.4 shows the results for six of the ten glaciers. Close-ups are used to better illustrate the variability. The white (square-shaped) outline is the result from the automated classification with the band ratio method (TM3 / TM5).

The overlay illustrates four important findings: (1) the variability of the outline positions is in general 1-2 pixels, (2) the white TM outline is located mostly within this variability, (3) in regions with a more difficult interpretation (snow patch in Fig. 2.4a, debris-cover in Fig. 2.4c) the outlines show larger differences, and (4) there is some variability in regard to the digitizing level of detail (the number vertices). For the glacier depicted in Fig. 2.4c the standard deviation is about three times higher (15%) than for all other glaciers.

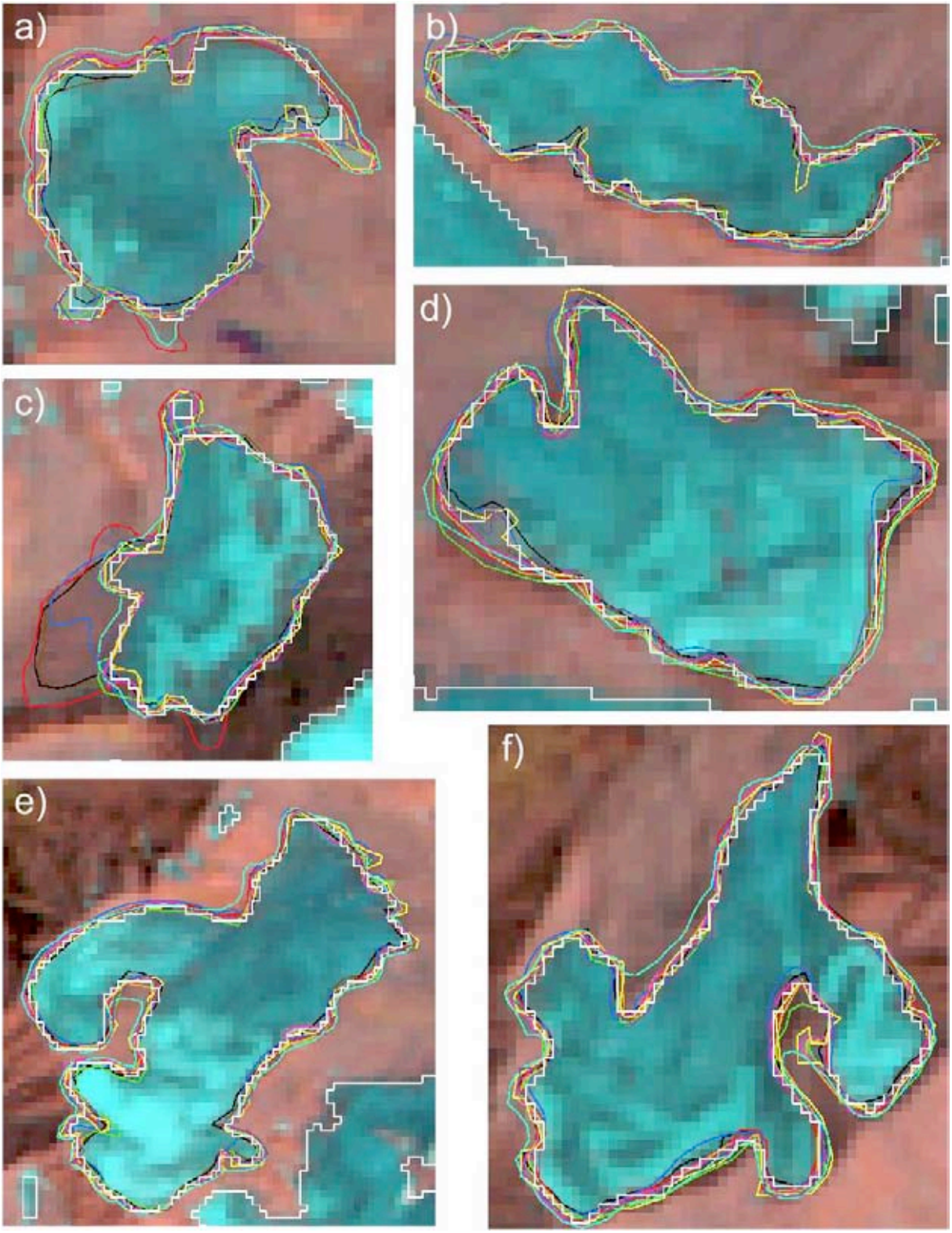

Fig. 2.4: Results from the multiple digitizations of glaciers on a TM scene (bands 543 as RGB) as performed by different analysts. The white and more roughly pixeled lines show the automatically derived glacier extents.

When comparing multiple digitizations of the same glacier by the same analyst, the outlines are more consistent in regard to their position and level of detail. This is also expressed in the standard deviations for individual glacier which are often smaller by a factor of two. However, the maximum standard deviation for a specific glacier by one of the analysts is as high as for the digitizing by different analysts. We thus conclude that multiple digitizations of the same set of glaciers by the same analyst provide a realistic assessment of the analysts accuracy. Selected overlays of the multiple digitizations are shown in Fig. 2.5.

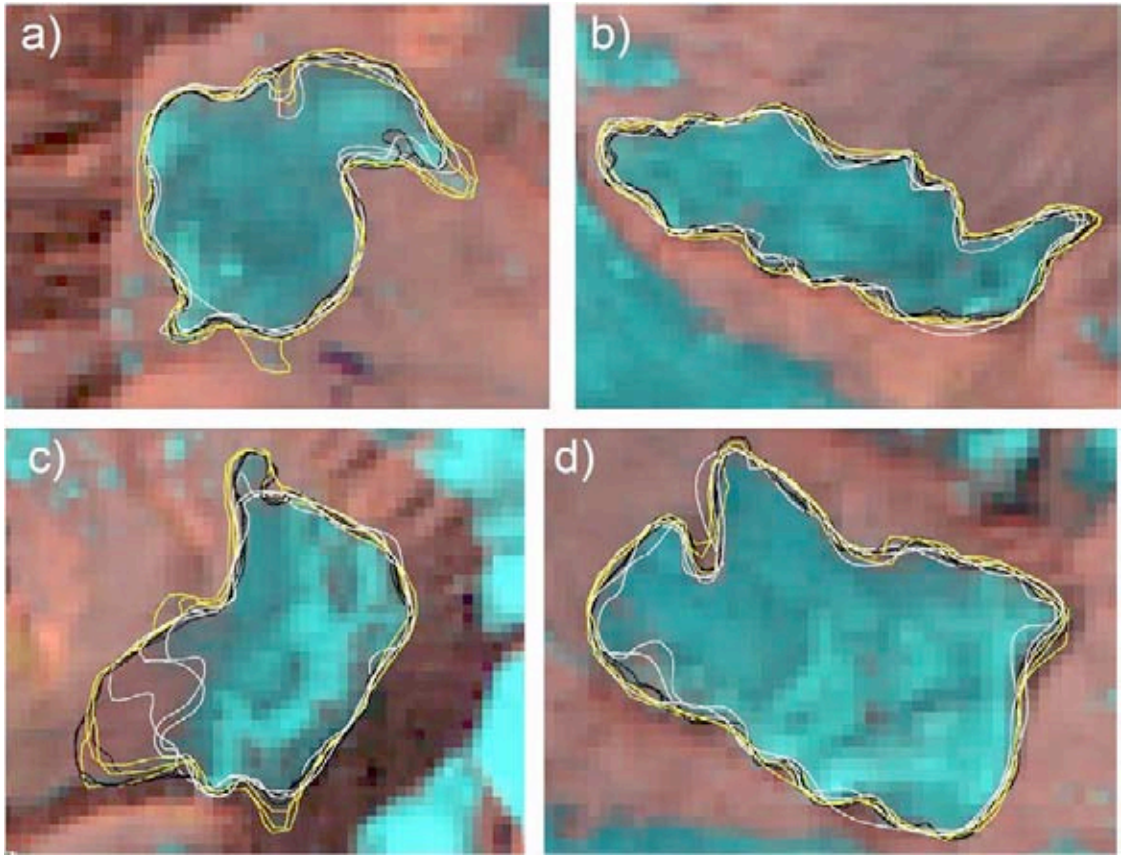

Fig. 2.5: Results from the multiple digitizations of glaciers on a TM scene (bands 543 as RGB) as performed by the same analyst (the different colours refer to three different analysts). A large variability in interpretation is also visible for the same analyst.

A validation of the glacier outlines as derived by the analysts and the automated method with a reference dataset was performed for three of the glaciers from GA2 using high-resolution Ikonos images that were acquired only two weeks after the Landsat scene. For two of these validation glaciers we show the manually digitized extends as an overlay with the high-resolution image in Fig. 2.6 and 2.7. Based on the results from the multiple digitizations, we decided to determine the reference area also from a multiple digitization. This was actually a good idea as the differences in interpretation clearly reveal. The mean standard deviation of the digitizations on the Ikonos data (2.6% of the area) was comparable to the mean difference of the automatically (-4.6%) and the manually derived outlines (-2.1%) of the TM image when compared to a reference value. Hence, a real or absolute ‘validation’ of the TM-derived outlines is not possible.

Fig. 2.6: Overlay of the manually digitized glacier extents for the Upper Guslarferner in the Austrian Alps. The white line is automatically, the green line manually derived from TM. . Image: screenshot from Google Maps, North is at top, one raster cell is 30 m.

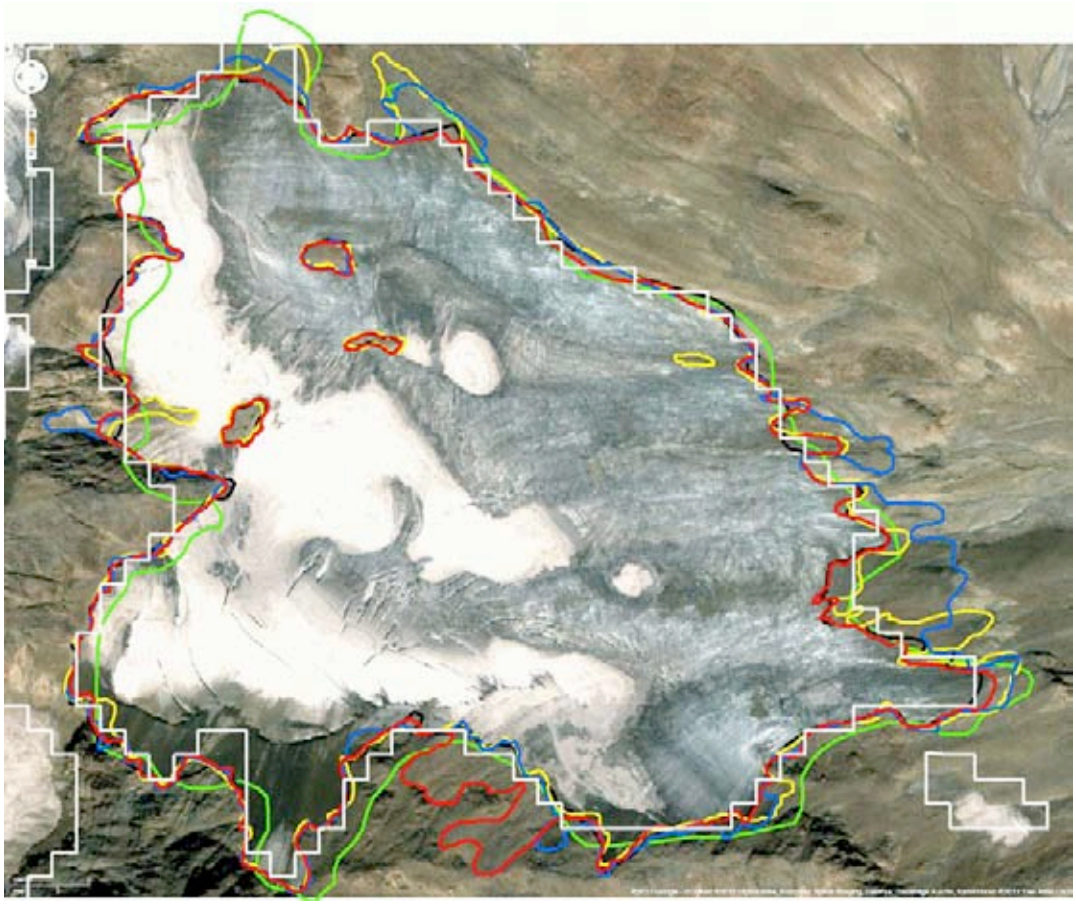

Fig. 2.7: Overlay of the manually digitized glacier extents for Steinschlagferner in South Tirol. The white line is automatically, the green line manually derived from TM. Image: screenshot from Google Maps, North is at top, one raster cell is 30 m.

For test region **GA3** in the Swiss Alps three glaciers as depicted on high-resolution (0.4-0.8 m) aerial photography were used for the digitization. The three glaciers include the typical problems (debris, shadow, seasonal snow) and will thus also include some variabilty due to a different interpretation by the analyst. Problems with visibility due to resolution are expected to play a minor role. The overlays are shown in Fig. 2.8, partly with outlines from TM on top (in white). The smallest of the three glaciers (Fig. 2.8a) is only 0.09 km^2 in size and heavily shadowed. However, apart from the regions above the Bergschrund (the large crevasse at the head of the glacier), all outlines are very similar (+/-1/2 TM pixel) and the standard deviation of the area differences is less than 2%. Also the TM image acquired in the same year (2009) is in good agreement with the outlines in the accumulation region and slightly larger in the ablation region (upper right part in the image). Overall, the area derived from TM is 12% larger, but actually this is only 0.01 km^2 in this case.

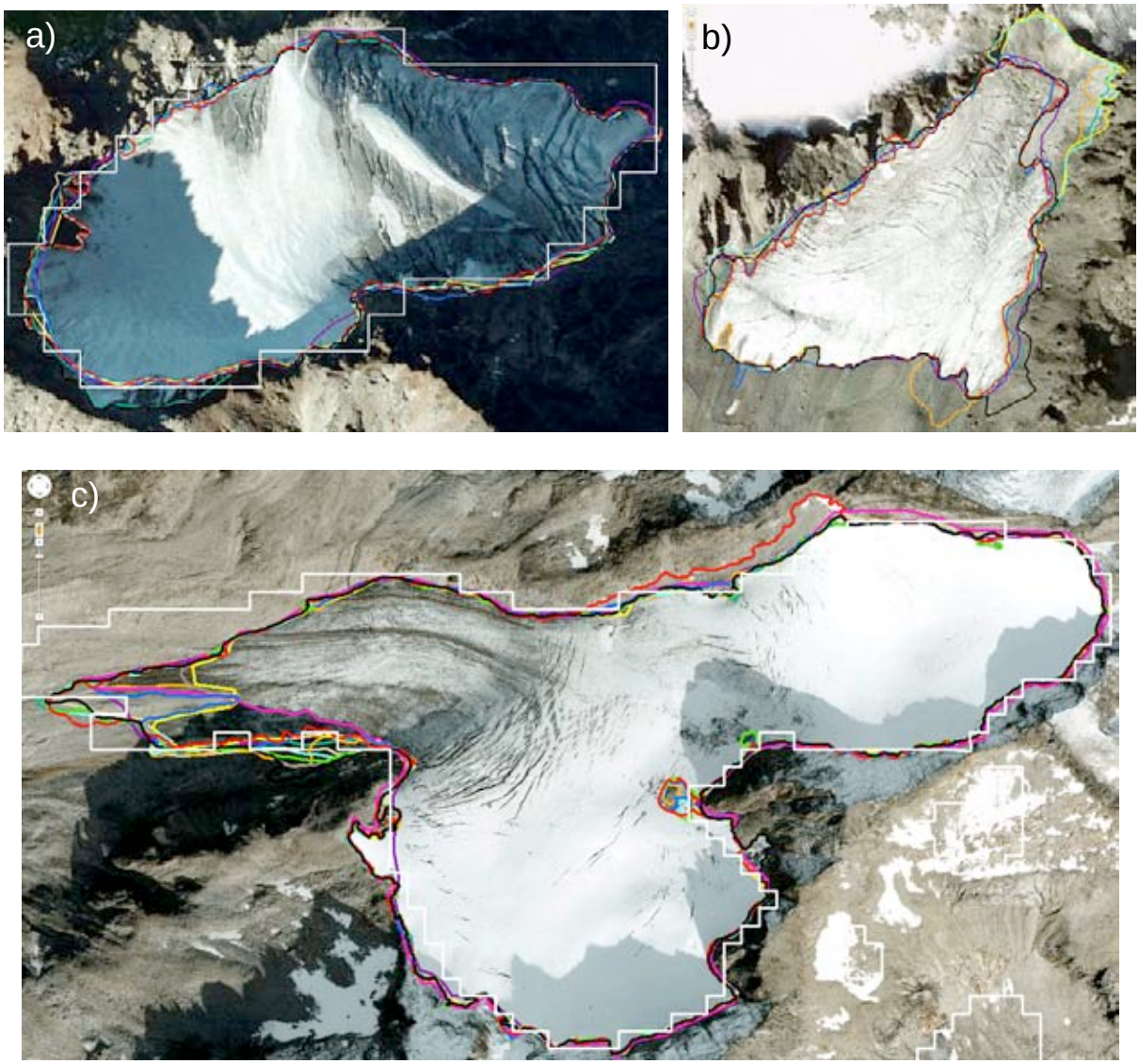

Fig. 2.8: Overlay of manually digitized glacier extents for the three test glaciers in the Swiss Alps. a) Vadret Futschöl, b) Vadret d'Urezzas, c) Geren glacier. Images: screenshots from Google Maps, North is at top, one raster cell is 30 m.

The second glacier (Fig. 2.8b) is slightly larger (0.3 km^2) and partly covered by debris (in the upper part and along the lateral sides). These regions caused again some variability in interpretation illustrating the difficulties with this issue despite the high resolution and with good optical contrast. In this case parts of the glacier was buried under rockfall that has a different appearance from a nomal moraine and is thus even harder to identify. The standard deviation of the digitizations by all analysts is hence larger than before (nearly 9%), but the difference of the mean to the TM derived size is smaller (-7%). However, it has to be noted that in this case the mean value of the manual digitizations is too high as large parts of a bare rock surface at the highest parts are included by four analysts. If these regions would have been digitized correctly, the agreement would be much better.

The largest glacier in this sample (Fig. 2.8c) is 0.7 km^2 in size and has some debris cover, shadow and seasonal snow as a mapping challenge. The participants had no problem with the snow, some differences in the deepest shadow part, and larger differences for the debris at the terminus. The ice-cored moraine was partly included and partly not resulting in a length difference of 270 m due to interpretation differences. This points to the necessity of working with digital reference points in the glacier forefield to measure length changes accurately and consistently. The TM outline agrees very well in the upper part, but is much larger at the terminus. In this case this is due to the different acquisition dates (2003 compared to 2010) and the glacier area loss during that period (about 0.05 km^2). The cumulative retreat over this period is about the same (300 m) as that due to the interpretation differences.

2.4.2 Statistical measures

Plotting the standard deviations of the area differences vs glacier size (Fig. 2.9) for the 8 glaciers from GA1 and 9 of the ten glaciers from GA2 (the largest one is not shown), reveals some dependency between the variables: Towards smaller glaciers the standard deviation (SD) of the relative area differences increases. For glaciers larger than 1 km^2 the values are in general smaller than 5%. In absolute terms the SD varies between 0.03 and 0.08 km^2 and the differences to the TM-derived areas is between 0 and 0.03 km^2 (see Table 2.3).

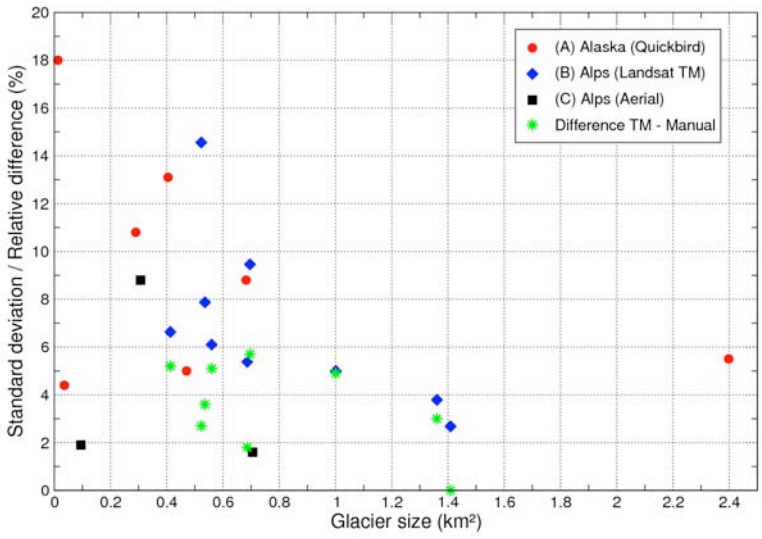

Fig. 2.9: Glacier size vs. standard deviation for 7 of the 8 glaciers from GA1 (A), 9 of the 10 glaciers from GA2 (B), 3 glaciers from GA3 (C), and the difference of TM with the mean.

For all individual glaciers of GA2, the difference of the TM-derived area to the mean of the manually digitized area is smaller than the standard deviation of the latter. This holds also for the comparison with the second digitizing of the same set of glaciers, and largely also with the third digitizing (the difference is larger for three glaciers). Apart from two glaciers where the difference is close to zero, the TM-derived areas are always (slightly) smaller. This is likely due to a more generous inclusion of mixed pixels in the manual digitization. Glacier #3 in GA2 had a considerable amount of debris cover at the terminus (Fig. 2.10) that was difficult to interprete as the high standard deviation in Table 2.2 shows. As visible in Fig. 2.4c, this part was missing in the 2003 dataset derived from TM. To have a comparable dataset, we have added this part to the TM-derived outline afterwards.

There is a slight tendency towards higher standard deviations in the second digitization of GA2 and somewhat lower ones in the third. However, this is not a systematic trend as different glaciers are considered. Moreover, the differences to the first digitizations are small. The multiple digitizations of the Ikonos images used for validation (GA2 VAL in Table 2.3) revealed standard deviations comparable to GA3. The mean area from GA2 round robin is very close to the reference dataset (only 0.2 km^2 or -1.3% smaller).

Region	Glacier	Area [m^2]			StDev	Automated with TM	
		mean	min	max	[%]	Area [m^2]	Diff. [%]
GA1 (n=13, t=40-400)	1	13161	10343	19557	18.0	-	-
	2	470309	391475	474862	5.0	-	-
	3	681970	544314	772594	8.8	-	-
	4	35890	33795	38286	4.4	-	-
	5	58448	26372	39294	30.1	-	-
	6	2397957	2196792	2588353	5.5	-	-
	7	289665	218184	326143	10.8	-	-
	8	404490	344254	503332	13.1	-	-
	All	4322091	4057342	4652627	5.7	-	-
GA2 RR (n=13, t=40-150)	1	413025	358260	470325	6.6	391500	-5.2
	2	1423065	1328811	1556307	3.8	1380600	-3.0
	3	528487	430703	659533	14.6	514400	-2.7
	4	559764	511161	609834	6.1	531000	-5.1
	5	1000940	932074	1128614	5.0	952200	-4.9
	6	686207	629529	738677	5.4	674100	-1.8
	7	535773	474604	649233	7.9	516600	-3.6
	8	1409021	1326903	1470863	2.7	1409337	0.0
	9	695791	602448	835277	9.5	656100	-5.7
	10	8918710	8558375	9397447	2.9	8936010	0.2
	All	16170784	15520158	17150047	2.1	15961847	-1.3
GA2 VAL	4	546858	539848	556046	1.3	531000	-2.9
	8	1511081	1460441	1570260	3.2	1409337	-6.7
	9	663232	644903	689287	2.8	656100	-1.1
	All	2721171	2656784	2815593	2.6	2596437	-4.6
GA3 (n=12, t=10-60)	1	94683	92611	97425	1.9	106256	12.2
	2	306672	271571	357903	8.8	284339	-7.3
	3	705369	680312	720165	1.6	754200	6.9
	All	1106724	1074279	1155421	3.6	1144795	3.4

Table 2.3: Comparison of glacier area values as derived for the three test regions by all analysts. For the regions were multiple digitizations were performed (GA2 RR), the mean value of the first digitization from each analyst is taken for the comparison. Abbreviations: n = sample size, t = time used for the manual digitizing in minutes, Diff. = Difference, RR = round robin, VAL = validation data set.

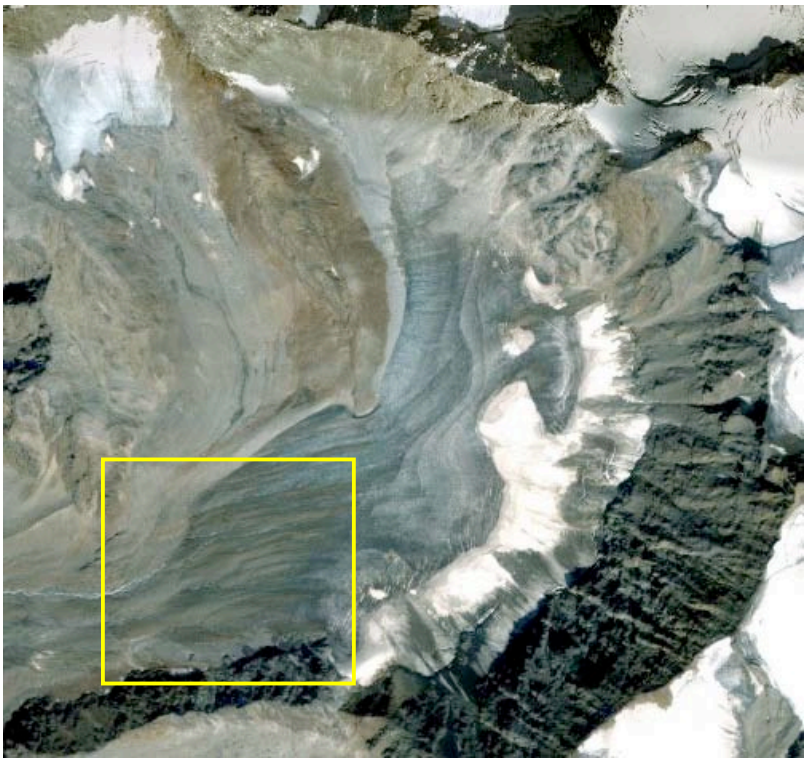

Fig. 2.10: The debris-covered terminus (yellow box) of Wannetferner in the Ötztal Alps (Austria) as seen with Ikonos on 13. August 2003. This part was added in the TM image to get comparable values for Table 2.2. Image: screenshot from Google Maps.

As mentioned above, the level of detail (number of tics per outline) varies with the analyst and with the source data. In contrast to automated assessments, manual digitizing averages through a group of pixels considering a large range of factors (colour, brightness, shape, glacier flow, shadow/debris, etc.). The degree of generalization increases with decreasing contrast and is only repeatable within the error bounds given above. We have thus also analyzed **the time required** for the digitization (as provided by the participants in the feedback form) and its relation to pixel size. In the mean, the digitization required 63 min for the 8 glaciers of GA1 (7.9 min per glacier), 28.5 min for the ten glaciers in GA2 (2.8 min per glacier), and 32 minutes for the three glaciers in GA3 (10.7 min per glacier). In other words, the digitization per glacier is about 3-4 times longer on the high-resolution images than with TM. This is likely due to the higher number of pixels that have to be included during the digitizing. Considering the slightly higher standard deviations of the higher resolution datasets, it seems that the higher resolution is not worth the effort. As the script-based (automated) delineation of all glaciers (could be several hundred) within a scene might take 5-10 minutes (incl. one or two adjustements of the threshold), we get a mapping duration of about 1 second per glacier (for 500 glaciers in a scene). As the accuracy is as good (or even slightly better) as with the manual digitization, manual digitization of debris-free glaciers is clearly not recommended.

2.5 Comparison of the applied algorithms

2.5.1 Results of the round robin for GA4

Although well-established simple and accurate algorithms for glacier mapping from optical satellite data exist (e.g. Albert, 2002; Paul et al., 2002, Racoviteanu et al., 2009), we took the opportunity with the round robin for glacier area to assess if any new and promising algorithms were developed since these papers were published. For this purpose we selected a small region in the Himalaya for free application of any available algorithm. In providing only ETM+ bands 1 to 5, we intended not to go too deep into potential methods that map also the debris-cocvered parts. However, apart from an uncorrected region, we also asked the participants to correct a small subregion within that scene for debris-cover. This was done by all participants using the provided band 543 composite as a background and guide (Fig. 2.11).

As is obvious from Fig. 2.11, there is very little variability in the glacier outlines for clean ice, but quite a large one for the debris-covered parts. Of course, these differences have a considerable impact on the derived glacier size. We thus decided to analyse debris-free and debris-covered glaciers separately. Their extents were derived by digital intersection of the respective outlines with a basin drainage map that was derived from watershed analysis of the SRTM DEM (in its void-filled version from CGIARS). The outlines from the glacier inventory created in the GlobGlacier project (Frey et al., in revision) are used as a reference dataset. In this dataset glacier outlines were verified with PALSAR coherence images (Strozzi et al., 2010). Table 2.4 provides an overview on the applied algorithms and the time required according to the feedback forms. The analysis of omission and commission errors was found to be not very useful as the differences in the location of the outlines were rather random.

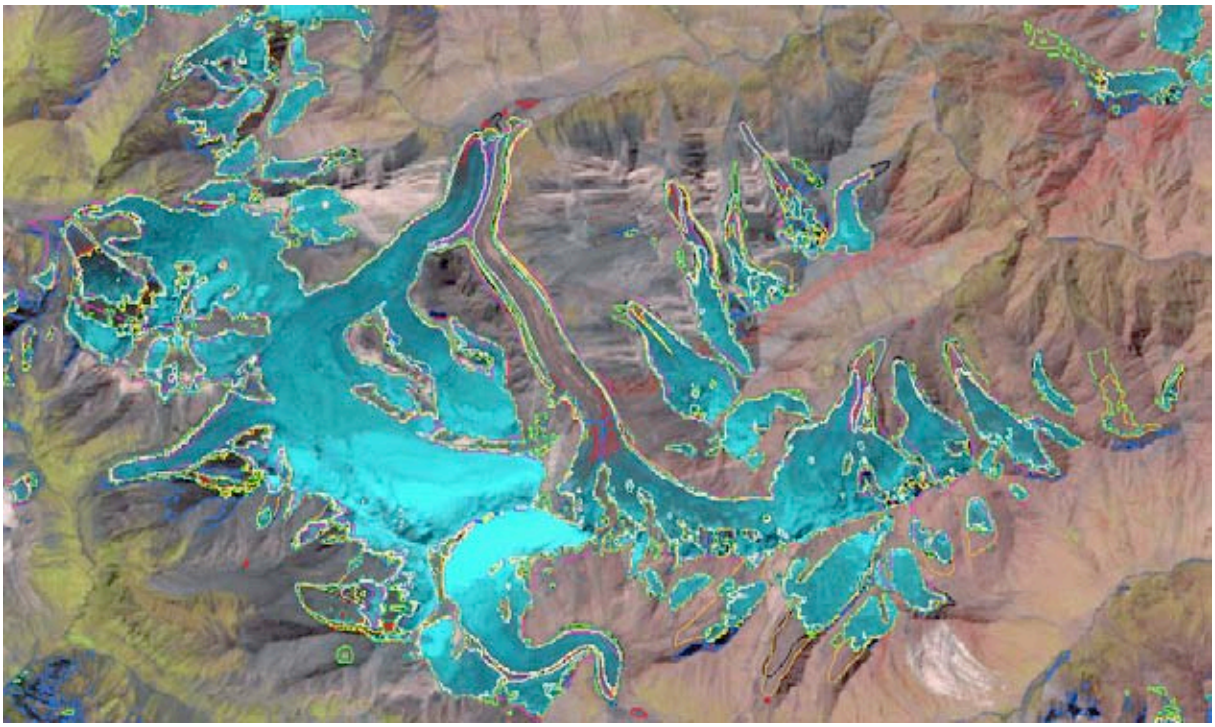

Fig. 2.11: Overlay of the glacier outlines from the different participants for the subset of test region GA4 where editing of wrong classification results was requested.

The overview in Table 2.4 reveals that most of the participants have applied a band ratio either with bands 3/5 (5 times), 4/5 (4 times), or the NDSI (3 times). Additionnally, ISODATA clustering, principal component analysis, unsupervised classification, and a decission tree classifier were applied. Some participants ignored the thermal bands, the panchromatic band or a DEM to apply their algorithm of choice, others also downloaded the original Landsat scene and included this information. It seems that we have thus not made sufficiently clear in the instructions that this possibility existed for those algorithms that require more input. There is only a small variabilty in the thresholds used (between 2.0 and 2.3), but for the NDSI the values are quite different. The overlay of the non-edited glaciers in Fig. 2.11 revealed indeed only very small differences among the applied algorithms. This confirmed that accurate mapping of clean ice is not a big issue for any of the algorithms and the problems start with the interpretation of the debris-covered parts. As this was done manually by all participants in a final stage, results might be different only due to the experience of the analyst and the additional information used (e.g. high-resolution imagery in Google Earth).

Nr.	Algorithm description	Time	Software
1	03 band ratio 3/5 (thr2.3)	40/25	Envi, ArcGIS
2	05 band ratio 3/5 (thr 2.0), thermal band (thr 0-148), slope from GDEM2 (thr 0-13), decision tree classifier, 3x3 filter	4h/2h	Envi, ArcGIS
3	10 NDSI (thr 0.58-0.92), 3x3 filter	n/a	ILWIS,
4	11 ISODATA clustering on Principal Components from bands 1-5, NDSI and 4/5 ratio	11/25	ERDAS
5	13 band ratio 4/5	90	ArcGIS
6	15 band ratio 3/5 (thr 2.0), median filter 3x3	?	?
7	16 band ratio 4/5, 20 categories	7 / 150	ArcGIS
8	17 NDSI (thr 0.4), 3x3 median filter	10 / 40	PCI Geomatica
9	18 band ratio 4/5 (thr 2.0), 3x3 filter, unsupervised class. with 15 classes of the segmented image, missed thermal and pan bands	30	Envi
10	20 band ratio 3/5 (thr 2.0), 3x3 low-pass filter	30 / 30	Envi/ ArcGis
11	01 band ratio 3/5 (thr n/a)	10 / 30	Arc/Info

Table 2.4: Comparison of the applied algorithms, their processing time, and the software used to perform the mapping according to the feedback forms of the participants.

A more detailed analysis of the manually digitized extents (Figs. 2.12 and 2.13) reveals no systematic differences in the interpretation of debris cover. Sometimes analyst A has digitized the largest extent, sometimes B, sometimes C. From these random differences we conclude that the interpretation is based on very subtle details and a matter of discussion in most cases. A strong obstacle in this region is certainly the high solar elevation that reduces contrast considerably and does thus not help to reveal the subtle differences in the shading of convex surfaces at lower solar elevations. For some glaciers only few participants spotted the debris-covered parts at all, leading to an error in the total size of about 200% in these cases. Though the PALSAR coherence images greatly help to identify such debris covered parts and constrain the position of the outline and the terminus, there is also some uncertainty in the extent derived from them. Considering the experience obtained so far, the mapping is best be done by visually controlling the digitization on the coherence images with optical images and, if available, high-resolution data and oblique perspective views (e.g. using Google Earth).

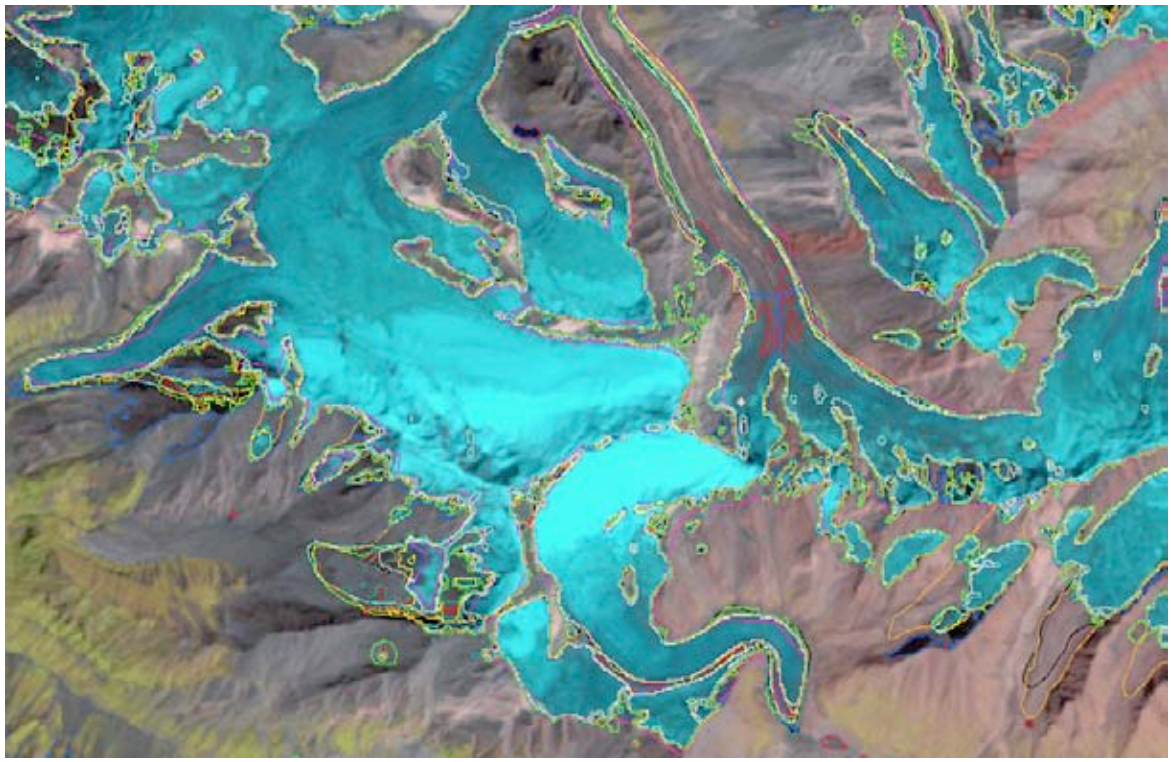

Fig. 2.12: Close-up of Fig. 2.11 (lower left). Only one resp. 2 analysts identified the debris-cover on the two glaciers in the lower tight.

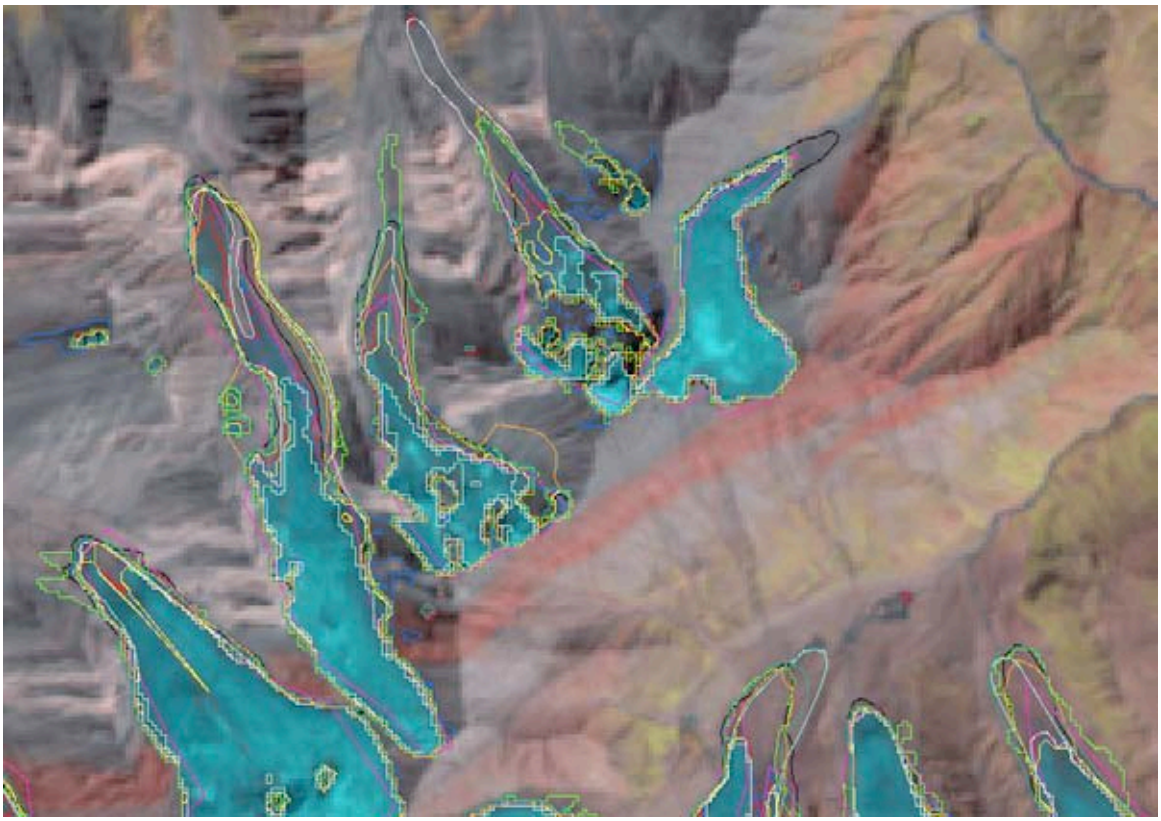

Fig. 2.13: Close-up of Fig. 2.11 (right from the centre). A large variability of the extents can be seen for the debris covered glacier parts.

When comparing glacier areas as derived from the various participants, a wide range of values emerged for individual glaciers. With debris free glaciers showing standard deviations of 4 to 10 percent, debris-covered glaciers had standard deviations larger 150%. The difference between the minimum and the maximum value can differ by up to 200%, i.e. glaciers can be up to 4 times larger when debris-cover is considered. This sounds quite high, but it is likely even higher differences are possible in some regions, as glaciers can also be more or less completely debris covered. In view of these large differences in interpretation of debris cover, we have not performed the originally foreseen more detailed analysis with PALSAR coherence images. Instead, we see our original statement confirmed that a most important product of the Glaciers_cci project for the glacier area product are illustrated guidelines helping to increase the consistency of the datasets in the GLIMS glacier database.

2.5.2 Results of the sensitivity analysis

After analysing the glacier mapping results for GA1 with TM, it was decided to perform the analysis in another, more complex region. This region is located on Landsat TM+ scene 232-089 (path-row) acquired on March 7, 1985 in the Parque Pumalin of Chile and across the border in Argentina (Fig. 2.14). Key to this decision was the more challenging mapping of ice and snow in cast shadow in this region. As shown before (Glaciers_cci, 2012b) there is very little change in mapped glacier extent with the band ratio method when changing the threshold by 0.1 (e.g. from 1.8 to 1.9). With the recommendation to select a threshold value in a most sensitive region and at the same time minimize the workload for manual correction (e.g. Racoviteanu et al., 2009), the focus of a sensitivity analysis must be on the mapping performance in regions of cast shadow (misclassified debris cover and water surfaces have to be corrected in the post-processig stage anyway). The mapping accuracy in shadow regions is largely governed by the additional threshold in band ETM+/TM1 (blue) and has to be selected interactively (Paul and Kääb, 2005).

To determine the changes of the outline due to a change in the TM1 threshold, we have systematically altered (steps of 10) the threshold value between 60 and 100 digital numbers (DN). The impacts on the resulting glacier outlines are visualized in Fig. 2.15 using outlines of different colour. From blue (threshold 100) to green, white, red and yellow (threshold 60) more and more regions with ice in cast shadow are included and the mapping quality is increasingly improved. Individual glaciers have area increases from 0 to 100% or might not have been mapped at all with a certain threshold. For the lowest threshold (yellow), however, glaciers are mapped at increasingly isolated places that might be noise. In this case a threshold below 60 would degrade mapping quality and a threshold value of 65 was finally selected. Hence, to get as much ice in shadow correctly mapped as possible, the threshold value has always to be tested against its lower limit. This makes automatic determination of a threshold value (e.g. by using reflectance values after atmospheric correction) problematic, as the increasing noise towards lower threshold values is only recognizable by visual inspection.

For a region like this with high amounts of seasonal snow outside of glaciers (and noise in shadow regions), it is always advisable to apply a low-pass filter for noise reduction (e.g. median with 3 by 3 kernel size) before the raw glacier map is converted to vector outlines. However, in regions with (near) perfect snow conditions and numerous very small glaciers this needs to be carefully checked (e.g. Paul et al., 2011b).

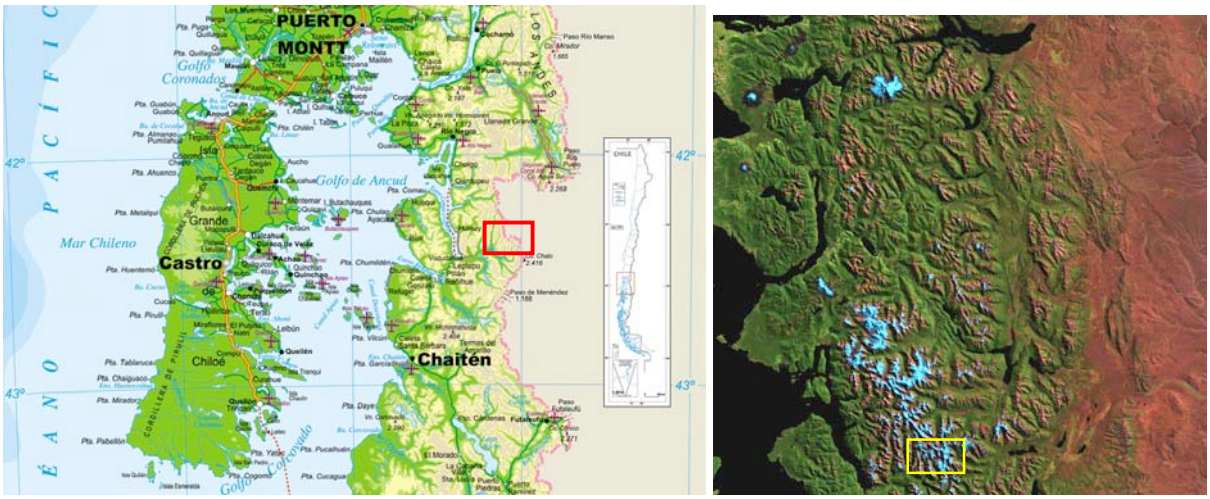

Fig. 2.14: Location of the test region depicted in Fig. 2.15 on a map of Chile (left) and on the Landsat quicklook of scene 232-089 (right). The map was taken from www.educarchile.cl.

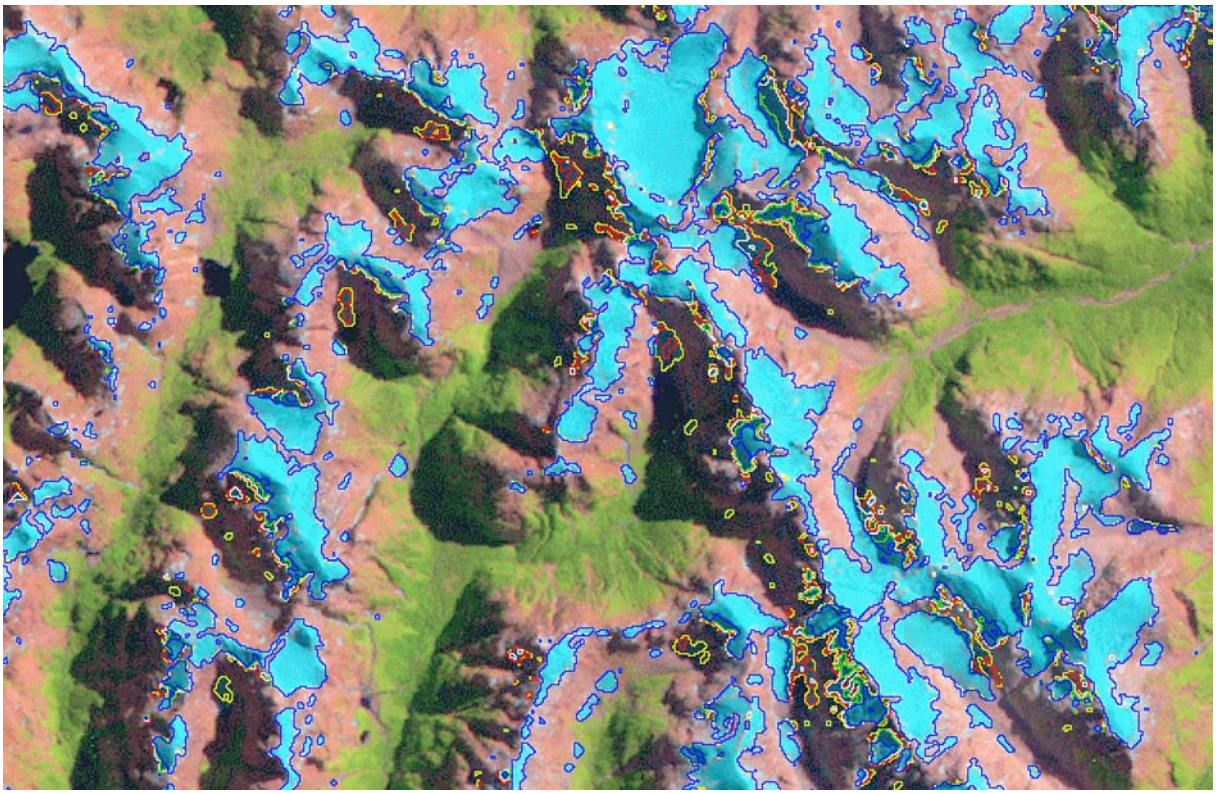

Fig. 2.15: Glacier outlines for the test site shown in Fig. 2.14 with five different values of the threshold in band TM1 applied (blue: 100, green: 90, white: 80, red: 70, yellow: 60). Substantial changes take only place in regions with ice and snow located in cast shadow. The finally selected threshold in TM1 is 65.

2.6 Glacier mapping accuracy and algorithm selection

The results of the round robin in regard to glacier mapping accuracy revealed the following main results:

- automatic delineation of glaciers is at least as accurate as manual digitizations when comparing the standard deviation of the results from the participating analysts to the difference of the automatically derived outlines to a reference value (i.e. the mean of all manual digitizations)
- the accuracy (and consistency) is not better when high resolution data are used as the methodological differences in interpretation remain and a lack of contrast can degrade the interpretation (the latter might change with the availability of a high resolution SWIR band)
- the accuracy of automatically derived glacier extents is difficult to determine as the glacier extents derived from a reference dataset have a similar uncertainty range (of a few percent)
- the accuracy of a glacier outline degrades with the difficulty in interpreting the debris-covered glacier parts
- the variability in the interpretation of the terminus position for debris-covered glaciers by different analysts is in the order of several hundred metres
- we recommended to determine the precision of digitized outlines by performing a multiple digitization experiment on a couple of differently sized glaciers (with and without debris cover)

From the algorithm comparison we draw the following **main conclusions**:

- The uncorrected outlines for clean ice varied by about 4-10% (standard deviation), so analysts can apply their algorithm of choice to map clean ice.
- Automated mapping is also recommended as a starting point for any glacier delineation as it is much faster and as accurate as manual digitizing.
- Considerable time should be spent on correcting the automatically derived outlines as good as possible in regions of misclassification (debris, shadow, water).
- Supporting information from high-resolution data, DEMs or coherence images from RADAR sensors should be taken into account whereever possible.
- Methodological uncertainties (e.g. due to seasonal or perennial snow, creeping permafrost bodies) should be handled pragmatically, but the rules applied should be reported in the literature
- Outlines should be made publicly available in their original vector format.
- Illustrated guidelines should be complied for the analysts to improve the consistency in the interpretation of critical cases (debris, shadow).

3. Elevation changes from altimetry

3.1 Overview of the round robin experiments

The aims of the round robin exercise for the elevation change (ELC) product obtainable using repeat altimetry were:

- the validation of the different repeat altimetry algorithms for the product generation;
- the selection of the 'best' performing algorithm based on the selection criteria described in the PVP (Glacier_cci, 2011).

The validation activity performed, whose results are described in the following sections, has been based on the following components:

1. comparison of temporally consistent airborne elevation changes with satellite altimeter elevation changes;
2. comparison of elevation changes derived from different sensors (e.g. radar vs. laser);
3. comparison of elevation changes derived from different algorithms (cross-track vs. repeat track).

The algorithm comparison has been focused on a representative set of glacier sites around the Earth, trying to cover also critical areas, such as glaciers with a rough surface. In particular, the regions chosen have been:

- Devon Ice Cap (EC1) on Devon Island in Nunavut arctic Canada
- Austfonna Icecap (EC2) in Nordauslandet Svalbard archipelago
- Antarctic Peninsula (EC3).

The potential participants to this experiment have been identified and were contacted directly for their contributions. They have been asked to generate the elevation change product using one or more of the algorithms described in the ATBDv0 (Glaciers_cci, 2012) and here briefly described for sake of simplicity:

- Cross-Over method, (abbreviation XO-RepAlt) as described in Zwally et al. (1989) and Wingham et al. (1998).
- Repeat Track method in their different variants as described in Moholdt et al. (2010 a and b), Howat et al. (2008) and Pritchard et al. (2009). The latter differ in the way the topography is considered between near repeat-tracks when comparing elevations from different tracks. In Moholdt et al. (2010a and b) a method which uses along-track interpolation to restrict the DEM slope correction to the cross-track distance between two repeat-tracks has been described (abbreviation DP-RT-RepAlt). Whereas, in order to avoid the use of external DEM, Pritchard et al. (2009) fitted a triangular plane to three elevation observations (abbreviation TP-RT-RepAlt), and used the plane as reference for measurements falling inside this triangle, while Howat et al. (2008) and Moholdt et al. (2010a and b) used rectangular reference planes determined by least squares fitting to segments of repeat-track of GLAS data (abbreviation RP-RT-RepAlt).

The following Table 3.1 gives an overview of the external and internal participants and the results obtained using the algorithms mentioned above over the different test sites based on the EO data available. The internal round robin exercise performed at the University of Leeds focused on the Devon Ice Cap and the Antarctica Peninsula. In regard to the first test site, the surface elevation change has been obtained using a repeat track algorithm, still based on the use of an external DEM for the topography correction, but based on a slightly different approach compared to the DP-RT-RepAlt algorithm described in Mohold et al (2010 b) (its abbreviation DS-RT-RepAlt stands for DEM Subtracting Repeat Track Repeat Altimetry, see section 3.3.1.2 for the description). Over the Antarctica Peninsula the XO-RepAlt algorithm has been applied.

Parti-cipant	Affiliation	EC1	EC1	EC2	EC2	EC3	EC3
		Laser	***Radar***	***Laser***	***Radar***	***Laser***	***Radar***
1	Scripps Institution of Oceanography	RP-RT-RepAlt		RP-RT-RepAlt DP-RT-RepAlt			
2	Nansen Envi-ronmental and Remote Sensing Centre						XO-RepAlt
3	University College of London		XO-RepAlt		XO-RepAlt		
4	University of Leeds	DS-RT-RepAlt					
5	University of Leeds						XO-RepAlt

Table 3.1: Overview of the participants and the performed analysis in the three test regions; the DS-RT-RepAlt is still based on a DEM for the topography correction, but it slightly differs from the DP-RT-RepAlt algorithm described in the ATBDv0 (Glaciers_cci, 2012).

As it can been seen from the above table, the exercises were based both on laser and radar altimetry data. In particular, the laser data used over the Devon Ice Cap (EC1) and Austfonna Ice Cap (EC2) were the ones acquired by the ICESat/GLAS instrument, whereas the radar data used on all the three test sites referred to the measurements acquired by the Envisat-RA2 altimeter. The following table provides for each algorithm applied and for each test site the information on the EO dataset used together with the temporal coverage.

Algorithm	EC1	EC1	EC2	EC2	EC3	EC3
	Sensor	***Temporal coverage***	***Sensor***	***Temporal coverage***	***Sensor***	***Temporal coverage***
DS-RT-RepAlt	ICESat/GLAS	2003-2009				
DP-RT-RepAlt			ICESat/GLAS	2003-2009		
RP-RT-RepAlt	ICESat/GLAS	2003-2009	ICESat/GLAS	2003-2009		
XO-RepAlt			Envisat/RA-2	2002-2010	Envisat/RA-2	2002-2010

Table 3.2: EO data used for each applied algorithm and each test site and relative temporal coverage.

3.2 Assessment of algorithm performance

3.2.1 Influences on the accuracy of glacier elevation changes from repeat altimetry

When a DEM is used to correct repeat-track data for the cross-track slope, it is not the absolute accuracy of the DEM that is important, but rather the reproduction of the relative local topography that is used to correct for the cross-track slope. In addition, the along-track interpolation error has to be considered. The combined *dh/dt* error varies greatly in space depending on the repeat-track separation distance, the quality of the DEM, the length of the time span, as well as the surface slope and roughness. On the other hand, the application of the plane method assumes that the regression scheme is able to separate between the slopes of a plane and the average elevation change rate (*dh/dt*). The along-track slope component is typically well resolved by each repeat-track, while the cross-track slope component of a plane is dependent on a number of non-coincident repeat-tracks which are influenced by *dh/dt*. A potential problem with the plane method for *dh/dt* calculation is the uneven temporal data sampling (Figs. 3.1 and 3.8). There is typically more data from the winter campaigns, and less data from the summer campaigns. The risk of a seasonal bias in *dh/dt* is especially high for planes where the earliest and latest satellite observations stem from different seasons.

3.2.2 Product validation strategy

The validation criteria are based upon the absolute difference of the elevation change maps obtained using different methods and/or sensors. In addition, the quantitative analysis is provided by means of the root mean square error, RMSE, and the correlation coefficient, R, defined as:

$$RMSE = \sqrt{E\left\{(x-y)^2\right\}}$$

$$R = \frac{E\left\{(x-\bar{x})(y-\bar{y})\right\}}{\sqrt{E\left\{(x-\bar{x})^2\right\}E\left\{(y-\bar{y})^2\right\}}}$$

The quantities *x* and *y* represent the two elevation change maps to compare.
This applies to the comparison between (a) airborne elevation changes and satellite altimeter elevation changes; (b) estimates of elevation changes using different satellite sensors and (c) estimates of elevation changes obtained by applying different algorithms to the same sensor data set.

3.3 Analysis and comparison of the round robin results

3.3.1 EC1 – Devon Ice Caps

3.3.1.1 Input data

The EO data used on this test site consist of the full ICESat/GLAS archive which covers the period 2003-2009, and the Envisat GDR data (version 2.1) in the period between 2002-2010. Table 3.3 reports more details of the ICESat dataset used by each participant. Figure 3.1 shows how the number of the ICESat acquisitions over the test site varies between the overall observation campaigns. Figure 3.2 shows the ICESat spatial data distribution in the whole acquisition period, i.e. 2003-2009.

	Participants	
	Scripps Institution of Oceanography	University of Leeds
Platform and sensor code (e.g. IceSat/GLAS, E1/RA, E2/RA, EV/RA-2)	ICESat	ICESat
Data product used	GLA06	GLA06
Release number (or version)	31	33
Repetition cycle	33 days campaigns repeated 2-3 times a year	33 days campaigns repeated 2-3 times a year
Number of repeat cycles	16	
List of cycles used	All	All
Number of the reference cycle used in the processing	Multiple, ascending and descending	N/A
Track 1 & mode: number of the track & ascending/descending (A/D)	N/A	N/A
Track 2 & mode: number of the track & ascending/descending (A/D)	N/A	N/A
Start date of the observations (YYYYMMDD)	20030924	20030222
End date of the observations (YYYYMMDD)	20091011	20091008
Projection (including zone number)	UTM 17X	UTM 17X
Region (Country and region)	Canadian Arctic	Canadian Arctic

Table 3.3: Summary of the ICESat data used by each participant.

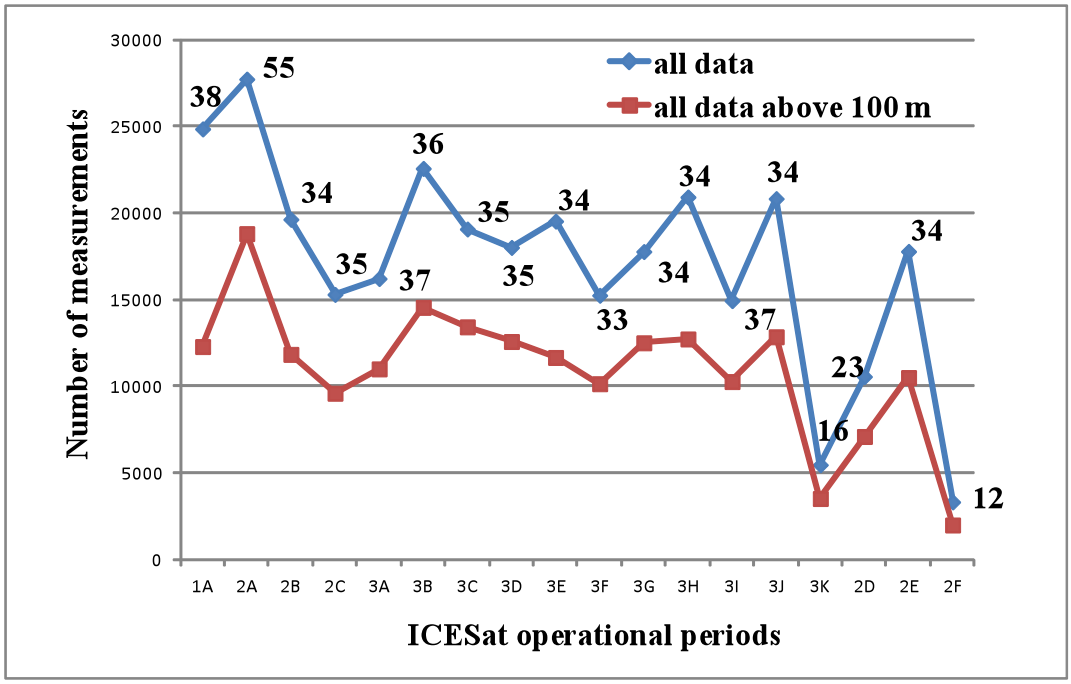

Fig. 3.1: Numbers of ICESat measurement for each operational period over Devon Ice Cap: the blue line represents the whole acquisition data; the red line represents all available data with measured elevations above 100 m related to the geoid; the labelled number refer to the number of operation days for each observation campaign.

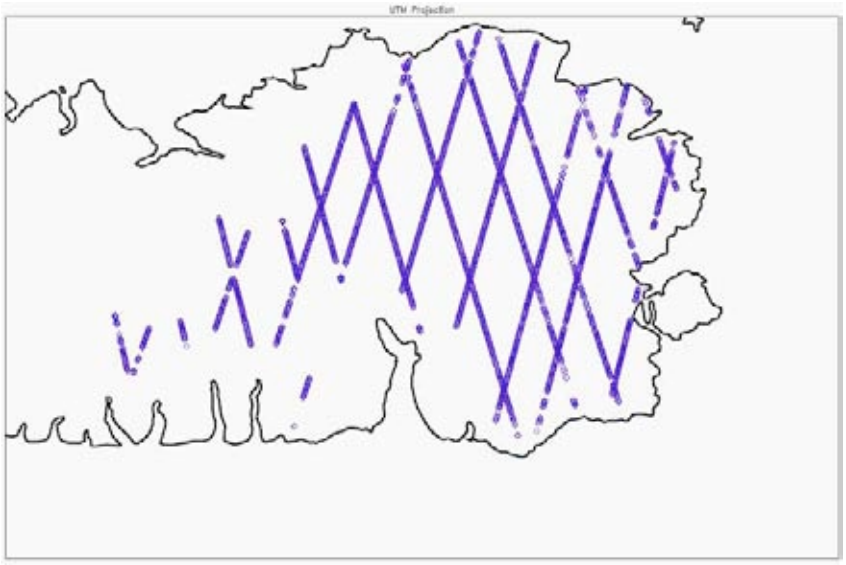

Fig. 3.2: ICESat spatial data distribution over Devon Ice Cap in the period 2003-2009.

3.3.1.2 Description of algorithms applied by the participants
The round robin participants used the following algorithms:

- RP-RT-RepAlt
- DS-RT-RepAlt
- XO-RepAlt

The first one is described in detail by Mohold et al. (2010 b) and reported in the ATBDv0 (Glaciers_cci, 2012). It is based on a least-squares regression technique that fits rectangular planes to segments of repeat-track ICESat data. For each plane the north and east slopes and the elevation change rate, *dh/dt* supposed constant, have been estimated.

The DS-RT-RepAlt algorithm is based on the use of an external DEM for correcting the slope effects between non-overlapping repeat tracks, but it slightly differs from the DP-RT-RepAlt algorithm described in Moholdt et al. (2010 b). In fact, the algorithm computes the difference between the GLAS measured elevations and the DEM elevations at each altimetry footprint location to assess the elevation change. The DEM elevations at these locations are obtained by interpolating the DEM grid. The differences achieved are then averaged in order to represent the spatial variation on grid cells of 1 km x 1 km. This procedure is repeated for each ICESat operational period and the elevation trend for each grid cell is determined by fitting a first order polynomial, so that the slope of the polynomial represents the elevation trend in the data bin considered (Rinne et al., 2011). To achieve the results shown in the following section, the DEM used has a spatial resolution of 250 m and has been obtained using the InSAR technique (Fig. 3.3). However, it has been re-sampled in order to create a grid of 1 km x 1 km.

The XO-RepAlt is described in Zwally et al. (1989) and Wingham et al. (1998) and reported in the ATBDv0 (Glaciers_cci, 2012). Table 3.4 summaries some processing characteristics of the methods applied on the ICESat data and the information on additional data used.

	Participants		
	Scripps Institution of Oceanography	University of Leeds	University College of London
Processing methods: XO-RepAlt, DP-RT-RepAlt, RP-RT-RepAlt, TP-RT-RepAlt	RP-RT-RepAlt	DS-RT-RepAlt	XO-RepAlt
Processing comments (e.g. filtering, or improvement of the processing methods)	Some basic filtering: Gardner et al. 2012	Some basic filtering: Rinne et al., 2011	
Estimated accuracy	0.40 m/yr	0.40 m/yr	0.50 m/yr
Spatial density	5%	4%	< 1%
Spatial coverage	75%	75%	< 5%
Processing time (approx. time for the algorithm to produce the elevation change at each point)	5-10 min after reading the data and selecting glacier points (i.e. the algorithm itself)	5-10 min after reading the data	1 day
Computer specification (e.g. Processer speed and RAM)	Standard desktop computer	Standard desktop computer	Standard desktop computer
Additional data (e.g. DEM for slope correction, DEM resolution)	No additional data	DEM at 250 m resolution	No additional data

Table 3.4: Summary of the processing characteristics of the algorithms used and on the additional data needed.

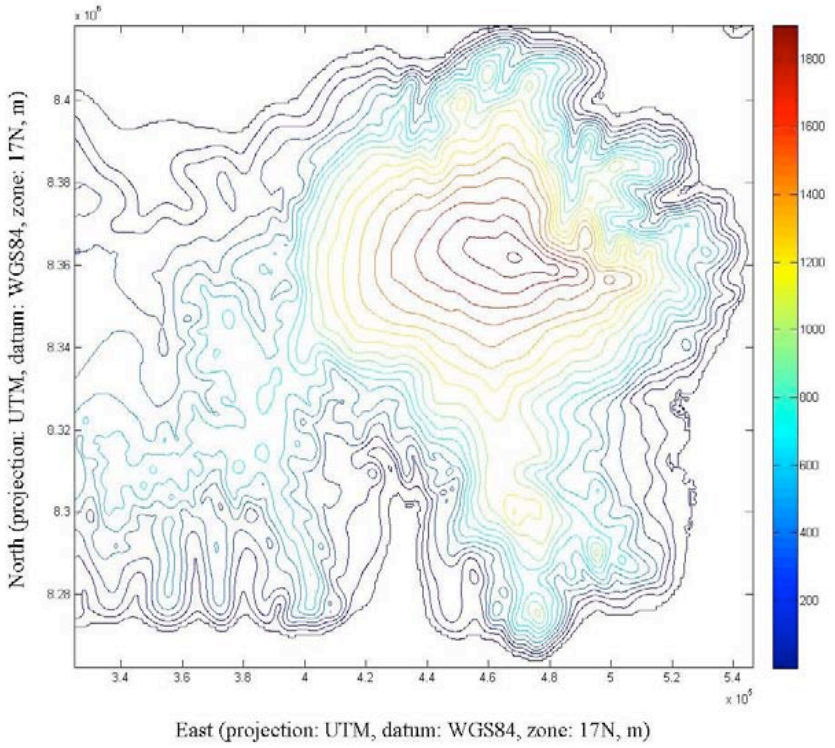

Fig. 3.3: DEM of Devon Ice Cap used when the DS-RT-RepAlt algorithm is applied, the iso-level curves correspond to increment of 100 m.

3.3.1.4 Comparison of the applied algorithms

Figure 3.4d, e and f report the trend of the elevation changes obtained by applying XO-RepAlt algorithms and taking into consideration three different retracking algorithms (landice1, landice2, landocean). From the figure it is evident that the application of this algorithm provides a really poor spatial coverage on the test area. For this reason, these results will not be considered for the inter-comparison validation activity. On the other hand, Fig. 3.2 shows the high spatial coverage of the ICESat/GLAS measurements. Therefore, the two outcomes of the RP-RT-RepAlt and DS-RT-RepAlt algorithms are considered further in the analysis. Panels a) and b) of the same figure show the elevation changes obtained using these two algorithms respectively. The panel c) shows the histograms of the occurrence of the values assumed by the obtained trends. It can be seen that the two histograms are overlapping in terms of the *dh/dt* range and that the first order statistic is of the same order between the two datasets. Indeed, the *dh/dt* trend obtained with the RP-RT-RepAlt algorithm has an average of -0.313 m/yr, whereas the *dh/dt* trend obtained from the DS-RP-RepAlt algorithm has an average value of -0.1307 m/yr.

The trend of the *dh/dt* versus the elevation for both the algorithms applied is shown in Fig. 3.5a and b. As it can be seen the trend has a quite similar linear fit. In addition, Table 3.5 reports the average and the standard deviation values obtained over elevation intervals of 100 m width for both the algorithms applied. The last column of the table shows the absolute difference between the average results of the two datasets. The maximum difference between the two algorithms is about 0.4 m/yr and occurs in the elevation interval between 200-300 m, decreasing after as the elevation increases. This shows that the error is mainly due to loss of data in regions of steep slope at the margin of the ice cap. Finally, another evidence of the good agreement between the two algorithms is reported in Fig. 3.6, which shows a scatter-plot between the two *dh/dt* output datasets. The good correlation between the two algorithms, expressed by the root mean square error, RMSE, of 0.47 m/yr indicates that the two compared algorithms are almost equivalent. In addition the value of the RMSE is comparable to the estimated measurement accuracy of 0.40 m/yr (Table 3.4), and so we conclude that there is no significant difference in the accuracy of either technique.

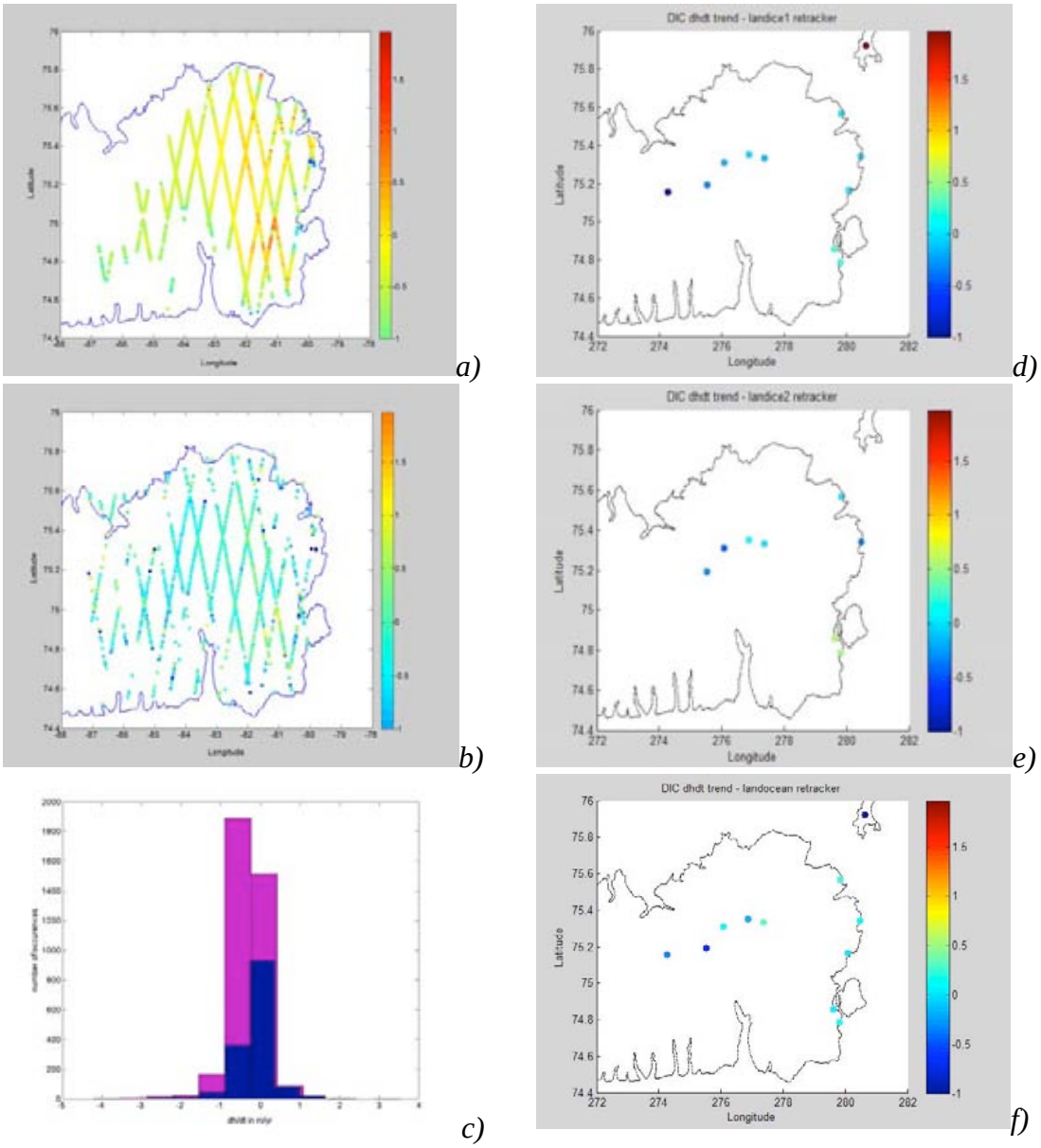

Fig. 3.4: a) dh/dt (m/yr) obtained using the full ICESat/GLAS archive and the RP-RT-RepAlt algorithm; b) dh/dt (m/yr) obtained using the full ICESat/GLAS archive and the DS-RT-RepAlt algorithm; c) histogram of the dh/dt occurrences obtained using: the RP-RT-RepAlt algorithm (purple) and the DS-RT-RepAlt algorithm (blue); d) dh/dt (m/yr) obtained using the Envisat data in the period 2002-2010 and the landice1 retracking algorithm; e) dh/dt (m/yr) obtained using the Envisat data in the period 2002-2010 and the landice2 retracking algorithm; f) dh/dt (m/yr) obtained using the Envisat data in the period 2002-2010 and the landocean retracking algorithm.

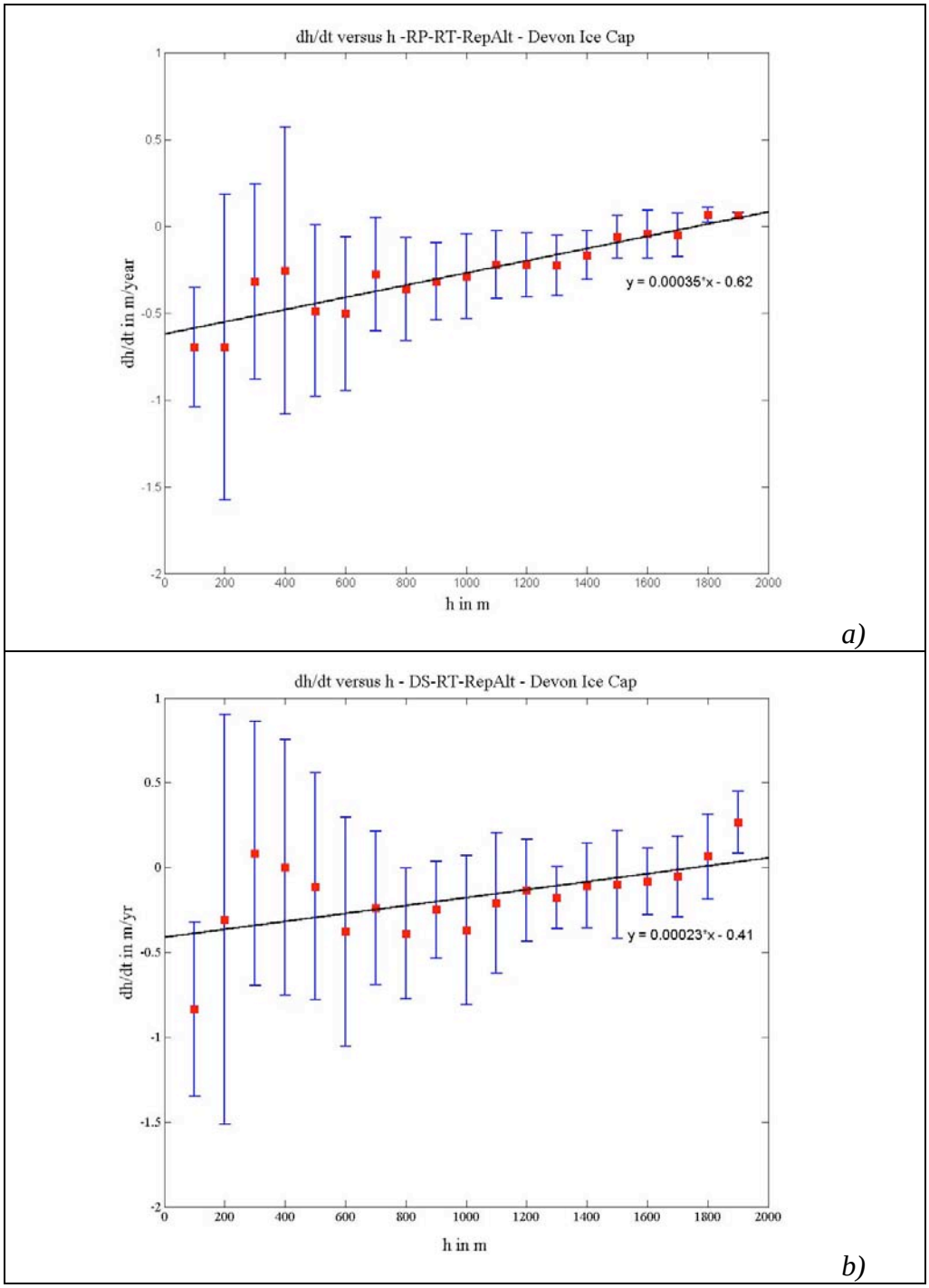

Fig. 3.5: Trend of dh/dt versus elevation h obtained using: a) the RP-RT-RepAlt algorithm and b) the DS-RT-RepAlt algorithm. In both panels, the dh/dt has been averaged over elevation intervals of 100 m width, the error bars representing the standard deviation and the black line is the linear fit.

h (m)	dh/dt (m/yr) RP-RT-RepAlt	dh/dt (m/yr) DS-RT-RepAlt	|Δ| (m/yr)
0 - 100	----	----	----
100 - 200	-0.6955±0.3426	-0.8349±0.5155	0.1394
200 - 300	-0.6979±0.8813	-0.3074±1.2089	0.3905
300 - 400	-0.3199±0.5656	0.0835±0.7808	0.2364
400 - 500	-0.2569±0.8269	0.0007±0.7552	0.2576
500 - 600	-0.4879±0.4931	-0.1118±0.6677	0.3761
600 - 700	-0.5036±0.4431	-0.3771±0.6744	0.1265
700 - 800	-0.2784±0.3263	-0.2380±0.4530	0.0404
800 – 900	-0.3610±0.2962	-0.3896±0.3867	0.0286
900 - 1000	-0.3186±0.2229	-0.2483±0.2856	0.0703
1000 - 1100	-0.2883±0.2440	-0.3693±0.4377	0.081
1100 - 1200	-0.2231±0.1955	-0.2079±0.4139	0.0152
1200 - 1300	-0.2237±0.1835	-0.1340±0.3015	0.0897
1300 - 1400	-0.2279±0.1740	-0.1773±0.1832	0.0506
1400 - 1500	-0.1674±0.1405	-0.1060±0.2489	0.0614
1500 - 1600	-0.0611±0.1223	-0.0998±0.3161	0.0387
1600 - 1700	-0.0461±0.1373	-0.0808±0.1967	0.0347
1700 - 1800	-0.0512±0.1240	-0.0530±0.2362	0.0018
1800 - 1900	0.0651±0.0416	0.0639±0.2501	0.0012
1900 - 2000	0.0619±0.0171	0.2672±0.1840	0.2053

Table 3.5: average and standard deviation values obtained on elevation intervals of 100 m width for the RP-RT-RepAlt and DS-RT-RepAlt algorithms applied; the third columns shows the absolute difference between the averages.

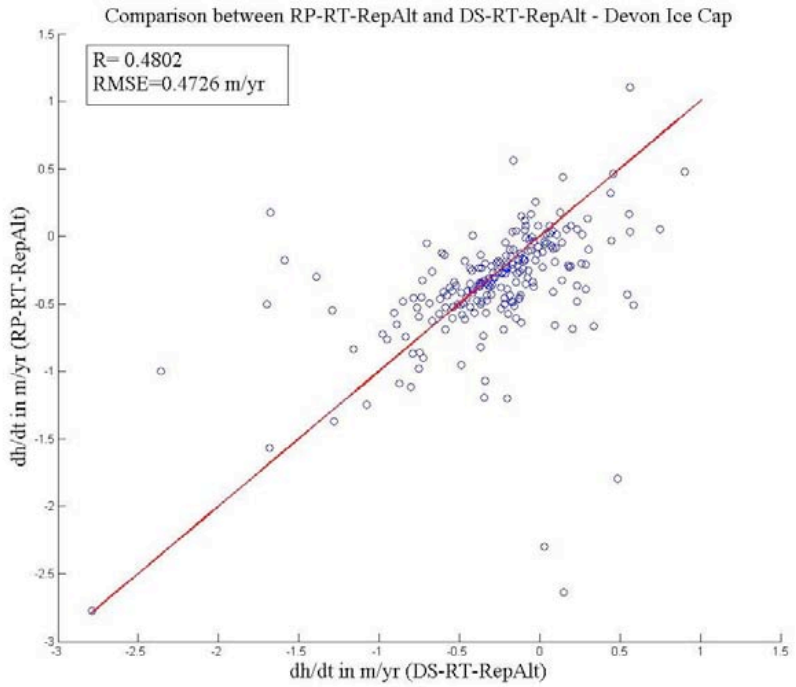

Fig. 3.6: scatter-plot of the dh/dt results obtained using the RP-RT-RepAlt and the DS-RT-RepAlt algorithms.

3.3.1.5 Comparison summary

In the following Table 3.6 we summarize the performance of the algorithms applied at the Devon Ice Cap, in terms of:

- spatial density;
- spatial coverage;
- temporal density;
- CPU and manpower time;
- accuracy.

These characteristics will inform the overall algorithm selection criteria (see section 3.4).

	Algorithms Applied		
	RP-RT-RepAlt	**DS-RT-RepAlt**	**XO-RepAlt**
Spatial density	5%	4%	< 1%
Spatial coverage	75%	75%	< 5%
Temporal density	1.6 $year^{-1}$	1.6 $year^{-1}$	10 $year^{-1}$
CPU	< 1 h	< 1 h	1 day
Manpower time	0.5 day	0.5 day	1 week
Accuracy	0.40 m/yr	0.40 m/yr	0.50 m/yr

Table 3.6: Algorithm performances over Devon Ice Caps.

3.3.2 EC2 – Austfonna Ice Caps

3.3.2.1 Input data

The EO data used on this test consists of the full ICESat/GLAS archive, which covers the temporal period between 2003-2009, and the Envisat GDR data (version 2.1) in the period between 2002-2010. Table 3.7 reports more details of the ICESat dataset used by the external participants (the temporal coverage of each ICESat operational period is summarised in Table 3.3). Figure 3.8 shows how the number of acquisitions over the test site varies between the overall observation campaigns.

	Participants
	Scripps Institution of Oceanography
Platform and sensor code (e.g. IceSat/GLAS, E1/RA, E2/RA, EV/RA-2)	ICESat
Data product used	GLA06
Release number (or version)	28
Repetition cycle	33 days campaigns repeated 2-3 times a year
Number of repeat cycles	13
List of cycles used	2A-3J
Number of the reference cycle used in the processing	Multiple, ascending and descending
Start date of the observations (YYYYMMDD)	20030924
End date of the observations (YYYYMMDD)	20080321
Projection (including zone number)	UTM 33X
Region (Country and region)	Norwegian Arctic
Processing methods: XO-RepAlt, DP-RT-RepAlt, RP-RT-RepAlt, TP-RT-RepAlt	DP-RT-RepAlt RP-RT-RepAlt
Processing comments (e.g. filtering, or improvement of the processing methods)	Some basic filtering: Moholdt et al. 2011

Table 3.7: Summary of the ICESat data used.

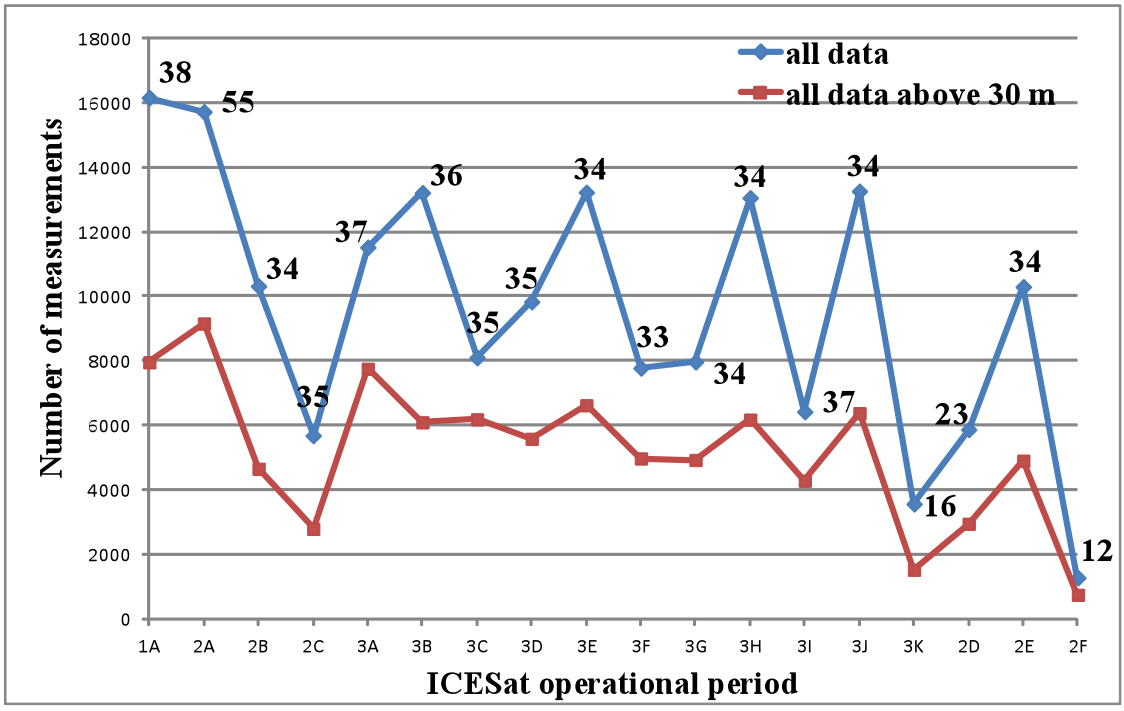

Fig. 3.8: Numbers of ICESat measurement for each operational period over Austfonna Ice Cap: the blue line represents the whole acquisition data; the red line represents all available data with measured elevations above 100 m related to the geoid; the labelled number refer to the number of operation days for each observation campaign.

3.3.2.2 Description of algorithms applied by the participants
The results received for the Austfonna Ice Cap test site have been obtained using:

- DP-RT-RepAlt
- RP-RT-RepAlt
- XO-RepAlt

Details on the two algorithms are provided in the ATBDv0 (Glaciers_cci, 2012). The DEM used when the DP-RT-RepAlt algorithm has been applied is reported in Fig. 3.9. Table 3.9 summaries some processing characteristics of the methods applied on the ICESat data and the information on additional data used.

	Participants	
	Scripps Institution of Oceanography	University College of London
Processing methods: XO-RepAlt, DP-RT-RepAlt, RP-RT-RepAlt, TP-RT-RepAlt	DP-RT-RepAlt RP-RT-RepAlt	XO-RepAlt
Processing comments (e.g. filtering, or improvement of the processing methods)	Some basic filtering: Moholdt et al. 2011	
Estimated accuracy	1 m/yr: DP-RT-RepAlt 0.30 m/yr: RP-RT-RepAlt	0.50 m/yr
Spatial density	5% for both	2%
Spatial coverage	75% for both	60%
Processing time (time required or at least its order of magnitude for the algorithm to produce the elevation change at each point)	5-10 min after reading the data and selecting glacier points (i.e. the algorithm itself)	1 day
Computer specification (e.g. Processer speed and RAM)	Standard desktop computer	Standard desktop computer
Additional data (e.g. DEM for slope correction, DEM resolution)	DP-RT-RepAlt: A DEM at 50 m resolution RP-RT-RepAlt: No additional data	No additional data

Table 3.9: Summary of the processing characteristics of the algorithms used and on the additional data needed.

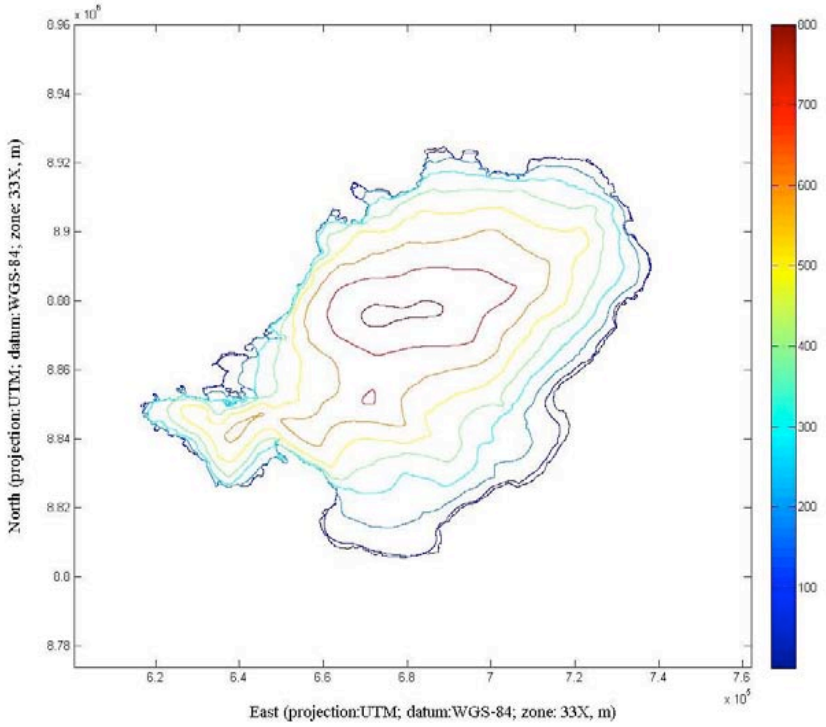

Fig. 3.9: Elevation contours of the Austfonna Ice Cap used when the DP-RT-RepAlt algorithm is applied, the iso-level curves correspond to increment of 100 m.

3.3.2.3 Comparison of the applied algorithms

Figure 3.10a and b reports the *dh/dt* trend obtained using the DP-RT-RepAlt and the RP-RT-RepAlt algorithm, respectively. Figure 3.10c shows the histograms of the occurrence of the values assumed by the two trends. From these histograms it can be noted that the number of valid trend measurements is higher in the case where the DP-RT-RepAlt algorithm is applied, presumably due to low-availability of data for computing local surface slopes when applying the RP-RT-RepAlt algorithm. However, the *dh/dt* mean values between the results of the two methods are of the same order: 0.20 m/yr in the case of the DP-RT-RepAlt algorithm and 0.125 m/yr in the case of the RP-RT-RepAlt algorithm. The panels d), e) and f) report the result of the XO-RepAlt algorithms applied considering three different retracking algorithms (landice1, landice2, landocean). Figure 3.11 shows the trend of the *dh/dt* values obtained from the airborne data versus the elevation.

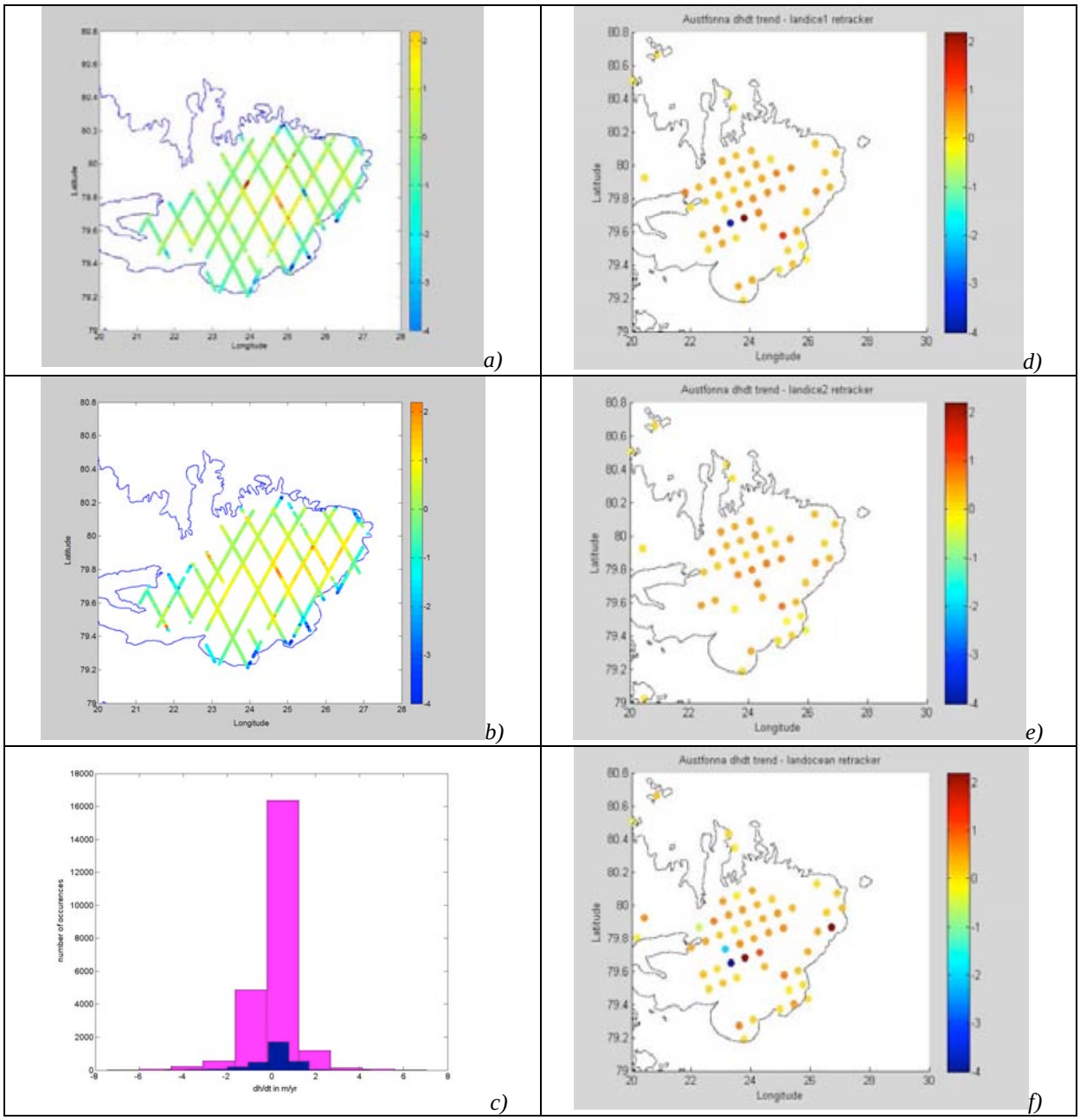

Fig. 3.10: a) dh/dt (m/yr) obtained using the full ICESat/GLAS archive and the DP-RT-RepAlt algorithm; b) dh/dt (m/yr) obtained using the full ICESat/GLAS archive and the RP-RT-RepAlt algorithm; c) histogram of the dh/dt occurrences obtained using: the DP-RT-RepAlt algorithm (purple) and the RP-RT-RepAlt algorithm (blue); d) dh/dt (m/yr) obtained using the Envisat data in the period 2002-2010 and the landice1 retracking algorithm; e) dh/dt (m/yr) obtained using the Envisat data in the period 2002-2010 and the landice2 retracking algorithm; f) dh/dt (m/yr) obtained using the Envisat data in the period 2002-2010 and the landocean retracking algorithm.

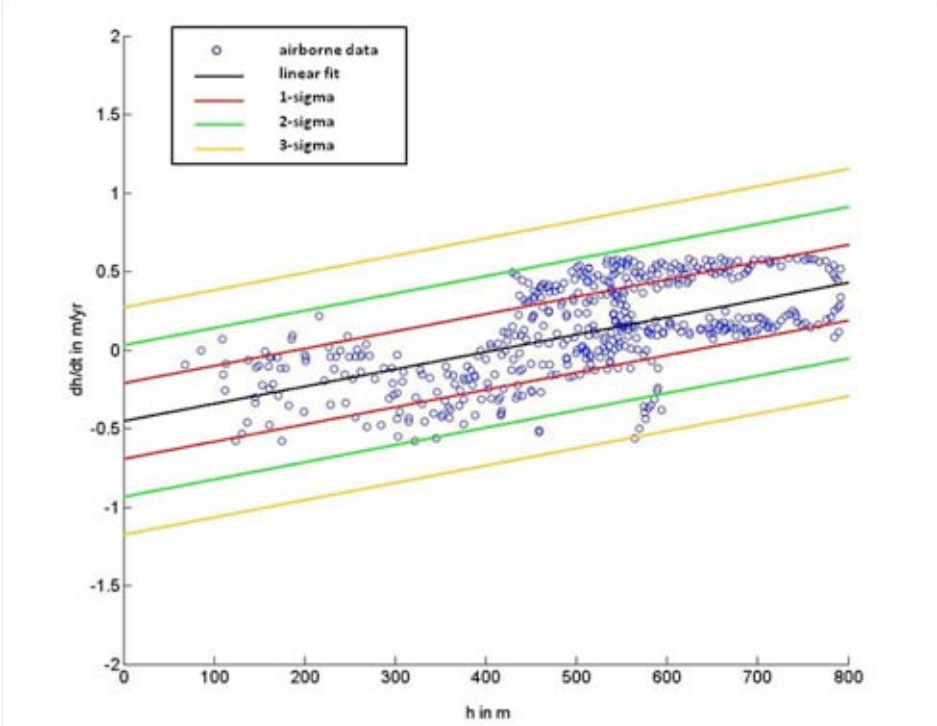

Fig 3.11: dh/dt (in m/yr) obtained from airborne data versus elevation h (in m); black line is the linear fit, whereas the red, green and orange lines represent the displacement from the linear fit of 1-σ, 2–σ and 3-σ, respectively.

The trends of the *dh/dt* versus elevation *h* for the two repeat track algorithms applied (DP-RT-RepAlt and RP-RT-RepAlt) and for the cross-over method (XO-RepAlt), are shown in Fig. 3.12a, b and c, respectively. The *dh/dt* has been averaged over elevation intervals of 100 m width. The figures include the error bars representing the standard deviation referred to the mean value obtained in each elevation bin. The overlapped lines representing the airborne linear fit and the displacement of 1-σ, 2-σ and 3-σ around it indicate that the averaged values obtained from the repeat track algorithms fall in the dynamic range interval of ±2-σ. This result is a sign of a good agreement between the repeat track algorithms applied and the airborne data. The averaged values obtained using the cross-over method still fall in the same dynamic range interval. However, this is not the case for the average values corresponding to the first elevation bins. The fact that these values are outside the 3-σ interval around the linear fit of the airborne data confirms the fact that the radar altimeter data are not measuring properly the elevation in correspondence of the ice cap margin where the degree of the slope is high. In addition, Table 10 reports the average and the standard deviation values obtained over the elevation intervals of 100 m width for the three methods applied. The last three columns report the absolute differences between the average values of the DP-RT-RepAlt, the RP-RT-RepAlt and the XO-RepAlt algorithms and the average values of the airborne dataset, respectively. In the first case, the DP-RT-RepAlt shows a maximum difference with the average values of the airborne data of about 0.6029 m/yr in the elevation band representing elevations from 100 to 200 m. In the same elevation interval, both the RP-RT-RepAlt and the XO-RepAlt algorithms show their maximum difference of 0.6594 m/yr and 0.792 m/yr, respectively.

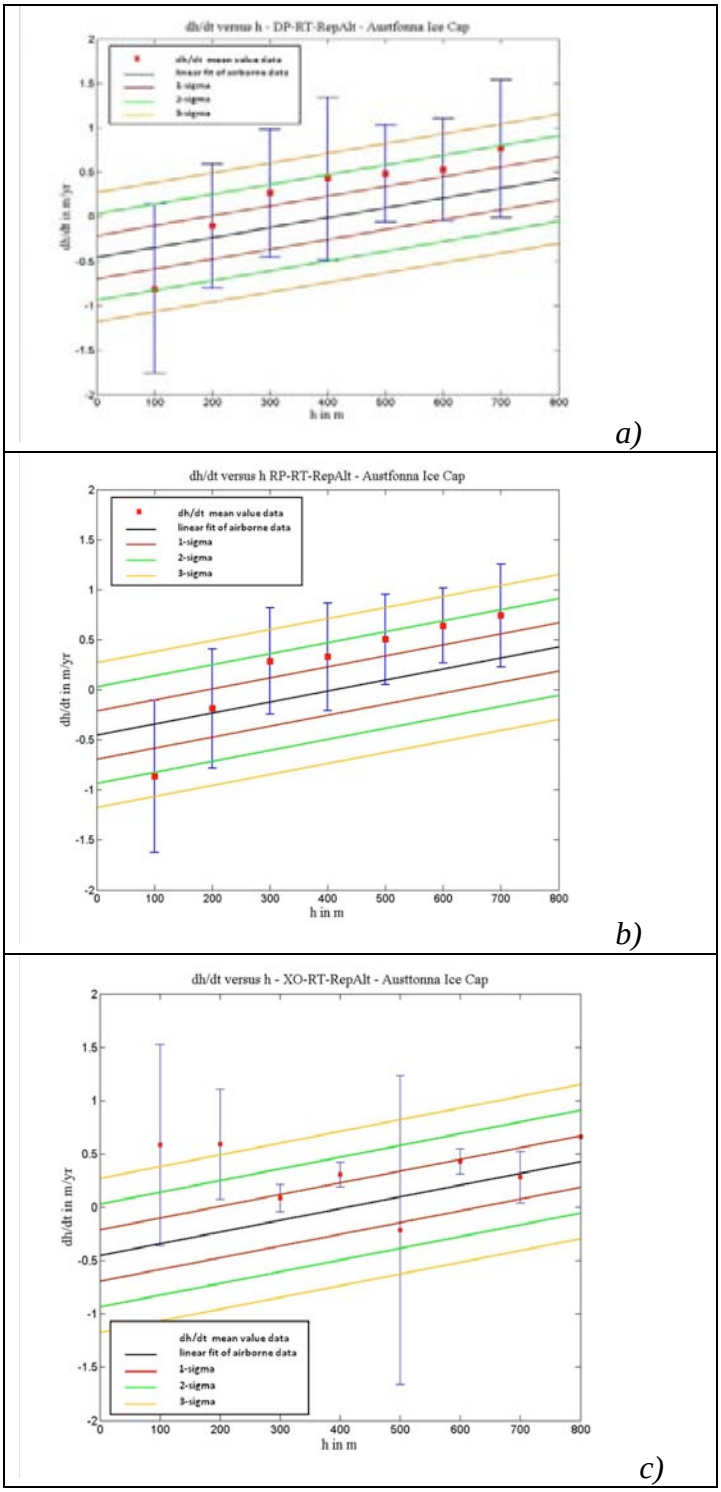

Fig. 3.12: dh/dt versus the elevation h obtained using: a) the DP-RT-RepAlt algorithm; b) the RP-RT-RepAlt algorithm; c) the XO-RepAlt algorithm; in all the panels, the dh/dt has been average over elevation intervals of 100 m width; the error bars represent the standard deviation referring to the mean value obtained in each elevation bin. The overlapped lines represent the airborne data linear fit, in black, and the displacement of 1-σ, 2-σ and 3-σ around it, in red, green and orange, respectively.

h (m)	dh/dt (m/yr) DP-RT-RepAlt (1)	dh/dt (m/yr) RP-RT-RepAlt (2)	dh/dt (m/yr) XO-RepAlt (3)	dh/dt (m/yr) airborne	|Δ| (m/yr) (1)	|Δ| (m/yr) (2)	|Δ| (m/yr) (3)
0 - 100	----	----	----	----	----	----	----
100 - 200	-0.8105±0.9495	-0.8670±0.7598	0.5844±0.9428	-0.2076±0.1949	0.6029	0.6594	0.792
200 - 300	-0.1048±0.6956	-0.1885±0.5958	0.5925±0.5156	-0.1329±0.1653	0.0281	0.0556	0.4596
300 - 400	0.2656±0.7179	0.2872±0.5310	0.0872±0.1285	-0.2838±0.1487	0.5494	0.571	0.371
400 - 500	0.4273±0.9135	0.3284±0.5348	0.3089±0.1175	0.0751±0.2671	0.3522	0.2533	0.2338
500 - 600	0.4839±0.5499	0.5032±0.4514	-0.2126±1.4501	0.2026±0.2654	0.2813	0.3006	0.4152
600 - 700	0.5275±0.5748	0.6394±0.3742	0.4304±0.1192	0.3359±0.1881	0.1916	0.3035	0.0945
700 - 800	0.7637±0.7755	0.7406±0.5135	0.2831±0.2409	0.3361±0.1827	0.4276	0.4045	0.053

Table 3.10: average and standard deviation values of the dh/dt obtained on elevation intervals of 100 m width for the DP-RT-RepAlt, the RP-RT-RepAlt, the XO-RepAlt algorithms and the airborne dataset; the last three columns report the absolute differences between the average values of the DP-RT-RepAlt, the RP-RT-RepAlt and the XO-RepAlt algorithms applied and the average values of the airborne dataset, respectively.

In Fig. 3.13 the scatter-plot between the *dh/dt* output datasets obtained applying the two repeat track methods (DP-RT-RepAlt and RP-RT-RepAlt) and the scatter-plots between each repeat track algorithm against the cross over method (XO-RepAlt) are reported. It clearly shows the strong agreement between the two repeat track algorithms. The good correlation between the two algorithms (R=0.894) and the RMSE of 0.41 m/yr indicate that the two compared algorithms are almost equivalent. Different is the result of the comparison between the two repeat track methods and the cross over one. In both cases the RMSE is more than 1 m/yr and the correlation coefficient assumes a small value close to zero when the cross-over method is compared with the DP-RT-RepAlt and a small negative value close to zero when compared with the RP-RT-RepAlt, indicating that the there is no correlation between the two methods. This result can be explained by the fact that the number of cross over points is relative small when compared with the elevation measurements used in the two repeat track algorithms. In addition it shows that the radar altimeter performs poorly over steep regions of the ice cap. Finally, in Fig. 3.14 the scatter-plots between the results obtained using the three different algorithms and the airborne *dh/dt* values are reported. Also from this figure it is evident that the repeat track algorithms applied using laser airborne data perform better than the cross-over method applied using radar altimeter measurements.

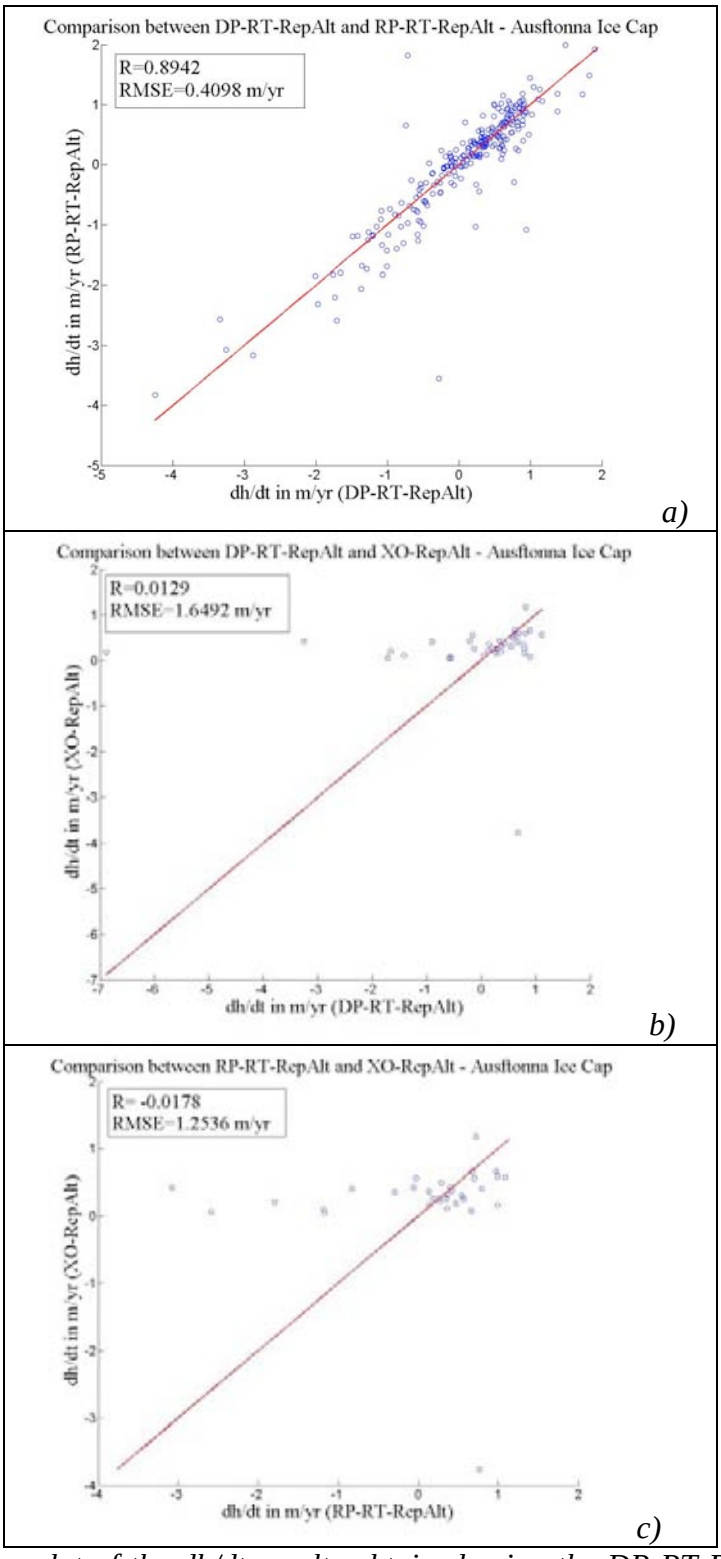

Fig. 3.13: a) scatter-plot of the dh/dt results obtained using the DP-RT-RepAlt and the RP-RT-RepAlt algorithms; b) scatter-plot of the dh/dt results obtained using the DP-RT-RepAlt and the XO-RepAlt algorithms; c) scatter-plot of the dh/dt results obtained using the RP-RT-RepAlt and the XO-RepAlt algorithms.

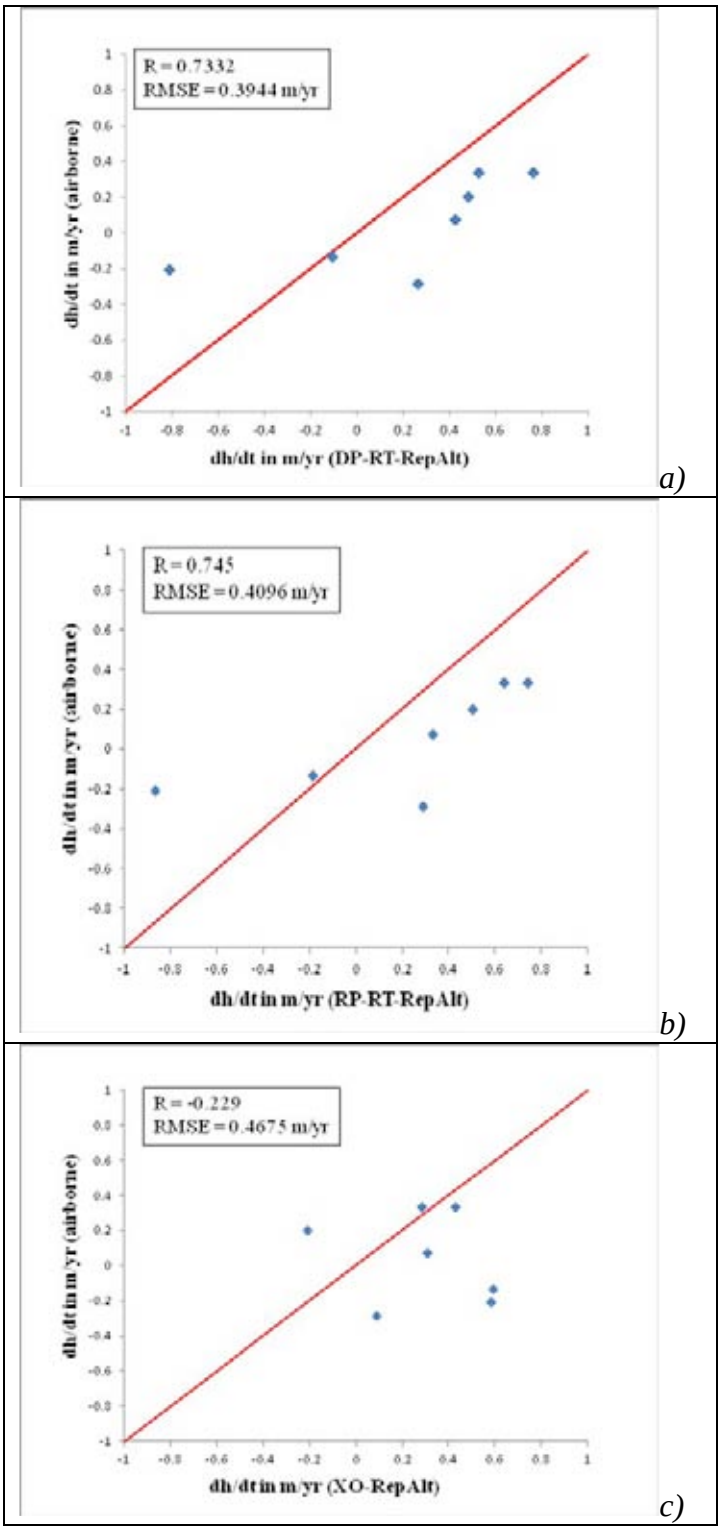

Fig. 3.14: Scatter-plot of the airborne dh/dt results versus: a) the dh/dt results obtained using the DP-RT-RepAlt; b) the dh/dt results obtained using the RP-RT-RepAlt; c) the dh/dt results obtained using the XO-RepAlt.

3.3.2.4 Comparison summary

In the following table we summaries the performances of the algorithms applied at the Austfonna Ice Cap in terms of:

- spatial density;
- spatial coverage;
- temporal density;
- CPU and manpower time;
- accuracy.

These characteristics will inform the overall algorithm selection criteria (see section 3.4).

	Algorithms		
	DP-RT-RepAlt	**RP-RT-RepAlt**	**XO-RepAlt**
Spatial density	5%	5%	2%
Spatial coverage	75%	75%	60%
Temporal density	1.6 $year^{-1}$	1.6 $year^{-1}$	10 $year^{-1}$
CPU	1 h	1 h	1 day w
Manpower time	0.5 day	0.5 day	1 week
Accuracy	1 m/yr	0.3 m/yr	0.5 m/yr

Table 3.11: Algorithm performances over Austfonna Ice Caps.

3.2.3. EC3 – Antarctica Peninsula

A round robin experiment was also defined for the Antarctic Peninsula region to assess the capability of conventional pulse-limited radar altimetry to deliver measurements of surface elevation change (SEC) in a region of Antarctica where the terrain was challenging. Although no independent validation data were available for this region, the round robin is of value in assessing the capability of the RA dataset to retrieve measurements of SEC.

3.2.3.1 Input data

Envisat data have been used over this test site, which cover the periods from October 2002 to December 2010.

3.2.3.2 Description of algorithms applied by the participants

Over this test site the XO-RepAlt algorithm has been applied from the participants. In particular, the external participant used the method reported in Khvorostovsky (2012), and the internal participant used the method reported in Wingham et al. (2006). A slope correction has not been applied to either datasets, due to the fact that DEMs present large errors over the steeper and rougher ice cap margins of the Antarctic Peninsula which, if applied, would have caused even greater levels of data to be discarded from the analysis.

3.2.3.3 Comparison of the applied algorithms

Figure 3.15a shows the Peninsula drainage basins and Fig. 3.15b and c report the *dh/dt* spatial distribution obtained by the two participants using the same methods. The *dh/dt* values (in cm/year) have been estimated for grids of 0.5 degrees latitude and 1 degree longitude.

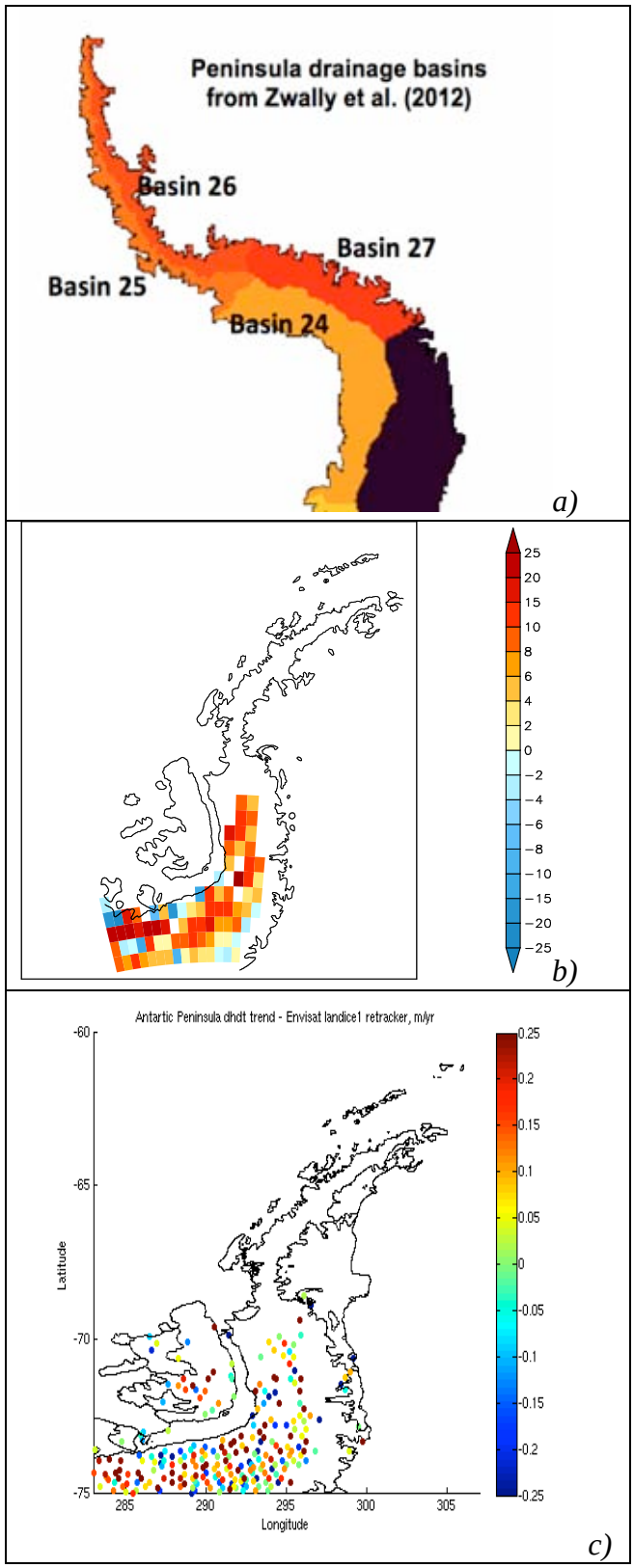

Fig: 3.15: a) Peninsula drainage basins; b) dh/dt spatial distribution obtained using the XO-RepAlt by the Nansen Environmental and Remote Sensing Centre (NERSC) participant; c) dh/dt spatial distribution obtained using the XO-RepAlt by the UL participant.

It is clear from the results that elevation rates are only derived for the inland basins (basins 24 and 27), where the surface is characteristically of low slope and similar to the interior of an ice sheet or ice cap. Very few elevation rates are derived for the northerly drainage basins (basins 25 and 26). In total, the spatial density of elevation rates acquired in basins 24, 25, 26, and 27 was 14%, 0.3%, 0.2% and 3%, respectively. Only one basin (basin 24) exhibits a relatively low drop out of date when compared to the original measurements acquired by the satellite altimeter. This is because much of the basin includes Dyer Plateau, a region of low slope, and this can be seen as an analogue for very flat ice caps. However, even in this basin the sampling is quite sparse (only 10-15 %) due to the latitude, and so validation of the results is an important next step in the process before we can assess viability.

No independent validation data are available for this test site, and so it is not possible to assess the absolute accuracy of the data. However, the distribution of the derived elevation rates allows us to assess the potential value of the radar altimeter data in the context of an end-to-end data processing system. From this experiment over the Antarctica Peninsula we are able to conclude that conventional, pulse-limited radar altimetry performs poorly over the Antarctic Peninsula due to a combination of the sparse ground tracks (about 25 km separation) and data drop out (>90 % for 3 of the 4 basins). This region can be seen as an analogue for mountain glaciers, which suggests that conventional radar altimetry should not be used there.

3.4 Selection criteria

We have reported the results of the validation activity where the elevation change trends derived from satellite repeat altimetry have been compared using: 1) different algorithms, cross-over and repeat track, and different sensors, laser and radar altimeter products, and using 2) independent airborne measurements. The comparison has been made on the absolute differences of the elevation change maps, and the RMSE together with the correlation coefficient R, have been computed to provide a quantitative analysis. The results of these quantities are summarised here for simplicity in Table 3.12 for both the inter-comparison between algorithms and/or sensors and the comparison with airborne data. In there, the repeat track algorithms, i.e. the DS-RT-RepAlt, the DP-RT-RepAlt and the RP-RT-RepAlt, have been applied using laser altimeter data, whereas the cross-over method, the XO-RepAlt has been applied using radar altimeter data.

As it can been seen, the DP-RT-RepAlt and the RP-RT-RepAlt algorithms show a RMSE of about 0.4 m/yr when compared with the airborne data indicating a good agreement and a result comparable with the estimated measurement accuracy. The good agreement with airborne data is confirmed also by the good value of the correlation coefficient (around 0.73). In addition, the two repeat track algorithms demonstrate a RMSE of the same order when inter-compared and a really high correlation coefficient (0.89), indicating that they are almost equivalent. Also the DS-RT-RepAlt algorithm shows a RMSE of the same order when compared with the RP-RT-RepAlt algorithm. Thus, it can be concluded that the three repeat algorithms investigated are almost equivalent.

With regard to the comparison between the elevation changes derived with satellite altimetry and the airborne measurements, the absolute differences obtained using the DP-RT-RepAlt, the RP-RT-RepAlt and the XO-RepAlt are plotted in Fig. 3.16. They show similar trends when the elevation becomes greater than 300 m. For elevation below this value, a dissimilarity between the absolute differences can be noted when the radar data are used, indicating that the radar performs better on regions with moderate slope than regions with high slope like the ice margin. Unfortunately, at these regions the validation dataset is unlikely to be representative.

RMSE (m/yr)/R	DS-RT-RepAlt (laser alt.)	DP-RT-RepAlt (laser alt.)	RP-RT-RepAlt (laser alt.)	XO-RepAlt (radar alt.)	airborne
DS-RT-RepAlt (laser alt.)	---	N/A	0.4726/0.4802	N/A	N/A
DP-RT-RepAlt (laser alt.)	---	---	0.4098/0.8942	1.6492/0.0129	0.3944/0.7332
RP-RT-RepAlt (laser alt.)	---	---	---	1.2536/-0.0178	0.4096/0.745
XO-RepAlt (radar alt.)	---	---	---	---	0.4675/-0.229
airborne	---	---	---	---	---

Table 3.12: Summaries of the RMSE and the correlation coefficient R obtained from the validation activity.

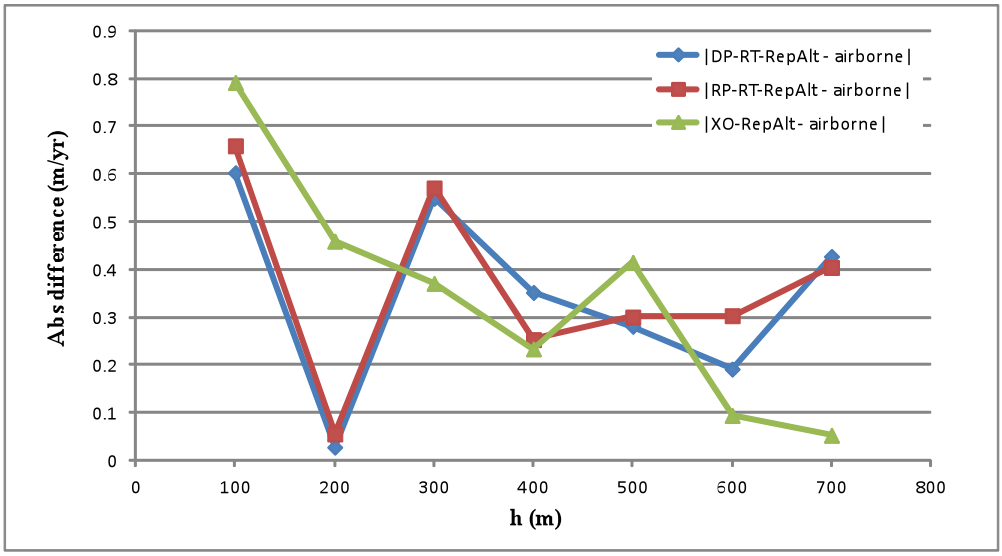

Fig. 3.16: Absolute differences in m/yr as a function of the elevation between airborne dh/dt and DP-RT-RepAlt dh/dt (blue); airborne dh/dt and RP-RT-RepAlt dh/dt (purple); airborne dh/dt and XO-RepAlt dh/dt (green).

The following general criteria are used for the selection of the best performing algorithm:

- spatial density of satellite derived SEC
- spatial coverage of satellite derived SEC
- temporal density of satellite derived SEC
- absolute accuracy relative to validation data
- processing time and manual interaction

Table 3.13 summarises the overall performance of the tested algorithms in relation to these selection criteria. The algorithm performances in relation to spatial and temporal density are detailed in Table 3.6 and Table 3.11. Additional information on the algorithm performance in relation to spatial density is provided in section 3.2.3. The algorithm performance in relation to absolute accuracy is detailed in Table 3.12.

	DS-RT-RepAlt	DP-RT-RepAlt	RP-RT-RepAlt	XO-RepAlt
RMSE	Good	Good	Good	Poor
Spatial density	Good	Good	Good	Poor
Temporal density	Moderate	Moderate	Moderate	Good
Processing time	Good	Good	Good	Moderate
SCORE	11	11	11	7

Table 3.13: Summary of the ice cap SEC algorithm performance in relation to the selection criteria.

From the algorithm comparison we draw the following **main conclusions**:

- The XO-RepAlt method provides ice cap SEC data of significantly higher temporal resolution than the RT-RepAlt methods, but both provide estimates of ice cap SEC using comparable levels of CPU and manpower.
- The RT-RepAlt methods provide estimates of ice cap SEC with significantly finer spatial sampling and greater accuracy than the XO-RepAlt method.
- Overall, the RT-RepAlt methods are selected as the most reliable technique for developing satellite based observations of ice cap SEC. Within this class of methods, there was no significant difference in the performance of the various SEC algorithms.

4. Elevation changes from DEM differencing

4.1 Assessment of algorithm performance

4.1.1 Influences on the accuracy of glacier elevation changes from DEM differencing

The reliability of glacier elevation changes derived from comparison of multi-temporal DEMs is influenced by the individual accuracies, precisions and resolutions of the DEMs to be differenced, the combined co-registration of the DEMs, and the resampling required to merge the DEMs into a single grid of elevation differences. DEM accuracy is dependent upon the data acquisition techniques used, mainly photogrammetric principles on optical images (i.e. aerial, ASTER or SPOT), interferometric techniques on repeat radar images (i.e. SRTM), or laser distance point clouds of measurements (i.e. LIDAR DEMs) and partly also the environmental conditions at the time of acquisition. In addition, the resolutions of the products from these techniques vary considerably depending upon whether data is acquired from the air or space. A number of studies have outlined various accuracies for the different DEMs and elevation data types (Kääb, 2005; Fricker et al., 2005; Rodriguez et al., 2006; Berthier et al., 2007; Toutin, 2008) mainly by comparison to other DEMs or measurements of elevation (i.e. GNSS, ICESat). The common approach is for comparison over terrain known or assumed to have not changed. Glaciers_cci will follow this standard for product validation and algorithm selection.

The comparison of two or more multi-temporal DEMs require that the models be horizontally and vertically aligned (*co-registered*) to ensure that multi-temporal pixels represent the same location on the Earth's surface. Methods for co-registration range from manual translations (VanLooy, 2011) to automated algorithms that minimize elevation residuals (Gruen and Akca, 2005; Schenk et al., 2005; Berthier et al. 2007; Miller et al. 2009; Nuth and Kääb, 2011). The round robin aimed at testing the co-registration approaches in search of the most reliable, robust and universal algorithm. An important consideration in terms of co-registration, DEMs of varying resolutions (pixel areas) depict different elevations at the same pixel centre location depending upon the acquisition technique (radar, lidar, photogrammetry) with the characteristics of the terrain (i.e. vegetation, surface roughness, visible contrast, material etc.) at the time of acquisition. Recent studies have emphasized the influence of varying DEM resolutions and resampling strategies on elevation-dependent biases detected within DEM differences (Paul, 2008; Gardelle et al., 2012). The datasets chosen for product validation and algorithm selection have varying resolutions to further investigate resampling and topographic effects on DEM difference accuracies.

Finally, the detection of significant glacier elevation changes is not only a function of DEM accuracy, but largely a function of time and the particular characteristics of the glaciers being measured in the environments they reside. Therefore, the data availability and the time span between DEMs have a major impact on glacier elevation change reliability. Choice of data is an important manual interaction step necessary to provide quality data products.

4.1.2 Algorithm development and product validation strategy

The first step to produce glacier elevation changes from multi-temporal DEMs is to ensure their horizontal and vertical co-registration, while the second step involves resampling for matrix differencing. Glaciers_cci focuses on the development of a universal co-registration

routine for consistent application in any assessment. In terms of product accuracy and validation, resampling techniques are also analyzed through the round robin.

The validation of glacier elevation changes from DEM differencing may be performed through external validation requiring two pairs of DEMs acquired at the same time, preferably of higher and lower accuracy/resolution, respectively. For most glaciers and glacier regions, coincident availability of both satellite and higher-accuracy DEMs is impossible. Therefore, external validation of glacier elevation changes is not practically obtainable and excluded from further investigation here. However, availability of a single high resolution/accuracy DEM (e.g. airborne LIDAR DEM) provides an alternate external validation strategy by comparing the individual satellite DEMs with higher accuracy and precision for validating the raw elevations rather than the elevation changes. If the datasets are temporally consistent, they may also be compared over glaciers. Most cases, however, involve comparisons over stable terrain (e.g. temporally constant; off-glacier) requiring the assumption that off-glacier elevation models are coherent in accuracy to on-glacier models. The validation plan for situations when high accuracy data are not available involves comparison between three or more datasets, also using stable terrain outside of the glaciers.

In this algorithm comparison, the focus of the round robin experiment is on selection and development of a universal co-registration routine for consistent application in any assessment. Therefore, the major criteria for algorithm selection are robustness, reliability and accuracy. The round robin data package was specifically designed to contain DEMs of varying resolution as well as alternate data types such as elevation profiles from the ICESat campaign. The DEM data was provided in a standard format, GeoTiff in which the pixel definition (cell centre vs. cell corner) is specifically defined with the GeoTiff TAGS.

4.2 Methods for accuracy determination

4.2.1 External Validation

The external validation involves verification of a satellite DEM using a high accuracy DEM. The process begins with co-registration of the datasets, and second resampling of the one dataset to the other. If the data is temporally consistent, the comparison can be made over the glacier to analyze/detect any glacier specific biases related to the acquisition strategy (i.e. radar wave penetration into snow/firn within the SRTM DEM). In most cases, however, the analysis will only be made over stable terrain (off-glacier terrain). After co-registration, visual analysis of the changes over stable terrain is performed to detect any internal scene biases that may exist. If detected, procedures for removal will be investigated. This step is however case-study specific and therefore cannot be universally standardized such as the co-registration pre-processing step.

Alternate components of the external validation involve comparing the topographic attributes such as mean, minimum and maximum glacier elevation and derivatives of slopes and aspects over glaciers (Frey and Paul, in press). For this purpose, the variability and mean difference between topographic parameters of the medium resolution DEMs (SRTM, ASTER, ASTER GDEM, etc.) and the high resolution DEMs have to be calculated. Of further importance is to understand how potential elevation changes are affecting the topographic parameters, especially when incorporating them into the glacier database (i.e. GLIMS).

4.2.2 Internal Validation

Internal validation refers to the quantification of random and systematic errors in elevation differences between the DEMs using the DEMs themselves. In principle, this is not different from the external validation, except that there may not be one DEM that is more accurate than the others. The two multi-temporal DEMs are first co-registered to remove a potential systematic linear bias (horizontal and vertical). A third elevation dataset is then co-registered to the other two DEMs. This may be elevation profiles acquired by satellite laser altimetry (i.e. ICESat) simply another DEM. Each of the individual DEMs is then co-registered to the third dataset. This process returns three co-registration vectors between the data products that should form a perfect triangle (vector sum). Any mismatch in a vertex of the triangle is an estimate of the remaining un-removed bias. In practice, this un-removed linear vertical bias can be added to the error budget for elevation change. The resulting parameters of internal validation are an estimate of the random error and an un-removed linear vertical systematic bias.

4.2.3 Processing protocol for algorithm assessment and product validation

The algorithm assessment and product validation is based upon the 'Internal Validation' strategy (4.2.2) and the triangulation of co-registration vectors. DEM co-registration refers to the horizontal and vertical linear adjustments from one DEM to another DEM or elevation data source (i.e. laser altimetry points). The general procedures for co-registration are:

- *Pre-Processing*: Selection of suitable terrain for input into the DEM co-registration routine. Suitable terrain is defined as stable and unchanging, i.e. outside glaciers, no water bodies, no clouds or data voids and assuming limited vegetation effects. This terrain is further used for both determining the co-registration parameters and later for validation of products and accuracy quantification. This choice of stable terrain is an implementation procedure that leads to variations in the accuracy.
- *Co-Registration*: DEM co-registration following algorithms described in the ATBDv0 (Glaciers_cci, 2012b) as well as algorithms applied/introduced by the round robin participants. Adjustment of the slave DEM to the master DEM using the co-registration parameters.
- *Resampling*: Resample one DEM to another, and matrix difference the two aligned DEMs. Use the selected stable terrain for product validation and accuracy assessment. Estimation of random and systematic uncertainties in the final DEM difference grids.

4.3 Comparison of the applied algorithms

4.3.1 Description of algorithms applied by the participants of the round robin exercise

Co-Registration using slope and aspect [**CR-SA**]

This algorithm is based upon the analytical relation that elevation differences of 2 mis-registered DEMs have with the direction and magnitude of the mis-registration vector and the terrain derivatives of slope and aspect, described in detail within the ATBDv0 (Glaciers_cci, 2012b) and in Nuth and Kääb (2011). The algorithm operates by minimizing residuals to a cosine function. Implementation of the algorithm may vary, including resampling procedures (if the DEMs are of difference pixel sizes), the measure for minimization and the thresholds for exiting the iterative process. Figure 4.1 shows the first and final co-registration between the ASTER 2006 with the SRTM and ICESat. Figure 4.2 shows the histograms of elevation differences before and after co-registration as well as their "false-hillshade" colored grids.

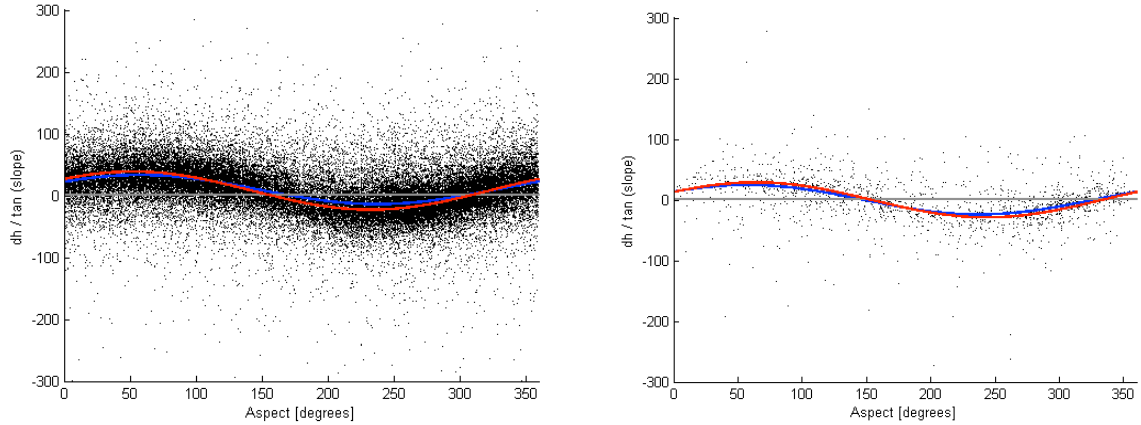

Fig. 4.1: Co-registration between the 2006 ASTER DEM and the SRTM [left] and between the 2006 ASTER DEM and ICESat [right]. The blue line is the algorithm solution after the 1st iteration and the red line is the algorithm fit after the 3rd and final iteration. ...

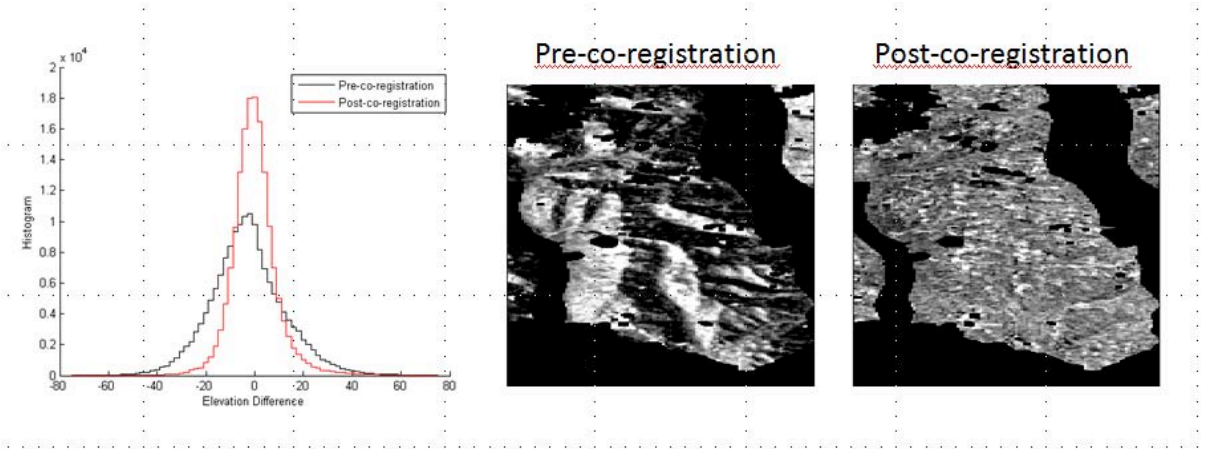

Fig. 4.2: Histograms of elevation differences over stable terrain for pre- and post- co-registration of the SRTM and 2006 ASTER DEM using the **CR-SA** *algorithm [left]. "False-hillshade" appearance of a subset of the elevation differences before co-registration of the same DEMs [centre] and the lack of "false-hillshade" effects after co-registration [right].*

Co-registration by image correlation [**CR-IC**]

An alternate method to determine the horizontal co-registration between two DEMs is using well-defined image correlation methods for the full DEMs (Rodriguez et al., 2006, Berthier et al., 2007, Tachikawa et al., 2011). In this 2-parameter approach, the x- and y- coordinates of the slave DEM are shifted around the original centre location and then differenced to the master DEM. Statistics such as the correlation coefficient, standard deviation or the root mean square error (RMSE) are used to determine the final slave DEM location in respect to the master (Fig. 4.3). The vertical adjustment is then obtained by removing an estimate of the mean difference between the co-registered datasets over stable terrain, sometimes only over low-sloped terrain.

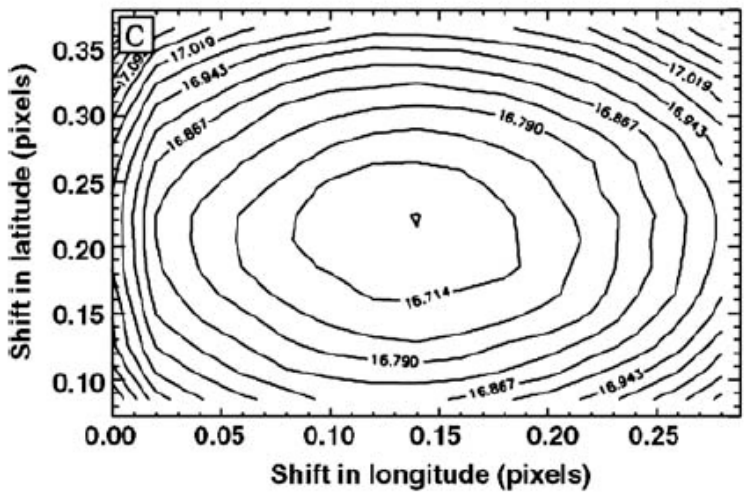

Fig. 4.3: Contour plot of the standard deviation of elevation differences over stable terrain after iteratively shifting in the latitude and longitude directions. The lowest residual ("bullseye") denotes the horizontal translation parameters in latitude and longitude as a result of **CR-IC** *(Fig. adapted from Berthier et al. 2007).*

Co-registration by robust surface matching [**CR-RSM**]
This method for co-registration is an advanced version of the **CR-IC** approach but uses a 7 parameter conformal function to solve for 3 translations (*Tx*, *Ty*, *Tz*), 3 rotations (ω, ϕ, $\varkappa$) and a scale factor (*s*) in a robust linear least squares manner (Miller et al., 2008 and 2009). The solution is derived by minimizing the Euclidean distance between two corresponding surfaces (Gruen and Akca, 2005). Similarly, vertical residuals between the DEMs on assumed stable terrain are used for minimization.

Co-registration by manual translation [**CR-MA**]
This co-registration technique is based upon the visual interpretation between DEMs (VanLooy, 2011). Various visualization tools can be used to detect horizontal translations between DEMs, including false- hillshades (Nuth and Kääb, 2011), or elevation profiles across the DEMs (Schenk et al., 2005). This technique contains little to no automation and is thus fully user dependent.

4.3.2 Intercomparison of the different algorithms

The algorithms for DEM co-registration rely on minimizing residuals between the DEMs on assumed stable terrain. The **CR-SA** algorithm solves the analytical solution to a mis-registration by exploiting the relationship between elevation differences, the direction of the terrain (aspect) and the direction of the mis-registration. Statistical approaches (i.e. robust least squares minimization) are used to solve the cosine-based algorithm. Thus, the co-registration parameters and their errors are directly returned from the algorithm solution. The **CR-IC** and **CR-RSM** algorithms rely on the statistical minimizations of the elevation residuals using a linearly combined 2- or 7-parameter function, respectively. The **CR-SA** algorithm minimizes the elevation residuals normalized by the slope tangent. The **CR-IC** and **CR-RSM** algorithm minimizes the raw elevation differences and may therefore be more sensitive to the slope distribution of the terrain. It has to be noted that internal pixel-to-pixel

horizontal location errors (i.e. from the instrument pointing accuracy) translate into vertical errors multiplied by the slope tangent.

The three co-registration algorithms described above can be implemented in various ways using different software products. For all algorithms, iteration is required to determine the final solutions. When to terminate the iterative process and according to what criteria will vary for different implementations and thus also the number of iterations. For **CR-SA**, 1-4 iterations are commonly required to achieve convergence, whereas **CR-IC** requires exponentially more iterations depending upon the parameter step size. The number of iterations required for the **CR-RSM** to converge on a solution is also dependent upon the minimizing parameter step size, and has been previously optimized to 5-6 iterations (Gruen and Akca, 2005). The software used in the round robin include Matlab, IDL, ENVI, PCI Geomatica, GDAL, ArcGIS and Excel. In some cases, only one software was used throughout the implementation and derivation of elevation changes. In other cases, multiple products were used for a single implementation which complicates the round robin results.

4.4 Statistical analysis of the results

4.4.1 Round robin results – New Zealand

The round robin experiment contains 6 participant results. Two of those implemented the **CR-SA** algorithm, 2 implemented the **CR-IC** algorithm, 1 implemented the **CR-RSM** algorithm and 1 participant performed a manual co-registration (**CR-MA**). Table 4.1 through 4.6 present the results from each contributor with the horizontal and vertical co-registration parameters between the DEMs, and the triangulation (vector sum) of these vectors. The co-registration parameters for the same datasets vary significantly between the contributions. This effect derives from a variation in pixel coordinate definition (pixel centre vs. pixel corner) within the processing software(s) used. In principle, it makes no difference as long as the definition is consistent throughout an individual processing routine. To determine whether the implementations are internally correct, the triangulation of co-registration vectors (the bottom halves of tables 4.1-4.6) can be analyzed. The round robin contributions in which the triangulation residuals were less than 1/10 an SRTM pixel includes those from the **CR-SA** algorithm implemented by the consortium (Table 4.1), the **CR-IC** algorithm implemented only on DEMs (Table 4.2; i.e. ASTER2002 - ASTER2006 - SRTM) and similarly for the **CR-RSM** algorithm applied to the DEMs (Table 4.5).

For the **CR-SA** method implemented by the consortium (in Matlab), the accuracy between the 4 round robin datasets is maximum 5 m horizontally and less than 1 m vertically (Table 4.1). A second implementation of the **CR-SA** (Table 4.3) resulted in different co-registration parameters for the datasets, and large triangulation residuals suspected to result from varying pixel coordinate definition within the implementation. The **CR-IC** approach resulted in small triangulation residuals (Table 4.2) despite significantly different co-registration parameters as compared to the **CR-SA** method (Table 4.1). The triangulation between the three DEMs (ASTER2002-ASTER2006-SRTM) using the **CR-IC** algorithm resulted in a horizontal residual less than 3 m and a vertical residual of ~2 m (Table 4.2). The vertical residual is larger than other methods and is suspected to be the result of varying slope distribution of the underlying elevation difference samples; i.e. the **CR-SA** algorithm normalizes for these slope distribution variations while the **CR-IC** algorithm relies solely on the raw samples. The **CR-**

RSM algorithm resulted in horizontal and vertical co-registration residuals of less than a metre for the DEM analysis, but is on the order of 10-15 m when ICESat is included.

	Datasets		ΔX	ΔY	ΔZ	$E_{\Delta X}$	$E_{\Delta y}$	$E_{\Delta z}$	Duration (s)
Co-Registration	SRTM	ICEsat	-1.71	-3.93	-3.80	7.08	6.55	1.45	2
	ASTER2006	ICESat	25.78	12.68	-0.25	14.24	17.45	4.22	7
	ASTER2006	SRTM	24.34	20.59	3.19	1.32	1.52	0.33	23
	ASTER2002	ICESat	-16.99	-11.67	-7.96	11.51	12.22	2.94	7
	ASTER2002	SRTM	-13.41	-9.10	-4.07	0.90	0.87	0.17	23
	ASTER2002	ASTER2006	-38.38	-26.19	-7.50	0.61	0.65	0.12	47
Triangulation	ASTER2006 - SRTM - ICESat		-3.15	3.99	-0.35				
	ASTER2002 - SRTM - ICESat		1.87	-1.36	0.10				
	ASTER2002 - ASTER2006 - SRTM		-0.62	3.50	-0.25				
	ASTER2002 - ASTER2006 - ICESat		4.39	-1.84	0.20				
	ASTER2002 - ASTER2006 - SRTM - ICE-Sat		1.24	2.14	-0.15				

*Table 4.1: Co-registration of the round robin data package using **CR-SA** algorithm by the consortium. The entire implementation is processed in Matlab. In the lower half the result of the triangulation of the co-registration vector sum is depicted.*

	Datasets		ΔX	ΔY	ΔZ	$E_{\Delta X}$	$E_{\Delta y}$	$E_{\Delta z}$	Duration (s)
Co-Registration	SRTM	ICEsat	-4.0	6.1	-3.8	-	-	-	< 300
	ASTER2006	ICESat	22.1	-14.6	-2.7	-	-	-	< 300
	ASTER2006	SRTM	-5.6	-50.8	2.5	-	-	-	< 300
	ASTER2002	ICESat	-5.6	2.4	-8.4	-	-	-	< 300
	ASTER2002	SRTM	-41.1	-24.5	-4.6	-	-	-	< 300
	ASTER2002	ASTER2006	-38.0	25.4	-4.9	-	-	-	< 300
Triangulation	ASTER2006 - SRTM - ICESat		-31.75	-30.07	1.37				
	ASTER2002 - SRTM - ICESat		-39.53	-20.88	0.08				
	ASTER2002 - ASTER2006 - SRTM		-2.53	-0.81	2.15				
	ASTER2002 - ASTER2006 - ICESat		-10.31	8.38	0.86				
	ASTER2002 - ASTER2006 - SRTM - ICE-Sat		-42.06	-21.69	2.23				

*Table 4.2: Participant contribution for co-registration. Co-registrations between DEMs (ASTER2002 – ASTER2006 – SRTM) were accomplished using a **CR-IC** algorithm while co-registrations involving ICESat relied on a modified **CR-SA** algorithm. Multiple software products are used including IDL/ENVI, PCI geomatica and Generic Mapping Tools (GMT).*

The round robin participants were asked additionally to provide their elevation difference grids after co-registration. Again, when multiple software products were used, the results were sometimes contaminated which made quantitative analysis difficult. It was also not possible to simply use their co-registration parameters since these reflected the pixel definitions that their software's used. Figure 4.4 shows the 6 elevation difference grids over the assumed stable terrain using a greyscale colour definition. If a horizontal mis-registration

exists within the elevation difference grid, then the terrain becomes visible as in a DEM hillshade map. Two of the contributions show significant "false-hillshade" appearences (Fig. 4.4). These grids were created using one software product for co-registration and ESRI ArcGIS for differencing. Within the in-built ArcGIS differencing routines, the resampling procedures are by default "nearest neighbour". This basic interpolation technique does not allow for any sub-pixel adjustments, and inherently creates, in a minimum, sub-pixel mis-alignments. Larger mis-registrations can occur if the two grids are of varying resolutions. So, despite the similarity in co-registration parameters between **CR-SA** in Table 4.1 and **CR-RSM** in Table 4.5, the nearest neighbour resampling in ArcGIS destroyed the CR-RSM co-registered difference grid.

	Datasets		ΔX	ΔY	ΔZ	$E_{\Delta X}$	$E_{\Delta y}$	$E_{\Delta z}$	Duration (s)
Co-Registration	SRTM	ICEsat	49.46	25.67	-9.68	-	-	-	16
	ASTER2006	ICESat	37.61	-2.13	-0.75	-	-	-	223
	ASTER2006	SRTM	-17.33	-44.99	4.08	-	-	-	34
	ASTER2002	ICESat	-1.18	22.85	-9.59	-	-	-	189
	ASTER2002	SRTM	-30.99	-22.91	-3.95	-	-	-	18
	ASTER2002	ASTER2006	-44.34	23.93	-7.5	-	-	-	88
Triangulation	ASTER2006 - SRTM - ICESat		-5.48	-17.19	-4.85				
	ASTER2002 - SRTM - ICESat		19.65	-20.09	-4.04				
	ASTER2002 - ASTER2006 - SRTM		-30.68	1.85	0.53				
	ASTER2002 - ASTER2006 - ICESat		-5.55	-1.05	1.34				
	ASTER2002 - ASTER2006 - SRTM - ICE-Sat		-11.03	-18.24	-3.51				

*Table 4.3: Participant contribution for co-registration. Co-registration performed using a modified version of the **CR-SA** algorithm in IDL. The algorithm solution was obtained by minimizing the RMSE using an advanced optimizing routine (Duan et al., 1994).*

	Datasets		ΔX	ΔY	ΔZ	$E_{\Delta X}$	$E_{\Delta y}$	$E_{\Delta z}$	Duration (s)
Co-Registration	SRTM	ICEsat	-	-	-	-	-	-	-
	ASTER2006	ICESat	-	-	-	-	-	-	-
	ASTER2006	SRTM	-15	-56.25	-2.01	-	-	-	159
	ASTER2002	ICESat	-	-	-	-	-	-	-
	ASTER2002	SRTM	-37.5	-22.5	-8.79	-	-	-	169
	ASTER2002	ASTER2006	-52.5	33.75	-9.42	-	-	-	192
Triangulation	ASTER2006 - SRTM - ICESat		-	-	-				
	ASTER2002 - SRTM - ICESat		-	-	-				
	ASTER2002 - ASTER2006 - SRTM		-30.00	0.00	-2.64				
	ASTER2002 - ASTER2006 - ICESat		-	-	-				
	ASTER2002 - ASTER2006 - SRTM - ICE-Sat		-	-	-				

*Table 4.4: Participant contribution for co-registration. Co-registration was performed using a **CR-IC** algorithm with free IDL code from the IDL Astronomy Library:* http://idlastro.gsfc.nasa.gov.

	Datasets		ΔX	ΔY	ΔZ	$E_{\Delta x}$	$E_{\Delta y}$	$E_{\Delta z}$	Duration (s)
Co-Registration	SRTM	ICEsat	11.6	-14.4	-8.5	2.149	1.969	3.181	6180
	ASTER2006	ICESat	25.3	18.4	9.9	1.526	1.448	2.638	780
	ASTER2006	SRTM	24.4	21.8	7.6	0.042	0.048	0.018	120
	ASTER2002	ICESat	-11.6	-11.6	-3.4	1.927	1.816	3.252	7800
	ASTER2002	SRTM	-12.0	-4.6	-1.0	0.043	0.049	0.019	3900
	ASTER2002	ASTER2006	-36.3	-25.8	-8.5	0.026	0.028	0.012	1620
Triangulation	ASTER2006 - SRTM - ICESat		10.71	-11.03	-10.75				
	ASTER2002 - SRTM - ICESat		11.23	-7.50	-6.16				
	ASTER2002 - ASTER2006 - SRTM		0.09	0.60	0.13				
	ASTER2002 - ASTER2006 - ICESat		0.61	4.13	4.72				
	ASTER2002 - ASTER2006 - SRTM - ICE-Sat		11.32	-6.90	-6.03				

*Table 4.5: Participant contribution for co-registration. Co-registration using the **CR-RSM** algorithm on a selected 20 km by 20 km section of the original provided dataset.*

	Datasets		ΔX	ΔY	ΔZ	$E_{\Delta x}$	$E_{\Delta y}$	$E_{\Delta z}$	Duration (s)
Co-Registration	SRTM	ICEsat	-	-	6.2	-	-	22.3	-
	ASTER2006	ICESat	-	-	3.6	-	-	24.13	-
	ASTER2006	SRTM	43.12	77.99	1.49	32.3	41	18.61	472
	ASTER2002	ICESat	-	-	8.8	-	-	16.5	-
	ASTER2002	SRTM	34.3	26.3	5.63	65.7	56.9	29.3	502
	ASTER2002	ASTER2006	-54.6	-19.9	6.79	44.4	51.9	22.6	463
Triangulation	ASTER2006 - SRTM - ICESat		-	-	4.05				
	ASTER2002 - SRTM - ICESat		-	-	3.07				
	ASTER2002 - ASTER2006 - SRTM		-45.78	31.79	2.65				
	ASTER2002 - ASTER2006 - ICESat		-	-	1.67				
	ASTER2002 - ASTER2006 - SRTM - ICE-Sat		-	-	5.72				

Table 4.6: Participant contribution for co-registration. Co-registration was performed using a manual technique within ArcGIS and Excel.

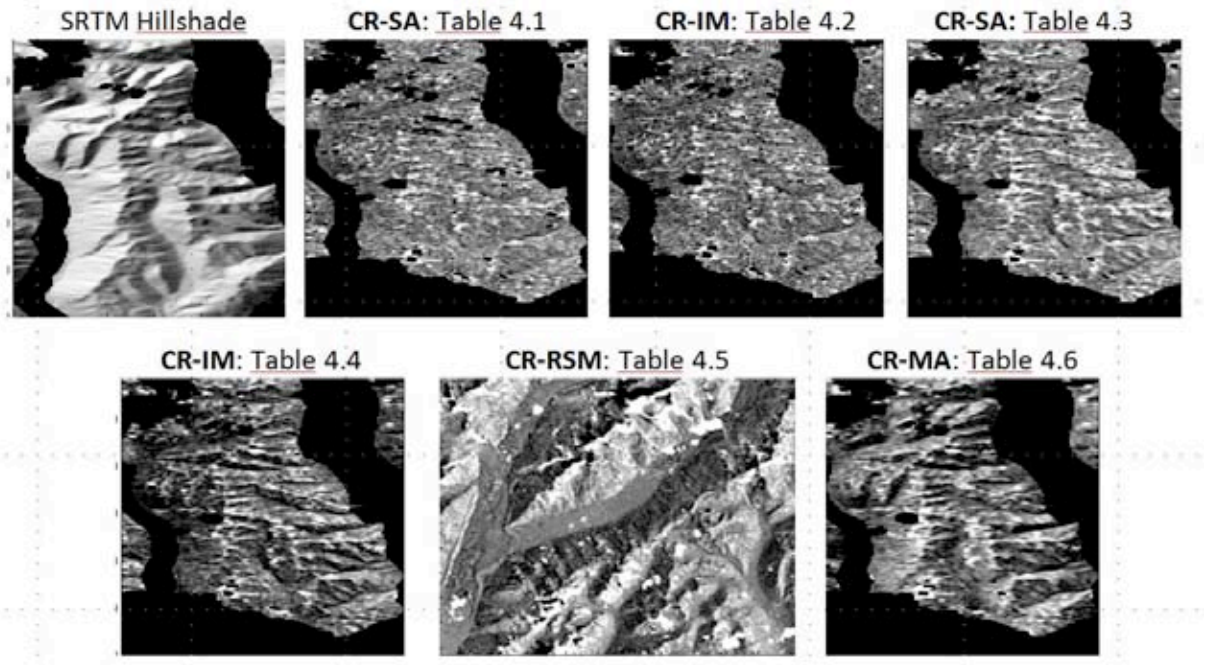

Fig. 4.4: The SRTM hillshade and elevation differences after co-registration (plotted as greyscale, "false-hillshades") from the six contributions to the round robin as detailed in Tables 4.1-4.6. Images associated with Tables 4.1 and 4.2 do not resemble topography and thus co-registration and matrix resampling and differentiation was successful. Slight topography can be seen from results associated with Tables 4.3-4.6. The image associated with Table 4.5 does not correspond with the area of interest since only a portion of the DEM differencing was returned by the participant. Note, Co-Registration seemed to be successful for the results of Table 4.5, however, the matrix resampling and differentiation in ArcGIS introduced a new horizontal translation.

4.4.2 Effect of the mean bias and standard deviation in relation to co-registration, resampling, elevation, slope and curvature

To further analyze elevation difference (residuals) between elevation models, the relationships between re-sampling procedures are investigated. DEMs of varying resolutions (pixel areas) depict different elevations at the same pixel centre location. In addition, as stated in Section 4.3.2, internal pixel to pixel horizontal location errors (i.e. from the instrument pointing accuracy) translate into vertical errors multiplied by the slope tangent. It has been suggested that an elevation bias may exist in the global SRTM elevation model (Berthier et al., 2006), or alternatively that an elevation bias may be caused by topographic derivatives such as slope and curvature (Paul, 2008; Gardelle et al., 2012). This is tested with the 90m SRTM DEM and the 30 m ASTER DEM of 2006 from the New Zealand round robin data package. Two elevation difference grids are generated after co-registration of the datasets (Fig. 4.1, 4.2 and Table 4.1). The first resamples (down sampling) the ASTER DEM to 90 m by applying a 3 by 3 pixel block mean filter to the original 30 m data, and then bilinear interpolation of the 90 m ASTER into the SRTM grid. The second resamples (up sampling) the 90 m SRTM to a 30 m resolution using bilinear interpolation into the ASTER grid. The elevation difference grids are then generated through matrix subtraction (SRTM minus ASTER). Figure 4.5 shows the relationship between the median and standard deviation of the elevation differences with elevation, slope and maximum curvature. It is apparent that up

sampling the 90 m SRTM DEM to 30 m results in a larger elevation bias with elevation, and further with a linear relationship to terrain curvature (similar to Gardelle et al., 2012). These relationships are not well defined or visible for the down sampled 90 m ASTER DEM. Of further interest is the standard deviation of the elevation differences is larger for the up sampled SRTM rather than the down sampled ASTER.

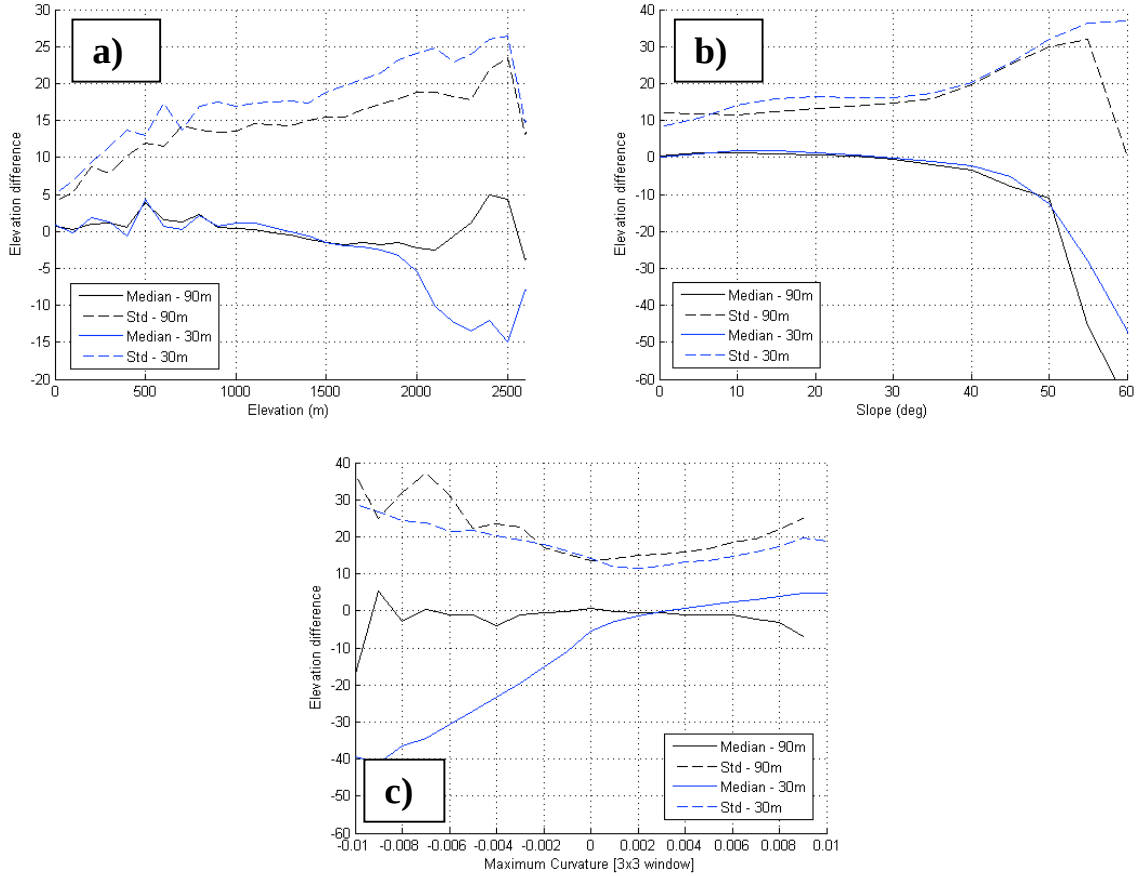

*Fig. 4.5. Median (solid line) and standard deviation (dotted line) of the elevation difference residuals on stable terrain (y-axis) between the SRTM and ASTER DEMs through up sampling the SRTM DEM to 30 m (blue lines) and down sampling the ASTER DEM to 90 m (black lines) plotted against elevation (**a**), slope (**b**), and maximum curvature (**c**).*

4.4.3 Accuracy of the ASTER DEM (AST14DMO product)

The ASTER GDEM (v2) is not suitable for deriving elevation changes from DEM differencing, as the year of DEM acquisition is spread over a 12-year period (2000-2011) without individual pixel time stamps. However, it is also possible to obtain from USGS a DEM from a single ASTER scene that is processed automatically, the AST14DMO product. Of course, this product has all the shortcomings the processing of a single ASTER scene has (e.g. Kääb, 2005) and the accuracy of this product has so far not been assessed. We have thus taken the opportunity of the round robin to compare the AST14DMO product to a DEM derived from LIDAR data acquired only 4 days later (on 4 Oct 2009) over Findelenglacier. Hillshades of both DEMs are shown in Fig. 4.6 and the elevation differences between the two DEMs are depicted in Fig. 4.7.

Before both DEMs were subtracted, the ASTER DEM was reprojected to the Swiss coordinate system, both DEMs were resampled to 10 m spatial resolution (bilinear) and co-registered using the approach described by Nuth and Kääb (2011) and in section 4.3.1. Glacier outlines from 2009 were also transformed to a 10 m raster to determine spatial statistics of the differences.

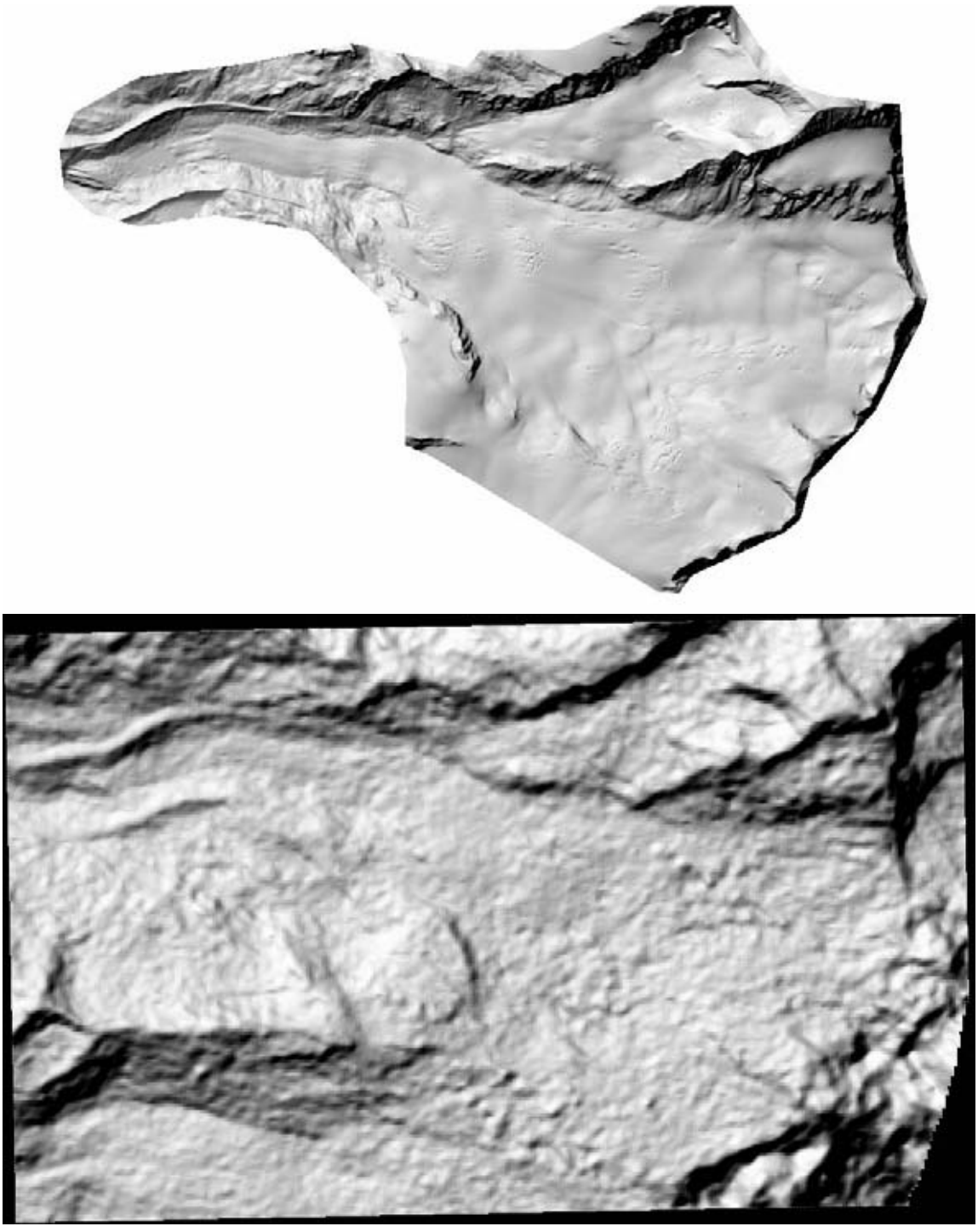

Fig. 4.6: Comparison of the hillshades of the LIDAR DEM (top) and the AST14DMO product (bottom) over Findelglacier. Both DEMs were provided by Philip Joerg, University of Zurich.

The hillshade of the ASTER DEM (Fig. 4.6, bottom) shows the typical bumpy structure known from previous studies but seems otherwise be free from artifacts. On the other hand, the LIDAR DEM (Fig. 4.6 top) shows all structures of the surface (crevasses, meltwater chanels, etc.) in great detail. Surprisingly, the differences over the glacier surface were in the mean over the entire glacier very small (0.47 m), but with a comparably large standard deviation of +/-8.74 m. Locally, differences are between 53 and -67 m. Considering also the region outside of glaciers, the differences vary between 167 and -321 m, largely driven by local artefacts. From this comparison we conclude that the AST14DMO product might be well suitable to derive glacier specific elevation changes, given that mapping conditions were excellent and a proper co-registration is performed.

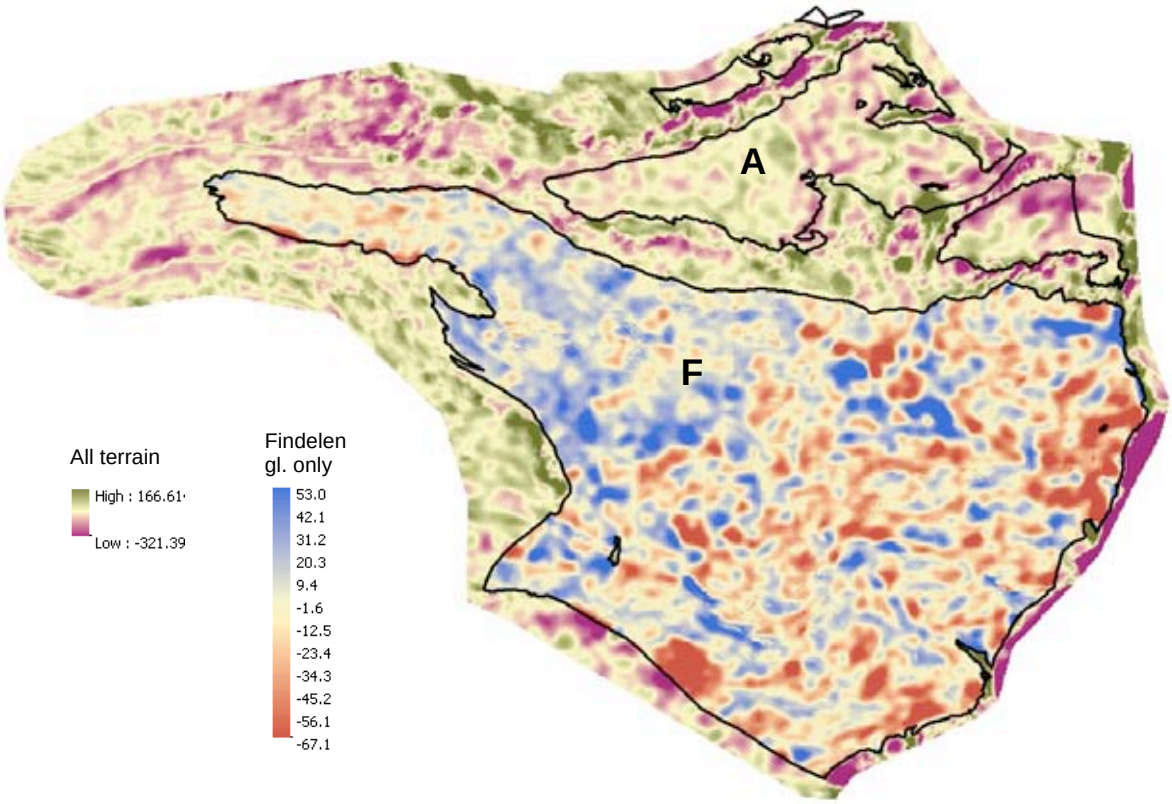

Fig. 4.7: Differences between a LIDAR DEM acquired on 4 Oct 2009 and the ASTER14DMO product derived from an ASTER scene acquired only 4 days earlier. Differences for all terrain are shown in green to purple, differences over Findelen glacier only are depicted in blue to red colours. Glacier outlines are depicted in black, F is Findelen and A is Adler glacier. The difference grid was provided by Philip Joerg, University of Zurich.

4.5 Discussion of the investigated algorithms

Three main algorithms (**CR-SA**, **CR-IC**, **CR-RSM**) for co-registering DEMs have been tested within the round robin experiment. All three algorithms operate on a single equation that minimizes elevation difference residuals over terrain assumed stable. The accuracy and precision of each of the three algorithms can be assessed through the vector sum of co-

registration vectors (triangulation provided in Tables 4.1-4.6). In most cases, these vector triangulations resulted in residuals almost as large as the DEM pixels themselves. The contributions where this occurred commonly used multiple software products. In those cases where the co-registration vectors are significantly different than Table 4.1 (i.e. Table 4.2, 4.3, 4.4 and 4.6), it is suspected that pixel definitions (pixel corner vs. pixel centre) and how the DEMs are imported into the various software products is most likely the cause. In the case of Table 4.2 (**CR-IC**), despite the variation of the co-registration vectors, the vector triangulation was successful, and therefore internal consistency was maintained. In summary, one contribution for each method resulted in acceptable triangulation residuals further providing accuracy estimates. All three methods resulted in horizontal accuracies greater than 2-3 metres and vertically as good as 2 m (i.e. better than 1/10th of a pixel).

A major difference between **CR-SA** and both **CR-IC**/**CR-RSM** is, that the elevation residuals are normalized by the slope tangent (cf. 4.3.2 and Glaciers_cci, 2012b) in **CR-SA** but not **CR-IC** and **CR-RSM**. Although this seems not to affect the horizontal co-registration significantly, it may affect the vertical co-registration because part of a single vertical point error is composed of horizontal mis-alignments (i.e. from individual pixel location accuracies) multiplied by the slope tangent. In practice, this will alter any mean/median elevation difference used as a bias correction (vertical co-registration in **CR-IC**/**CR-RSM**) depending upon the slope distribution of the samples used to determine the mean/median. This may explain some of the variation in the vertical co-registration adjustment, especially between the ASTER 2006 DEM and the SRTM DEM from most of the round robin contributions (Tables 4.1-4.6). Furthermore, the selection of stable ground from which to make these analyses ultimately leads to variability in the results. How much variability and bias will be induced by this depends upon the nature of the terrain and the data at the time of acquisition(s).

Two methods were implemented for co-registering DEMs to ICESat, **CR-SA** (Table 4.1) and **CR-RSM** (Table 4.5). For the **CR-SA** algorithm, triangulated residuals were all greater than 5 metres horizontally and 1 metre vertically. The **CR-RSM** algorithm resulted in triangulated residuals of less than 15 metres horizontally and 11 metres vertically. The potential reason for the better results from **CR-SA** is that the analytical solution to a mis-registration requires less samples to achieve a statistically robust estimate as compared to the **CR-RSM** algorithm which operates at similar quality/accuracy when sample sizes are large (i.e. over a million).

4.6 Algorithm selection

In this algorithm comparison, the focus of the round robin experiment is on selection and development of a universal co-registration routine for consistent application in any assessment. Various software products are used within the experiment (Matlab, IDL, GDAL, ENVI, PCI, ArcGIS) and their importing routines generated less ideal results that led to two important findings. First, the development of universal DEM co-registration tool for DEM differencing should be implemented in its entirety within one processing chain / one software. The round robin contributions that used multiple software products within their processing chain generally resulted in internal triangulation residuals (vector sum of co-registration parameters between three datasets) of similar magnitude as the co-registration parameters themselves. This effect is the result of importing definitions and routines within the various software products that apparently do not take into consideration the extended

information within the Geotif TAGs. While apparent in the results of the round robin, it was not detrimental to the experiment and algorithm selection as one result for each of the three available algorithms resulted in acceptable internal triangulation residuals. The second important finding is that despite the quality of the co-registration, if the final matrix differentiation is performed using ArcGIS software without controlling that the DEM grids are exactly coincident, then horizontal translations are incorporated in the final elevation difference grid due to the default 'nearest neighbour' interpolation technique used within the software's tools. A good example is that of the contribution of the **CR-RSM** algorithm in which co-registration parameters coincide with the consortium's implemented **CR-SA** algorithm, but the returned elevation difference grid resulted in a "false-hillshade" appearance and thus mis-registration. These two findings reinforce that development and implementation of a DEM co-registration tool should be within one processing chain, one software, and the output from the tool (either co-registration parameters, 2 co-registered DEMs or even an elevation difference grid) requires consideration that the users of the products use other software's that may or may not handle the pixel definition correctly. Subsidiary recommendations as a result from this experiment is that all software should be upgraded to handle properly the GeoTIFF Tags, and all individual users should be aware exactly how their software operates (i.e. in the case of ArcGIS).

In terms of accuracy, the three applied co-registration algorithms generally resulted in similar parameters and internal triangulation residuals of similar magnitudes. The most important consideration is the residual about the z-axis as this will directly transfer into the glacier elevation change products and estimates of mean glacier elevation changes from these products. The range of co-registration elevation adjustments (ΔZ in Tables 4.1-4.6) applied by the participants was ~5 metres for the co-registration between DEMs (i.e. 2006 ASTER and SRTM). This variability is significant for a glacier elevation change product, and we suspect that it results from the slope distribution of the stable ground sample used for the algorithm. These implementation variations require further investigations, especially to determine whether autonomous rules can be generated for stable terrain selection. Nonetheless this affect is less important for the **CR-SA** algorithm as it normalizes the elevation differences by slope tangent, and in fact, operates more significantly with steeper terrain than does the other algorithms.

It is important that the universal co-registration tool maintains consistent application in any assessment. For this goal, the round robin experiment included ICESat satellite laser altimetry, as it is the most consistent global elevation data product available to date (Nuth and Kääb, 2011). In terms of robustness, only the **CR-SA** and **CR-RSM** algorithms returned results for co-registering the DEMs to ICESat. The internal triangulated residuals of the **CR-SA** involving ICESat were of similar magnitudes as those using only DEMs (Table 4.1, i.e. less than a few metres horizontally and 1 metre vertically). The **CR-RSM** algorithm resulted in slightly lower accuracies (Table 4.5).

In terms of algorithm efficiency, the participants of the round robin were asked to also include computational times. These statistics provide a general overview, but are however biased in terms of variation on computers used and programs applied. For example, routine times for completion vary depending upon individual computer specifications and whether they are developed and operated with parallel processing capabilities. Nonetheless, the most efficient routines in our experiment were those operated by the consortium in Matlab running the **CR-**

SA algorithm (Table 4.1) which co-registered all the different products to each other individually less than 1 minute each. The **CR-IC** algorithm implemented in IDL from two participants resulted in processing times of 3-5 minutes (Tables 4.2 and 4.4) while the **CR-RSM** required sometimes up to 2 hours (Table 4.5). In terms of the number of iterations required to find a global minimum solution, **CR-SA** requires 1-4 iterations whereas **CR-IC** requires exponentially more depending upon the parameter step sizes. **CR-RSM** may be optimized to obtain solutions in 6 iterations (Gruen and Akca, 2005), though the processing times for the results returned in this round robin imply many more iterations, similar to that of **CR-IC** but requiring solutions to 7 parameters rather than 2.

In conclusion, we consider the application of the **CR-SA** algorithm to be the most robust and efficient algorithm with similar reliability to other published approaches and thus is the most appropriate for glaciers_cci. In addition, the analytical basis of the **CR-SA** algorithm allows consistent application for multiple data types and sources in which the other algorithms experienced slightly worse reliabilities. In terms of product accuracy, further investigation is required into the choice of stable terrain implemented and whether this choice may be accomplished automatically. This choice of terrain affects the individual accuracies of the co-registration, the estimated errors for elevation difference grid products and the mean elevation changes that may be derived from them. We estimate the reliability of the **CR-SA** algorithm to return co-registration parameters as accurate as 1/10th a DEM pixel but certainly not worse than 1/3rd a pixel. Finally, the resampling procedures used for the final glacier elevation change product requires more investigation, though first results imply that up-sampling to higher resolutions will lead to artefacts related to terrain curvature while down sampling to lower resolutions minimizes these effects.

From the product validation and algorithm comparison we draw the following **main conclusions**:

- If the ASTER source images have good contrast over glaciers, the automatically generated standard product (AST14DMO) has a sufficient quality to derive overall elevation changes (a mean value per glacier).
- The CR-SA algorithm is the most robust and efficient algorithm.
- The reliability of the CR-SA algorithm is estimated to return co-registration parameters as accurate as 1/10th of a pixel and certainly not worse than 1/3rd a pixel.
- For resampling DEMs, we recommend to resample to the largest pixel size through averaging procedures rather than downsample to higher resolutions using interpolation procedures.
- A processing system and co-registration tool development should be based within one single system / software as this reduces the potential for biases within processing.
- Software packages should upgrade to handle GeoTIFF Tags properly and completely.
- Individual users need to check and know how their softwares import and interpolate data as biases may be induced within these two basic steps.

5. Glacier velocity - optical

5.1 Context of the algorithm development and product validation

A large number of archived and upcoming optical and SAR satellite missions make it now possible to operationally map and monitor glacier flow on a nearly global scale. Deriving glacier displacements globally will provide unique glaciological information. It will make it possible to compare spatio-temporal variations of glacier velocities both within regions and between regions. Such knowledge will enable better understanding of a wide range of processes related to glacial mass fluxes, such as glacier response to climate and climatic changes, glacier physics and flow modes, glacier flow instabilities (e.g. surges), subglacial processes (e.g. erosion), supra- and intra-glacial mass transport, etc. Knowledge about glacier ice supply helps to understand the development of glacial lakes and associated hazards. Mapping and monitoring glacier flow globally perfectly complements current attempts for mapping and monitoring glacier areas and glacier volume changes on a global scale (GLIMS, GlobGlacier, Glaciers_cci).

Within the Glaciers_cci project, the generation of ice motion fields of glaciers is performed using repeat-pass images acquired by SAR and medium to high-resolution optical satellite data. In the case of images from a single-look direction none of these methods is capable of directly measuring the full 3-D ice velocity vector, which is of interest for ice dynamic studies. Additional assumptions on the ice flow can be introduced to provide ice velocity information directly applicable for glaciological tasks. In general surface-parallel ice flow is assumed, where the slope is derived from a low-pass filtered DEM. As a consequence, displacements from SAR and optical methods cannot be compared directly, but only if projected to a common geometry. This step involves assumptions and auxiliary data (in particular DEMs).

Product validation and algorithm selection concerning repeat optical data is discussed in this chapter. Offset tracking in repeat SAR data is covered in Chapter 6, though we sometimes refer to repeat SAR data also in this section (e.g. when introducing the round robin in section 5.3.2).

Surface displacements from repeat optical data are usually performed on orthorectified and projected images so that the displacements are directly in the coordinate system aimed at. The method applied in most cases is block-matching (in contrast to feature matching) where a maximum similarity of an image template is searched for in the second image. The techniques used for that purpose are similar, in parts identical, to the above amplitude offset tracking. A main difference between optical and SAR matching algorithms lies in their ability to cope with the large noise level in SAR images (radar speckle) (Debella-Gilo and Kääb, 2012). A second major difference relevant for offset tracking, is the different sensor and thus image geometry. Most optical data used for this purpose are nadir looking, with, for instance, similar pixel resolutions over the image and little effect of elevation changes and errors, whereas both effects play a significant role in the side-looking geometry of SAR data. In their raw form, optical orthoimages provide therefore the 2-dimensional horizontal displacement component in a given map coordinate system. Optical data are sensitive to surface features only.

It is very difficult to validate glacier displacements from space strictly against independent data with equal or better resolution, accuracy and precision, because:

(1) Glacier movement is temporally highly variable, at scales from hours to seasons or years. Glaciers often have diurnal to seasonal movement cycles and year-to-year variations, among others a consequence of constantly varying subglacial hydrology. A strict validation of glacier velocity products would therefore require perfectly simultaneous acquisitions of product generation and validation data, which could only be achieved by continuous ground measurements.
(2) Glacier displacement measurement from repeat images requires image windows (or features) to be compared. The obtained displacement is then representative not for a certain infinite point, but rather for an area. This representativeness is, though, not a strict analytical function of the real displacement field but a statistical relation to the displacement field, its gradients, image features and contrast, and the algorithm and the implementation used.

As a consequence, validation of the glacier velocity product should not only be based on external data, as these provide only a limited reliability. Glacier displacements from repeat optical and SAR imagery should thus also be validated internally (i.e. from the product itself). Algorithms can also be tested against synthetic images. Further, it should be noted that image matching (or: offset tracking for the entire procedure) strictly provides displacements. Glacier velocities (GV) are estimated by dividing the displacements through the time period between the acquisitions used. Thus, the GV product is the mean velocity over the observation period, and does not take into account velocity variations inbetween. This fact is of particular importance when analysing time series of glacier velocities.

This overall context led to the set-up of the round robin in the following way: In order to compare algorithms used for image-matching of repeat optical and SAR images and analyse various implementation schemes, we have chosen three test sites with different characteristics where we provided different datasets for the participants. The selected regions are:

- Karakoram (Himalaya)
  - Panchromatic Landsat image pair separated by ~1 year.
  - 1 Envisat ASAR image pair separated by ~1 year.
- Vest- Austfonna (Svalbard)
  - 1 ALOS PALSAR image pair separated by 46 days.
- Breiðamerkurjökull (Iceland)
  - 1 TerraSar-X image pair separated by 11 days.

For validation, results from three permanent GPS stations are available for Breiðamerkurjökull, 35 GPS stakes can be used for Vest-/Austfonna, while Baltoro glacier in the Karakoram is considered for both optical and SAR data. The comparison of all algorithms is based on the participants selection, while the statistical analysis of the results is based on the validation with GPS and on the SAR/optical intercomparison.

The round robin for glacier velocity was launched on 07/03/2012 and the last results were received by 08/05/2012. Instead of sending out a broad invitation to mailing lists (e.g.

Cryolist), we first identified and collected potential participants and approached candidates directly asking if they are willing to participate in the round robin. In the invitation to the round robin, participants were asked (Fig. 5.1) to perform a specific task; however, the final choice was left with the candidates according to their preferences and interests.

Fig. 5.1: Contacted candidates for the round robin on glacier velocity (green dot = no results; red dot = results delivered).

In total, 15 groups were invited to participate in the round robin, where seven provided results and three groups returning results for two tasks leading to a total number of ten datasets from external participants for the round robin inter-comparisons. Table 5.1 gives an overview of the participants and the results received for the different tasks. A more detailed description of the algorithms applied and results achieved for the optical sensor is provided in section 5.3.2.

Parti-cipant	Affiliation	Registered (DBT2 access)	Task 1	Task 2	Task 3	Task 4
			Baltoro Optical	Baltoro SAR	Vestfonna SAR	Iceland SAR
1	Swansea University	08.03.2012	x	x	-	-
2	University Strasbourg	08.03.2012	-	-	x	-
3	NORUT	08.03.2012	-	-	x	x
4	GFZ Potsdam	12.03.2012	-	x	-	x
5	University Erlangen	15.03.2012	-	x	-	-
6	Danish Technical University (DTU)	15.03.2012	-	-	-	x
7	Ohio State University (OSU)	19.03.2012	x	-	-	-

Table 5.1: Institutions participating round robin on glacier velocity, and corresponding products.

5.2 Methods for accuracy determination

5.2.1 External Validation

- Comparison against other image data: Glacier velocities from repeat image data can be compared against those from image data of equal or better resolution, accuracy and precision. The discrepancy between both velocity fields is then a function of (error budget):
  - the accuracy of both matches;
  - the co-registration between both image sets (i.e. same georeference), which can be tested by matching stable ground. Typically, discrepancies are related to absolute image orientation and orthoprojection;
  - the representativeness of the displacement obtained compared to the real displacement;
  - temporal, real velocity variations between the acquisition dates of the two image sets.
- Ground-based measurements: Satellite derived displacements can be compared to ground measurements such as those from GNSS, radar, lidar, tachymetric survey, etc. Though highly precise, the temporal and spatial representativeness of such data compared to the area and time covered by the image data to be validated will vary and is not strictly known.

5.2.2 Internal Validation

- Visual interpretation of the derived velocity field by a glaciologist: Check for unnatural outliers or other features in the field; check for coherence and consistency; check for unnatural patterns; check for (roughly) downslope direction. These checks are subjective, but will rely on basic physical laws such as the incompressibility of ice. Although subjective, this type of validation should be done in any case.
- Matching quality measures: Most matching algorithms provide directly, or after additional processing, quantities that describe the degree of similarity between the matching image windows, e.g. the correlation coefficient (CC) or signal-to-noise rations (SNR). These parameters are an indication for the reliability of an individual match. However, the measure is not strict, i.e. bad matches might actually accurately reflect the true displacement, or vice-versa. As a consequence, the measure cannot be used alone for validation.
- High and low pass versions of the velocity field: Due to the physical properties of glacier ice, such as incompressibility and stress transfer, and the low spatial variations of gravity that drives glacier flow, glacier velocities are usually smooth and coherent. This experience can be employed to compare different frequencies of the velocity field, and to disregard results that differ too much from a value expected from a field version at lower frequency. Practically, the original result can be compared to a low-pass filtered result and individual measurements be kept or disregarded based on the differences between both versions of the velocity field. Whereas, this validation or filter gives often good results, it fails where entire zones of the measurements are actually inaccurate, or where a glacier actually shows in reality high local velocity gradients
- Inversion of displacement: An image 2 can be inversely deformed using a displacement field between image 1 and 2, and the reconstructed image 1r compared to the actual image 1. The similarity between both can be quantified e.g. by using the cross-correlation (CC) coefficient. This method is less suitable to judge velocity products as the overall CC level depends on the content of the individual images, but the method

useful for judging the performance of different algorithms applied to the same set of images.

- Stable ground: Matching stable ground in the image set, if present, gives a good indication for the overall co-registration of the repeat images, and some general idea of the matching accuracy under the specific image conditions. The representativeness of the latter indication for the glacier velocities depends on the image content similarity between the stable ground and the glacier areas.

5.2.3 Protocol for algorithm assessment and product validation

Validation with ground-truth data (absolute)

- The first step is to interpolate displacements from image matching at location of ground point displacement data
- Secondly, apply Glacier_cci standard internal validation procedures (see below) in order to determine whether the image matching was successful for the extracted positions.
- Compute differences between displacements measured by image matching with those measured through means such as GNSS.
- Finally, compute the magnitude of difference vectors and their statistical summary.

Validation with image data (relative or absolute)

- Bring reference images (or already matched displacements if provided from externals) into same geometry as product images (into map geometry, or into acquisition geometry through polynomial co-registration). The validation is absolute if the map geometries have been derived independently, else relative.
- Match displacements in both image sets
- Apply Glacier_cci standard internal validation procedures (see below)
- Interpolate (if not matched at same spacing) displacement fields to same spacing
- Compute differences and magnitude of difference vectors and their statistical summary

Internal validation

- Match displacements in image set
- Investigate non-zero matches over stable ground; potentially improve image co-registration
- Apply a cross-correlation (CC) or signal-to-noise ration (SNR) threshold; threshold set interactively
- Compare raw displacements with low-pass filtered ones; apply interactive threshold on differences
- Visual inspection of remaining displacements; careful removal of obvious outliers
- statistical summary of stable ground matches

Algorithm comparison

- Match displacements in image set using the different algorithms
- Perform internal validation procedure (see above)
- Summary statistics (number of successful matches, etc.)
- If available, compare to validation data and compile discrepancy statistics (see above)

The criteria for selecting a “best“ algorithm are (often a combination of criteria is applied):

- Largest number of successful matches
- Largest distribution of successful matches (maximal fraction of glacier covered by successful matches)
- Smallest discrepancy to zero over stable terrain
- Smallest discrepancy to validation data or nominal displacements in case of synthetic images (measure: difference vector length)
- Minimal processing time and manual interaction necessary
- Degree of maturity: Established and published/peer-reviewed algorithm?
- Robustness against different image types and environmental conditions

5.3 Comparison of the applied algorithms

The algorithm comparison is based on two different exercises using Landsat data:

- The comparison of all algorithms listed in the ATBDv0 (Glaciers_cci, 2012b) based on a representative set of glacier sites around the Earth. This comparison is performed by the Glaciers_cci consortium (Heid and Kääb, 2012) and discussed in 5.3.1.
- A round robin exercise based on a Karakoram test site where participants used algorithms and implementations of their choice (described in section 5.3.2).

5.3.1 Algorithm comparison by the consortium (quantitative and qualitative)

The algorithms used in the consortium-internal comparison are described in the ATBDv0 (Glaciers_cci, 2012b).

Satellite data and regions

We choose a set of glacier regions and glaciers with different characteristics that we suggest to be globally representative in terms of glacier dimensions, topography, surface features, flow rates, climatic setting, etc. The European Alps (Landsat path 195 row 28) are chosen because the glaciers in this area are small valley glaciers with high surface transformation due to intense summer melt. Karakoram (path 148 row 35) is chosen because many of the glaciers in this area have a thick debris cover in their lower parts. Alaska (path 63 row 18) is selected because of the large mass flux and the high velocities. Pine Island in Antarctica (path 233 row 113) is selected because of the very little visual contrast in this area, and because of the very large glacier size. Southwest Greenland (path 9 row 11) is chosen because the velocity differences between fast ice streams and slower moving parts of the ice cap are very large, and because the glaciers here contain very little debris. Table 5.2 gives an overview of the images used, and Table 5.3 about the matching window sizes.

Area	Path/Row	Date image t = 1	Date image t = 2	ID image t = 1	ID image t = 2
European Alps	195/28	12 Aug 2000	30 Jul 2001	LE71950282000225EDC00	LE71950282001211EDC00
Karakoram	148/35	16 Jun 2000	21 Jul 2001	LE71480352000168SGS01	LE71480352001202SGS00
Alaska	63/18	31 Aug 2000	03 Sept 2001	LE70630182000244AGS00	LE70630182001246EDC00
Pine Island	233/113	13 Jan 2001	15 Dec 2001	LE72331132001013EDC00	LE72331132001349EDC00
Greenland	9/11	07 Jul 2001	08 Aug 2001	LE00901120011188EDC00	LE70090112001220EDC00

Table 5.2: Overview of Landsat image pairs used in the internal algorithm test.

Area	Window size, pixels	Filtering threshold, *m*
Karakoram	32	± 45
European Alps	16	± 45
Alaska	128 - 64	± 300
Pine Island	512 - 64	± 150
Greenland	64	± 100

Table 5.3: Size of matching windows and thresholds for the deviation between the raw matching results and the low-pass filtered matching results that are used to filter the displacement fields in the different test regions.

General results

Table 5.4 shows the percentage of assumed correct matches over glacierized areas for the different matching methods. The results vary both from method to method and from area to area. Generally, COSI-Corr obtains the highest percentage of assumed correct matches, whereas NCC and PC-O obtain the lowest. Alaska is the area which obtains the highest percentage of assumed correct matches, whereas Pine Island obtains the lowest percentage of assumed correct matches.

Area	NCC	CCF	PC	CCF-O	PC-O	COSI-Corr	n
Karakoram	46%	59%	56%	61%	54%	**67%**	21950
European Alps	49%	56%	51%	57%	48%	**67%**	5476
Alaska	*76%*	*86%*	*83%*	*87%*	*73%*	***91%***	13393
Pine Island	16%	**18%**	17%	17%	11%	13%	28822
Greenland	14%	25%	22%	23%	17%	**30%**	23790

Table 5.4: Percentage of assumed correct matches over glacierized areas for the different matching methods. Maxima by region are shown in bold, maxima by method are in italic.

Performance on stable ground

The RMSE of the matching measurements in both x and y directions over stable ground in the European Alps is given in Table 5.5. The results are comparable and better than 1/10 of a pixel for NCC, PC and PC-O. CCF and CCF-O perform slightly worse, but still close to 1/10 of a pixel. The matching method from COSI-Corr has the highest RMSE with about 1/4 pixel. Due to several small clouds in the Landsat scene from 2001 the number of erroneous matches is relatively high. NCC and CCF have more erroneous matches than the other matching methods.

Matching method	σ_x, m	σ_y, m	n
NCC	1.2	1.2	2917
CCF	1.5	1.7	3128
PC	1.0	1.2	3271
CCF-O	1.4	1.7	3288
PC-O	1.1	1.4	3252
COSI-Corr	2.9	3.9	3260

Table 5.5: RMSE of the matching measurements for the 6 different matching methods over stable ground in the European Alps.

Karakoram

The difference between CCF and CCF-O is striking. CCF-O obtains assumed correct measurements over most of the glacierized areas, whereas CCF has problems revealing that the fast valley glaciers have moved over the time period. The displacement field created by CCF is therefore very coherent, but wrong. However, in the upper and slower moving parts of the glaciers, the performance of the two methods is similar.

The performance of NCC is comparable to the performance of CCF-O on the lower parts of the glaciers where the visual contrast is good. However, NCC produces fewer correct matches compared to the CCF-O over all. This is mainly because NCC produces fewer correct matches above the snowline where the visual contrast is poor.

PC performs similarly to CCF, which means that it does not capture the velocity of the fast valley glaciers, but that the other glacierized parts are well matched. The number of correct matches in total is, however, lower compared to CCF. PC-O manages, like CCF-O, to capture the displacement both of the fast valley glaciers and of the glacierized area with poor visual contrast, but the number of correct matches is lower for PC-O compared to CCF-O.

COSI-Corr performs similarly to CCF-O on the glacier tongues, which have good visual contrast, but above the snowline COSI-Corr obtains a higher number of assumed correct matches than CCF-O. Therefore, the percentage of assumed correct matches for COSI-Corr is higher than the percentage of assumed correct matches for CCF-O.

European Alps

NCC is the only matching method to match most of the area below the snowline of the glaciers in this region. COSI-Corr creates fewer assumed correct matches below the snowline, but above the snowline, where the visual contrast is poor, COSI-Corr creates more assumed correct matches than NCC. When it comes to the number of assumed correct matches in total, COSI-Corr outperforms both NCC and the other four matching methods. This is because of its superior performance in the areas with low visual contrast. CCF-O obtains more correct matches than the three other similar Fourier methods.

The displacement field in the European Alps needs more manual editing than the displacement fields in Karakoram because the matching methods produce fewer correct

matches. Since we know that the effective window size for COSI-Corr is less than the actual window size, we also try to match the images from the European Alps using COSI-Corr and a window size of 32 pixels by 32 pixels. NCC still performs better using a window size of 16 pixels by 16 pixels than COSI-Corr does using window sizes of 32 pixels by 32 pixels.

Alaska

In this region, NCC experiences trouble not just with areas with poor visual contrast, but also with thin clouds and snow in crevasses. In the 2001 image there is a thin cloud covering parts of Malaspina Glacier. CCF-O and to some extent also COSI-Corr manage to match this area in spite of the thin cloud, whereas NCC does not succeed. At Bering Glacier some of the crevasses are filled up with snow in the 2000 image but not in the 2001 image. This also causes the NCC method to fail, whereas CCF-O and COSI-Corr succeed.

CCF-O performs better than the three similar Fourier methods also in this area. It both captures the displacement of the outlet glaciers better than the other methods and obtains more correct matches above the snowline.

Pine Island

Even if the glaciers_cci project does not cover ice sheets, we include Pine Island glacier in our evaluation in order to simulate an extreme case in terms of surface features. Since the displacement of the front of Pine Island Glacier is close to 3000 m over the time period used (approximately 1 year), large window sizes are needed in this area in order to capture the movement. The ice stream itself is therefore matched with a window size of 512 pixels, aligned depending on the measured displacements, and matched again using a window size of 64 pixels. The area outside of the ice stream is matched using a window size of 64 pixels since the movement here is much smaller.

All six of the methods manage to match Pine Island Glacier relatively well. This is an area with many crevasses that makes it easy to match. The parts surrounding Pine Island Glacier, however, lack visual contrast and are hence more difficult to match. This is the reason for a low percentage of correct matches in this area. The performance of the different matching methods is relatively similar.

CCF obtains the highest percentage of assumed correct matches in this area. However, many of these matches come from the slow flowing area surrounding Pine Island Glacier. On the ice stream itself, this method obtains fewer correct matches compared to the other methods. Because CCF measures smaller displacements over parts of the slow flowing area compared to other methods, we strongly suspect that many of the assumed correct matches are actually incorrect because the method does not normalize. It is however difficult to quantify this effect and filter the measurements correctly.

The different window sizes that are used in the matching over Pine Island Glacier make it possible to investigate the difference in the displacements derived. At the margins of the ice stream, the largest window size measures a larger displacement than the smaller window size. The difference between the methods is in extreme cases more than 400 m. This indicates that there is a strong velocity gradient in this area. In the middle of the ice stream the two window sizes perform much more similarly, indicating small velocity gradients in this area.

Greenland

Also in our Greenland case study, the percentage of assumed correct matches is relatively low because a large section of the images contain areas above the snowline and hence has poor visual contrast. COSI-Corr resolves the velocity differences between the fast flowing ice streams and the slower flowing parts of the ice sheet better than CCF-O. COSI-Corr also has the highest percentage of assumed correct matches in total.

CCF obtains a high percentage of assumed correct matches compared to the three other similar Fourier methods also here, but many of these matches stem from the slower flowing parts above the snowline. Many of these matches are therefore, as for the areas surrounding Pine Island Glacier, probably incorrect.

In some areas it is difficult to identify the ice streams in the Greenland images visually before the matching is conducted. There are large velocity variations over short distances, and velocity measurements are necessary to separate fast flowing areas from more stagnant areas.

Other tests: SLC-off images, 15 m vs. 30 m resolution, adaptive windows

To test how the three methods NCC, CCF-O and COSI-Corr perform on striped Landsat images from 31 May 2003 and onwards after a failure of the Scan Line Corrector (SLC-off), two images covering Larsen B, Antarctic Peninsula, are chosen. The images are from path 217 row 106, taken on 8 January 2005 and 11 January 2006. Matching results were obtained using windows of 64 pixels by 64 pixels for a 43 km by 43 km section covering both an area with stripes and an area without stripes. Both NCC and COSI-Corr fail in the striped area, whereas CCF-O ignores the stripes and obtains correct matches. The striped area contains strong visual contrast features like a crevassed glacier and several nunataks, so all methods are expected to perform well in this area if the images lacked stripes. In addition, COSI-Corr obtains very few correct matches outside the striped area, whereas NCC and CCF-O return approximately the same number of assumed correct matches. This area has some surface transformation because the 2005 image contains many small-scale snow dunes that are not present in the 2006 image. Also several melt ponds are present in the 2006 image but not in the 2005 image.

We also match the Karakoram images with CCF-O and COSI-Corr using Landsat 7 (SLC on) band 4 (near infrared, 0.77-0.9 μm) which has a spatial resolution of 30 m. Using a window size of 16 pixels, compared to 32 pixels for the panchromatic images (0.52-0.9 μm) with 15 m spatial resolution, it is possible to obtain matches in similar spectral wavelengths using windows covering the same ground sections. The matches are compared based on the difference in displacement between the assumed correct matches using 15 m and the displacement obtained for the same ground section using images with 30 m spatial resolution. Assuming an uncertainty using CCF-O of about 1/9 of a pixel, and statistically completely dependent matches, the two methods are assumed to measure the same displacement if the displacement difference is within 5 m. A total of 59 % of the matches are within this range. Extending this threshold to 10 m (2 sigma), 75 % of the matches are within this range. For COSI-Corr the accuracy is about 1/3 of a pixel, therefore the methods are assumed to measure the same displacement if the displacement difference is within 15 m. A total of 57 % of the matches are within this range, and 70 % are within the 2 sigma range.

For Baltoro glacier, Karakoram, (also the external round robin site) we obtain the displacements using normalized cross-correlation and adaptive matching window sizes based on the signal-to-noise ratio (SNR) of the digital numbers in the image and the cross-correlation coefficient. The SNR of the digital numbers is the variance of the signal divided by the noise variance. This means that the window sizes are varying over the image based on the level of noise in the images and the maximum correlation coefficients for each location. If the SNR is lower than a given threshold even at maximum window sizes, the method does not match. Hence, filtering is performed automatically before the matching is conducted. Comparing the results from the adaptive matching and COSI-Corr reveals that 89 % of the matches agree within 15 m.

For CCF-O and the adaptive method 90 % agree within 15 m. Mean size of the reference image for all the matchings is 29 pixels by 29 pixels. All of the matches that do not agree within 15 m are situated at the glacier margins or on the small tributaries to Baltoro. For almost all of these cases, the adaptive method measures higher velocities than the two other methods, due to less susceptibility to velocity gradients. In total, CCF-O and COSI-Corr give a higher number of assumed correct matches than the adaptive method does. CCF-O returns 1124 assumed correct matches, COSI-Corr 1175 assumed correct matches and the adaptive method 701 assumed correct matches in this area. The lower number for the adaptive method is mainly because it is, here, based on the NCC and returns hence fewer matches in the areas with low visual contrast (cf. above for NCC).

5.3.2 Algorithm comparison from round robin exercise

The round robin participants used the following two algorithms (see Table 5.6 for key parameters):

(1) Swansea University: Modified version of the GAMMA software. GAMMA uses in principle NCC for offset tracking, solved in the Fourier domain. Details of the modification are not published and not known to us (publication planned), but will among others include a software change to cope with optical images, which is not possible in the original GAMMA radar software.

(2) Ohio State University (OSU) uses in principle also NCC (Ahn and Howatt, 2011), but applied on various versions of the input images (different bands and principle components, if available; edge and gradient filters etc.) and decides on the final match based on comparison between these different results. Thus the method uses a standard matching algorithm but sophisticated pre- and postprocessing procedures.

Parameter	Swansea	Ohio State
Preprocessing (co-registration)	10 min	10 hours*
Matching window	30 x 30 pixel	31 x 31 pixel
Spacing	5 pixel	8 pixel
Oversampling	factor 2	-
Processing time	2 hours	36 hours
Postprocessing (SNR filter)	1 min	30 min

Table 5.6: Key parameters of the applied algorithms, according to the feedback forms of the participants (generation of pre-processed image versions).*

The results from the different algorithms are shown in Fig. 5.2 to 5.6 The Swansea method performs similar to the CCF-O method, with a relatively high number of correct matches in the glacierized regions. Most areas are matched as well as CCF-O matches, including areas with low visual contrast. The only areas that seem to be less well matched are very narrow glaciers, which are in some cases not found to move using the Swansea method, but which are found to move using CCF-O. The RMSE over stable bedrock seems to be comparable to values for CCF-O.

The OSU method is unlike CCF-O and the Swansea method in several ways. First of all it matches only points inside a mask indicating flowing glaciers. The effect is that some of the smaller glaciers are neglected and unmatched, and so are also stagnant glacier fronts and large parts of the accumulation area. Because no stable bedrock is matched, it is difficult to asesss the accuracy of this method, but points outside of digitized glacier outlines indicate that the RMSE is comparable to values for CCF-O and the Swansea method. The OSU method obtains more correct matches over Biafo glacier compared to CCF-O and the Swansea method. Also NCC manages to obtain correct matches in this area, and it is likely that it is high strain-rates that makes this area difficult to match using CCF-O and the Swansea method. Also, the surface changes on this glacier between the two images matched is striking. Presumably due to different melt conditions, significantly different supraglacial dust and debris features are visible in both images.

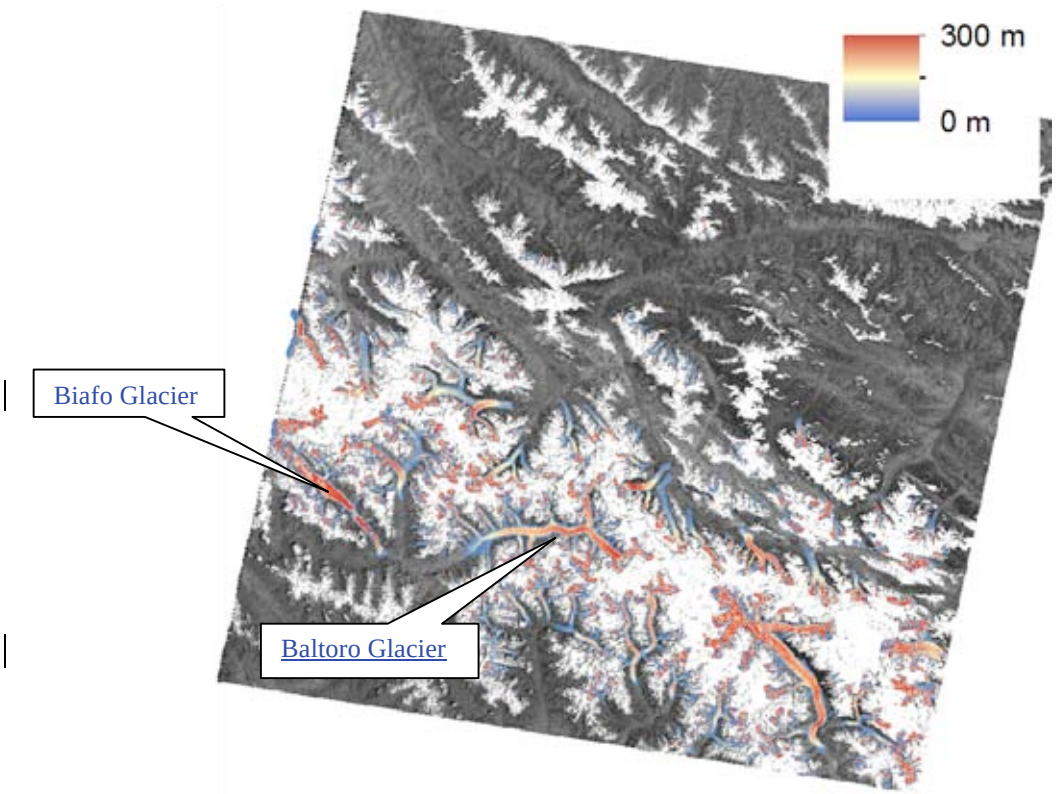

Fig. 5.2: Displacement magnitudes over Karakoram: OSU method, unfiltered. Biafo glacier is to the lower left.

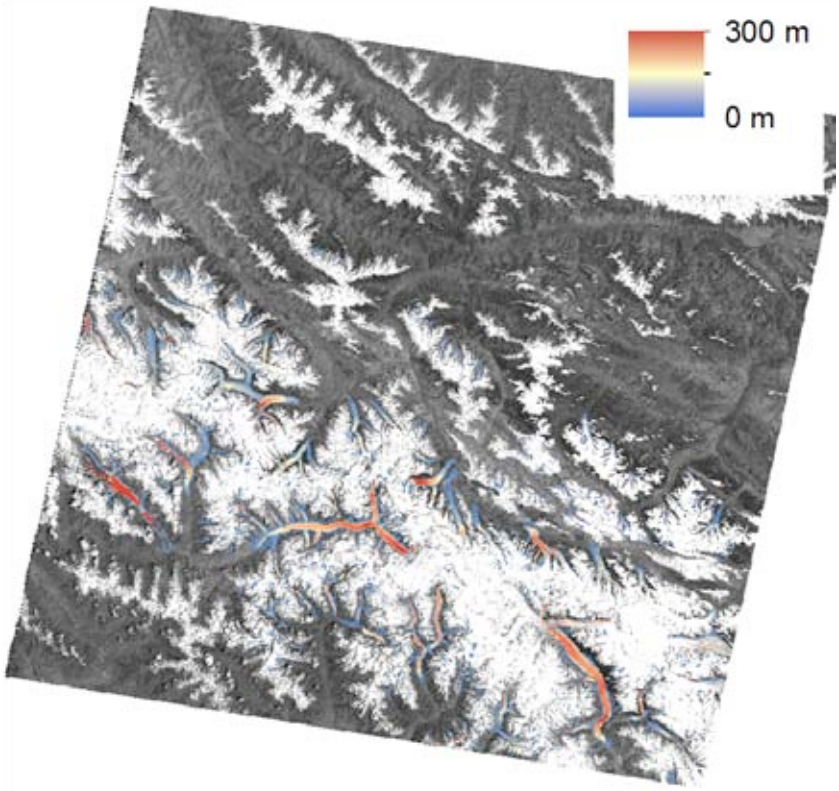

Fig. 5.3: As Fig. 5.2: OSU method, filtered. Biafo glacier is to the lower left.

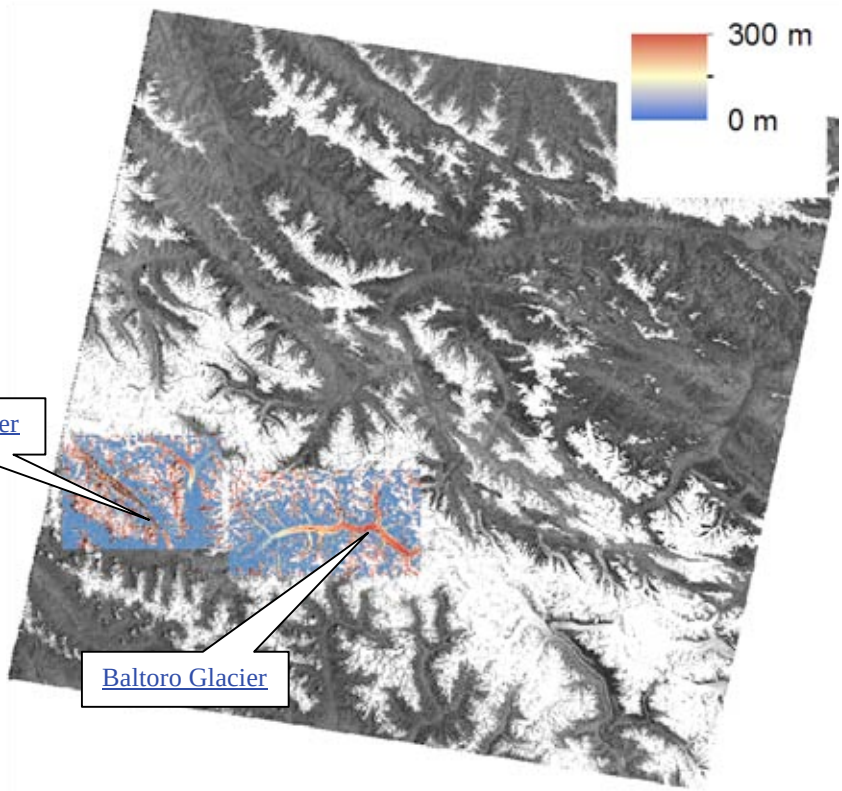

Fig. 5.4: As fig. 5.2 but with Swansea method. Biafo glacier is the left matched area, Baltoro glacier the right one. Note: Only two sections of matches are shown from the entire image matched due to the high density of points and according visualisation problems.

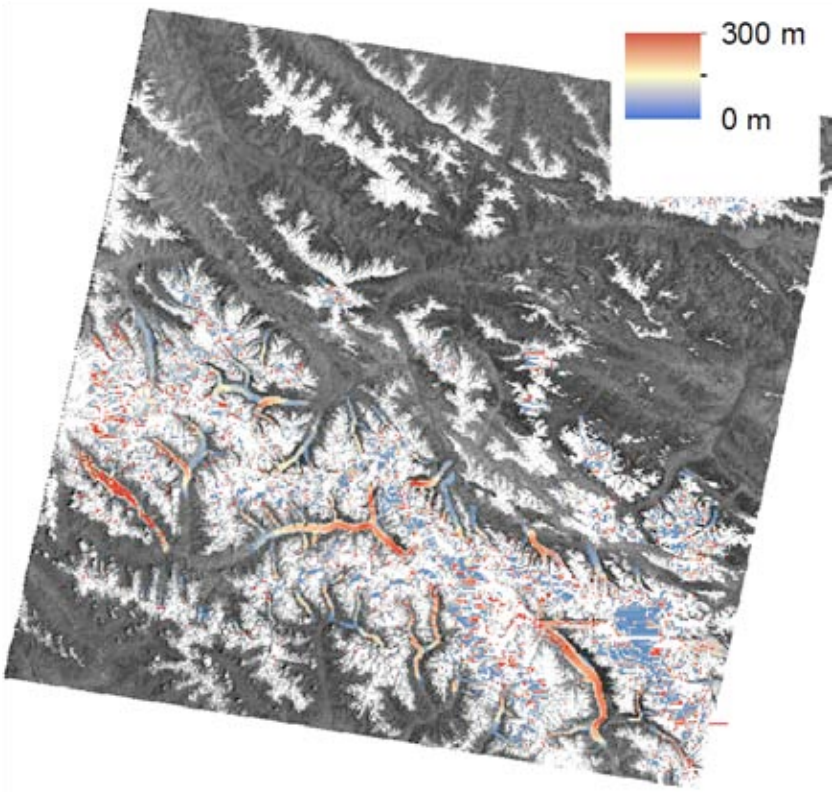

Fig. 5.5: As Fig. 5.2: NCC, unfiltered, University of Oslo implementation.

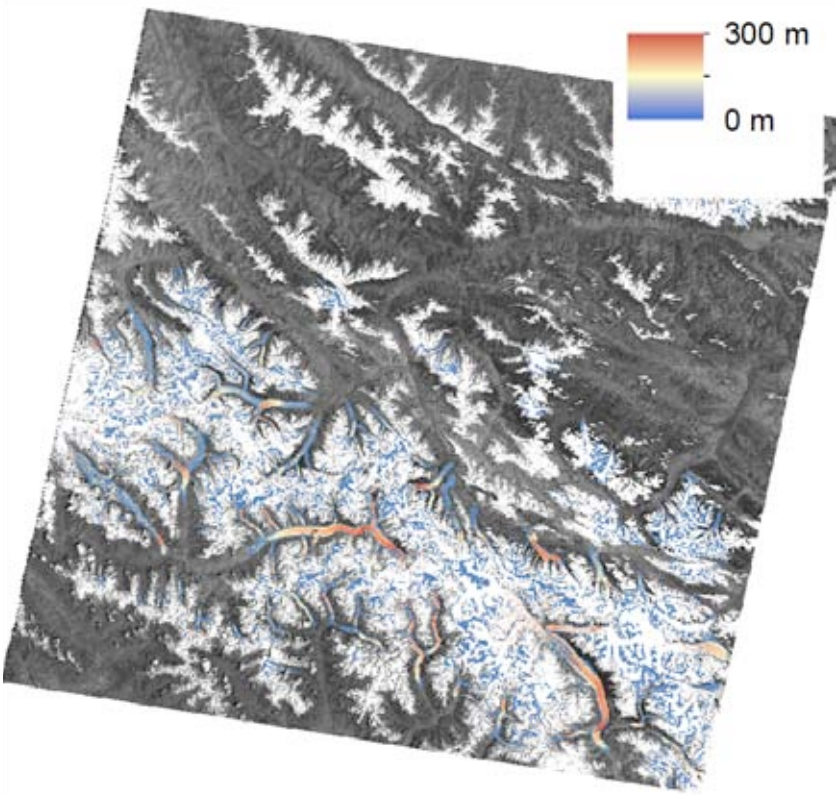

Fig. 5.6: Displacement magnitude over Karakoram: CCF-O, University of Oslo implementation.

Thirdly, Biafo glacier (Fig. 5.2) contains many longitudinal features (stripes) with few locally distinct features, leaving one matching dimension (transverse with respect to the glacier flow) to a large extent undefined (see image Fig. 5.7 over Biafo glacier compared to Fig. 5.8 over Baltoro glacier). NCC is more sensitive than CCF-O, and it is this increased sensitivity that makes it possible for NCC to match areas with high strain-rates. Further, the normalisation within NCC seems to cope better with the changes in surface reflectance between the two images matched. It is therefore possible that also the OSU method is more sensitive than CCF-O and the Swansea method, but for a representative conclusion this would have to be tested in, for example, areas with thin clouds or areas with snow filled crevasses in one of the images. Larger regions with such conditions were (un-)fortunately not present in the images matched in the round robin exercise. On glaciers other than the very difficult Biafo glacier, the results from the OSU method were similar to the CCF-O and the Swansea results.

In the post-processed version of the OSU method, many of the correct matches are removed. This is especially visible in the lower part of Biafo glacier. We therefore evaluated the raw results of this method instead of the post-processed results.

Because both NCC (Glaciers_cci/University of Oslo implementation) and the OSU method manage to match Biafo glacier, we assume that this is related to NCC and not the pre-processing step of the OSU algorithm. We therefore suggest that CCF-O in combination with NCC in areas with high strain rates is the most efficient method, because of the long processing time of the OSU method.

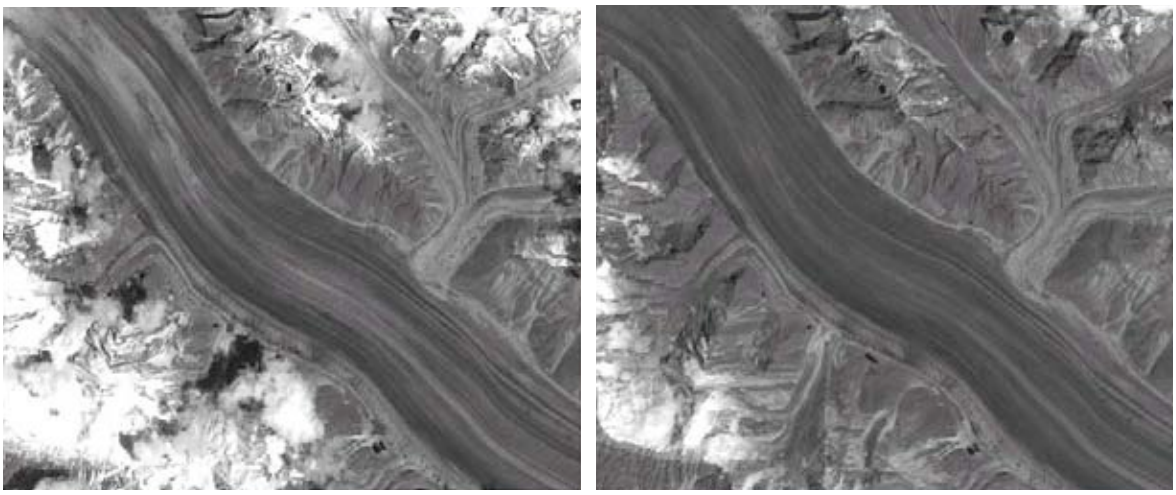

Fig. 5.7: Section of Biafo glacier, June 2000 (upper panel), July 2001 (lower panel).

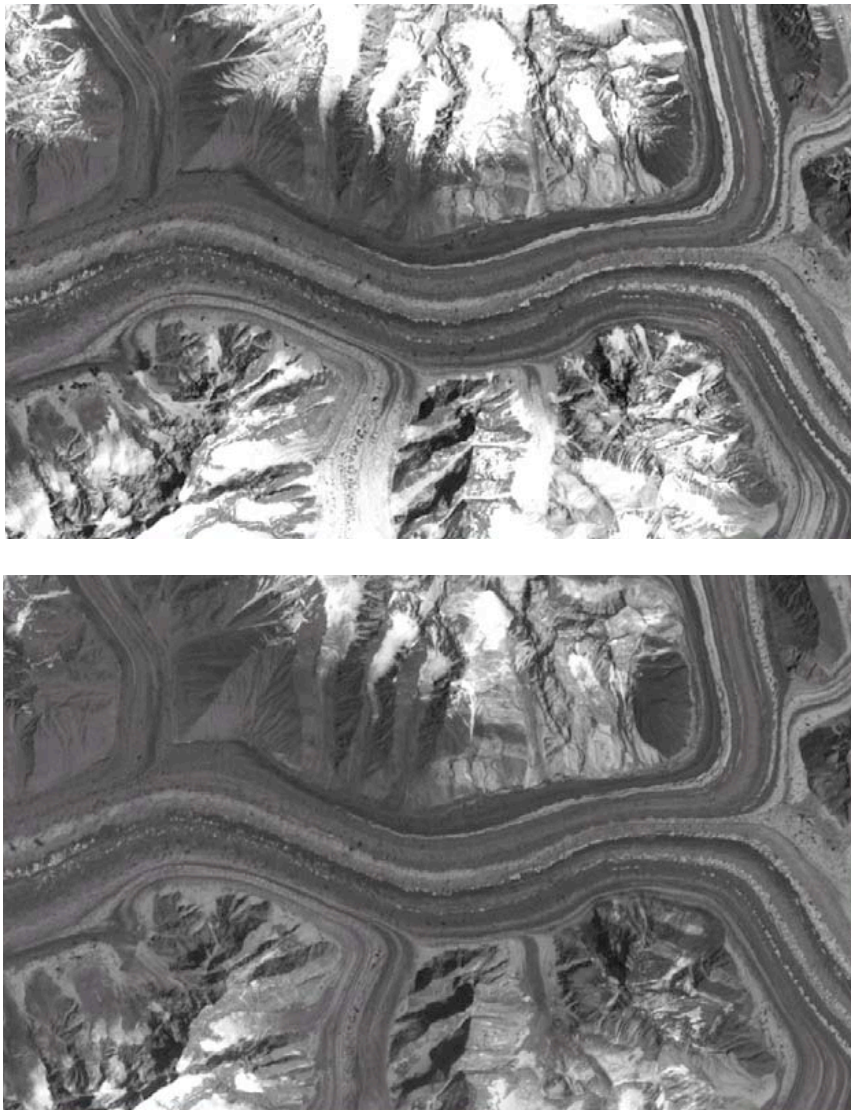

Fig. 5.8: Section of Baltoro glacier, June 2000 (upper panel), July 2001 (lower panel).

5.4 Discussion of the investigated algorithms

Matching methods with high RMSE values over stable ground are in general considered to be less accurate than matching methods with low RMSE values. The lower RMSE for PC and PC-O compared to CCF and CCF-O is because removing the amplitudes from the signal sharpens the correlation peak and thereby makes the sub-pixel determination more precise. The higher RMSE of COSI-Corr is due to the smaller efficient window size, and hence does not mean that the accuracy of COSI-Corr is lower, but that the implementations are different. However, since the accuracy is down to fractions of a pixel, also attitude variations and accompanying pixel geolocation errors of the Landsat sensor, and erroneous topographic corrections may influence the RMSE. It is clear that all methods obtain higher accuracies than what is possible to get using Landsat images due to the sensor noise. Therefore, the sub-pixel level accuracies in this study seem to be dominated by the sensor noise and not the accuracy of the matching methods.

Surprisingly, the un-normalized matching method CCF gives more correct matches than the normalized matching methods PC and PC-O. This is probably because it performs well above the snowline where the surface is very homogeneous, both spatially and temporally. Normalization is therefore not so important in this area. Below the snowline and outside of the glacier area it experiences problems, and it therefore proves that normalization is important in areas that are not homogeneous. This is especially clear in Karakoram where CCF does not register at all that the glacier has moved. Crevasses at one geographical place are more like the crevasses at the same geographical place the year after than the corresponding but displaced crevasses, probably due to different illumination conditions and surface degradation. Since CCF does not capture movement well, it is also difficult to use in slower flowing areas where it seems to return correct results, because there will always be transitions between fast and slow flowing areas. In zones where there is a transition from slow to faster flow it is difficult to know at which value to start filtering the matches. For the same reason there might also be erroneous matches that are not discovered in the present study. This could be part of the reason why the percentage of correct matches for this method is so high, especially for the Pine Island and Greenland studies.

CCF-O performs best compared to the three other Fourier methods (CCF, PC and PC-O) that are operated in the same way but with different normalizations. PC has problems capturing the displacement of fast outlet glaciers, so this kind of normalization seems to have little effect in such areas. In addition, it in general gives fewer correct matches compared to CCF-O. PC-O does not have any clear weaknesses compared to CCF-O, but in general gives fewer correct matches.

A previous study on the Larsen C ice shelf on the Antarctic Peninsula showed that few correct matches could be obtained in this area using the NCC method. The Larsen C ice shelf has a relatively homogeneous snow surface with few surface features like crevasses and flowlines. The present study confirms that the NCC method performs worse than other tested methods in areas with poor visual contrast. Hence, other methods should be chosen for matching such areas. COSI-Corr gives the most correct matches in low visual contrast areas, and this is mainly the reason why COSI-Corr obtains a higher percentage than other matching methods.

In areas of high visual contrast, the performance of NCC is comparable to the performance of CCF-O and COSI-Corr. In the European Alps, where the glaciers are small, NCC outperforms CCF-O and COSI-Corr in the narrow parts of the glaciers that are being channelized down the valleys. This is because NCC performs better with small window sizes, and in this case increasing the window size does not improve the results because the velocity varies significantly over short distances. For small glaciers where small window sizes are preferred, NCC can be a better choice than CCF-O and COSI-Corr. COSI-Corr performs better than CCF-O in such cases because it mainly uses the centre part of the window to match due to weighting of the central pixels, whereas CCF-O uses the entire window with equal weight. Large velocity gradients within the window therefore make it more difficult for CCF-O to match, whereas COSI-Corr and especially NCC are less sensitive in such cases.

Increasing the window size from 16 pixels by 16 pixels to 32 pixels by 32 pixels in the European Alps did not improve the COSI-Corr displacement measurements much. We tried this because of the smaller efficient window size of COSI-Corr. It is still difficult to get correct measurements on small glaciers with much deformation because COSI-Corr is less

sensitive to outliers, as explained in the Methods section. This method therefore needs the window to move coherently, which is not the case in areas with deformation.

Not only poor visual contrast, but also thin clouds and snow filled crevasses in one of the images can disturb the NCC method. CCF-O and to some extent also COSI-Corr are more robust under such conditions. This difference, like the difference between the methods for the small glaciers in the European Alps, arises because the matching methods work differently. Because CCF-O uses the entire window to match, it is less sensitive to noise. NCC, however, is easily dominated by large spatial gradients in digital numbers, like the difference between dark ice and bright snow for windows with snow filled crevasses, and searches for a similar jump in digital number in the image without snow filled crevasses. Since this difference in digital number is not present at the same feature in the image without snow filled crevasses it may find another feature with a similar difference in digital numbers and hence create a mismatch. Because COSI-Corr mainly uses the centre part of the window to match due to weighting of the central pixels, this method is more sensitive to noise constrained to few frequencies compared to CCF-O, but less sensitive to noise constrained to few frequencies compared to NCC. The example with snow filled crevasses and clouds is therefore connected to the example from the small glaciers in the European Alps, but the sensitivity of NCC therefore has an unwanted effect and the insensitivity of CCF-O has a wanted effect. This highlights the fact that one single matching method cannot be expected to perform well under all circumstances.

All the methods obtain a relatively similar number of assumed correct matches in the Pine Island area. This is probably because this area has too little visual contrast in most areas for the methods to obtain correct matches. But in the areas where the methods obtain assumed correct matches the visual contrast is very good due to crevasses, and all the methods manage to match these areas. As discussed above, normalization is not so important here because the area is very homogeneous.

The difference in the velocity derived using a large and a small window size over Pine Island Glacier highlights how important it is to use the smallest window size possible to avoid large velocity gradients within one window. The derived velocity may stem from anywhere inside the window depending on the digital numbers and the velocity distribution. It is therefore important to be aware of the fact that the displacement derived is not necessarily the displacement for the centre pixel of the window. This gets even more important when using the displacement measurements to derive strain rates. The smaller the windows are and the smaller the velocity gradients are, the more correct it will be to assume that the displacement derived is actually the displacement for the centre pixel of the window.

Of the three matching methods NCC, CCF-O and COSI-Corr, only CCF-O manages to match striped Landsat images successfully. Hence, only this method is considered to be useful on Landsat images from the ETM+ sensor after 31 May 2003 (SLC-off).

COSI-Corr obtains fewer assumed correct matches than CCF-O and NCC on the non-striped section of Larsen B. The 2005 image here contains many small-scale snow dunes that are not present in the 2006 image. These features are likely to corrupt the phase differences at many of the frequencies and hence inhibit successful matching. NCC is less sensitive to this kind of noise because other features represent a greater difference in digital numbers. The

performance of CCF-O was assumed to be lower because of these snow dunes since it normally outperforms NCC in areas with low visual contrast, but still it performs as well as NCC.

The difference between using images of 15 m spatial resolution and 30 m spatial resolution is generally within the 1 sigma uncertainty of the methods. According to statistical theory, 68 % of the differences are expected to be within this range, and since 59 % are within this range for CCF-O and 57 % for COSI-Corr the results are slightly worse than expected. There can be several reasons for this. Firstly, the different spatial resolutions may enhance or suppress different features so that actually different features are matched in 30 m spatial resolution images compared to 15 m spatial resolution images. If velocity gradients are present within the windows, this may also result in different displacements at different resolutions. Many of the glacier tongues show larger displacements using 15 m spatial resolution compared to 30 m spatial resolution. It is likely that small scaled features on glacier surfaces are better captured using finer spatial resolution so that especially for glacier margins, where the velocity gradients are large, images with finer resolution capture the displacement of the glacier whereas images with coarser resolution capture the more stagnant areas. Secondly, erroneous matches are present in both displacement fields. Most of the erroneous matches stem from the 30 m spatial resolution matching since this result is not filtered. But it is also possible that not all erroneous matches are filtered in the 15 m spatial resolution displacement field. However, since the general agreement is good, both CCF-O and COSI-Corr can be expected to provide accurate displacement measurements for images with 30 m spatial resolution. For areas with low visual contrast in medium resolution imagery the results can actually be better using 30 m spatial resolution, because the noise level is known to be higher for Landsat images with 15 m spatial resolution compared to Landsat images with 30 m spatial resolution. This is the case for some areas with low visual contrast in Karakoram. High resolution imagery, e.g. Ikonos, may be able to provide improved quality results (e.g. Bolch et. a. 2008) from the finer object detection of smaller pixel sizes, though is however not yet feasible on global scales.

The reason why the adaptive NCC method disagrees with COSI-Corr and CCF-O by more than 15 m for some points at the glacier margins is probably because the adaptive method uses smaller window sizes in this area compared to the other methods. The windows then include smaller velocity gradients so that the displacements measured are more representative for the centre pixel. Since the adaptive method implemented here is based on NCC, it is not surprising that it returns fewer matches than CCF-O and COSI-Corr in areas with low visual contrast. In the future, or using faster computers, the adaptive method can certainly be used for deriving glacier displacements globally. If SNR thresholds are tuned, locally adaptive window sizes can also be used in combination with other methods than the NCC. This will probably increase the number of correct matches in areas with low visual contrast, and at the same time areas containing large velocity gradients can be matched with smaller window sizes so that the velocities obtained are more representative for the centre pixel. However, it is difficult to use the adaptive method in combination with COSI-Corr since the latter method only accepts window sizes of 2^n where n is an integer. For all window sizes used in typical glaciological studies the difference from one window size to the next possible window size will probably be too large.

5.5 Algorithm selection

The methods have been evaluated according to criteria stated above, i.e. their applied suitability for global-scale mapping and monitoring of glacier flow using repeat medium-resolution optical satellite imagery, and their robustness. Landsat data of 15 m resolution (ETM+ pan) have been used for this test as they are probably the best suited data source for such global-scale application at present.

Of the four Fourier methods with different normalizations, CCF-O outperforms the other methods, both in areas with good and poor visual contrast. NCC is outperformed in areas of poor visual contrast, areas with thin clouds or changing snow conditions from one image to the next, but it performs well in high visual contrast areas and performs better than all other tested methods on narrow glaciers where small window sizes (about 16 pixels by 16 pixels) are needed. CCF-O has problems on narrow outlet glaciers. It also obtains fewer correct matches than COSI-Corr in areas with poor visual contrast, but more correct matches than NCC. COSI-Corr has problems on narrow outlet glaciers where small window sizes (about 16 pixels by 16 pixels) are needed, when snow dunes cover one of the images and also to some extent where thin clouds cover the surface. Of the three methods CCF-O, NCC and COSI-Corr, only CCF-O manages to match the striped Landsat images that are available after the failure of the SLC in May 2003.

The most robust matching methods of the six matching methods tested in this study are COSI-Corr and CCF-O. CCF-O produces fewer correct matches above the snowline, making COSI-Corr slight better there, although COSI-Corr has more problems with areas covered with thin clouds and snow dunes. NCC however, has clear advantages over small glaciers and for high deformation rates. Though this evaluation is not a complete overview of all matching methods and variations of them, it focuses on methods that are commonly used and that have performed well in published glaciological studies and the round robin exercise.

In summary, our study suggests that no one matching method clearly outperforms all others investigated under all circumstances, but rather that a set of two or three methods should be combined depending on the image conditions and the glacier characteristics. The evaluation criteria and algorithm comparisons suggest to base the further steps (round robin, algorithm implementation, etc.) on NCC and CCF-O. The COSI-Corr algorithm, performing similarly good as CCF-O is not publicly available. The NCC is given by (explanation see ATBDv0):

$$CC(i,j) = \frac{\sum_{k,l}(s(i+k,j+l)-\mu_s)(r(k,l)-\mu_r)}{\sqrt{\sum_{k,l}(s(i+k,j+l)-\mu_s)^2 \sum_{k,l}(r(k,l)-\mu_r)^2}}$$

In CCF-O orientation images

$$f_o(x,y) = sgn\left(\frac{\partial f(x,y)}{\partial x} + i\frac{\partial f(x,y)}{\partial y}\right)$$

$$g_o(x,y) = sgn\left(\frac{\partial g(x,y)}{\partial x} + i\frac{\partial g(x,y)}{\partial y}\right)$$

$$\text{where } sgn(x) = \begin{cases} 0 & \text{if } |x| = 0 \\ \frac{x}{|x|} & \text{otherwise} \end{cases}$$

are matched using cross-correlation (explanation see ATBDv0):

$$CC(i, j) = IFFT\,(F(u, v)G^{*}(u, v))$$

The OSU method shows that also the (automated) comparison of results using different preprocessing techniques can be useful, in particular in reducing erroneous measurements. As the CCF-O has an inherent bi-directional gradient filter, comparison of its results to NCC results resembles to some extent also the comparison of differently pre-processed images.

In summary, we consider the application of both NCC and CCF-O and a combination of both results as most appropriate for glaciers_cci, possibly extended by simultaneous matchings based on different pre-processing steps (cf. OSU method). The latter pre-processing steps may, however, to be refined depending on the sensor employed.

Further improvement can be achieved by filtering the matching results automatically by comparing the displacement matrix to its low-pass filtered version as proposed here. Matches over stable ground can be used to check and improve the co-registration between the images compared. Using these approaches, the matching process can be automated to a large degree for deriving glacier velocities with minimal user interaction, a prerequisite for being able the measure ice velocities over large areas and many regions. Though, we believe that a final expert check and edit of displacements should be performed and will still be necessary for some time to come in order to obtain glaciologically sound and useful results.

The algorithm evaluation showed that most algorithms and implementations are in principle able to achieve precisions far into the sub-pixel range. That makes clear that the actual error budget is in practice and in most cases not dominated by the algorithm itself.

The error budget of glacier displacement measurements from optical image matching or SAR intensity offset tracking consists therefore of:

- the algorithm precision;
- image co-registration (to be checked over stable terrain; accuracy close to matching precision possible);
- geometric sensor noise (sub-pixel level, but usually larger than algorithm precision);
- surface changes and transformations (i.e. representativety of surface features for ice particle displacement; e.g. influence of different illuminations, shift of surface features; sub-pixel or pixel level);
- mismatches due to similar but not corresponding features (e.g. self-similar ogives, crevasses or seracs; errors of many pixels possible);
- ability of post-processing procedures to eliminate measurement noise and mismatches;

In summary, the accuracy of individual glacier displacement measurements from repeat satellite data using image matching/offset tracking is on the order of one pixel, with areas with better accuracy, but also areas and points of much less accuracy. Reliability, or outliers, are frequent and require special attention.

From the product validation and algorithm comparison we draw the following **main conclusions**:

- No single matching method clearly outperforms all others investigated under all circumstances, but rather that a set of two or three methods should be combined depending on the image conditions and the glacier characteristics.
- The application of both NCC and CCF-O algorithms and possibly a combination of both results is considered as most appropriate for Glaciers_cci.
- The accuracy of individual glacier displacement measurements from repeat satellite data using image matching/offset tracking is on the order of one pixel. Outliers are common, requiring special attention.

6. Glacier velocity from microwave sensors

6.1 Context of the algorithms and accuracy determination

The general context of glacier velocity algorithms and possibilities to determine their accuracy are described in sections 5.1 and 5.2 and are thus not repeated here, besides the following details specific to SAR.

SAR data have the advantage of an active sensor that is not affected by solar illumination (day/night) or cloud coverage. Moreover, the acquisition geometry, satellite orbit and the technical sensor properties are well defined and stable, enabling precise analysis of repeat-pass images which are required for retrieval of glacier motion. The penetration depth of microwaves in dry snow and ice is on the order of a few metres to more than 100 m, depending on the radar frequency and snow and ice purity and structure (Ulaby et al., 1982; Maetzler, 1996; Rignot et al., 2001). While radar methods detect a mixture of surface and sub-surface features, they are generally referred to as measurement of glacier surface flow.

Depending on the available SAR data in terms of spatial resolution and temporal sequence of repeat passes, two primary methods evolved during recent years to retrieve the surface ice motion field:

- Across-track Interferometric repeat-pass SAR (InSAR) analysis, delivering the velocity component in radar line-of sight (LOS).
- SAR image offset tracking: this encompasses various methods, including the cross-correlation of chips in amplitude or complex SAR images and coherence optimization.

These methods deliver velocity components in LOS and along-track and do not require phase-unwrapping, but are less sensitive to displacement than InSAR. Depending on the employed algorithm they require distinct features like crevasses or coherence. In their raw form, SAR offset-tracking provides 2-dimensional displacements in SAR geometry, i.e. in the azimuth and slant-range system, and radar interferometry provides 1-dimensional displacements in the line-of-sight (LOS) direction between the ground points and the satellite positions.

Typical temporal baselines suitable for optical data are up to 1-2 years, for SAR offset-tracking up to a few weeks (depending on magnitude of ice flow and stability of available features), for SAR interferometry a few days (see PSD, Glaciers_cci, 2011b). In the optical domain, tracking methods are usually called ‘image matching’, in the microwave domain ‘offset tracking’. Here we use the term offset-tracking for both optical or SAR data. As explained in the ATBDv0 (Glaciers_cci, 2012b), the focus of Glaciers_cci and thus of this chapter is on SAR offset tracking.

6.2 Overview of the round robin experiments

An overview of the round robin set-up for velocity, both optical and radar, is given in section 5.3.2

Product validation and round robin exercise for glacier velocity using radar data are based on three experiments defined to consider different EO sensors, validation data, geographical areas and glacier types:

- Breidamerkurjökull, Iceland, with three permanent GPS stations and based on TerraSAR-X data;
- Nordaustlandet, Svalbard , with 35 GPS stakes and based on ALOS PALSAR data;
- Baltoro, Karakoram, same site as the optical round robin, with Envisat ASAR data.

The comparison of all algorithms is based on the round robin participants selection, while the statistical analysis of the results is based on the validation with GPS and optical data

6.3 Intercomparison of displacement fields from SAR data

6.3.1 Karakoram

Two repeat pass ENVISAT ASAR images acquired on 04/04/2004 and 24/04/2005 with a temporal separation of about one year were provided as raw data or in Slant range Single look Complex (SSC) images in the Gamma format. Four groups participated in the round robin over Karakoram using ENVISAT ASAR data. Details are listed in Table 6.1.

Parti-cipant	Organi-sation	Pre-pro-cessing	Software	Matching Window Size	Matching Spacing	Oversamp-ling Factor	Time	Post-pro-cessing
1	Swansea University	10 min	GAMMA (modified)	40 x 40	4 pixel	2 x 2	1 h	1 min
2	University of Erlangen	-	GAMMA	64 x 192	5 x 25	2 x 2	-	30 min
3	GAMMA	-	GAMMA	64 x 256	12 x 48	2 x 2	-	-
4	GFZ Potsdam	< 1 hour	ENVI SARscape	32 x 64	5 x 25	16 x 16	1.5 d	30 min

Table 6.1: Participants of the round robin for the Karakoram test region.

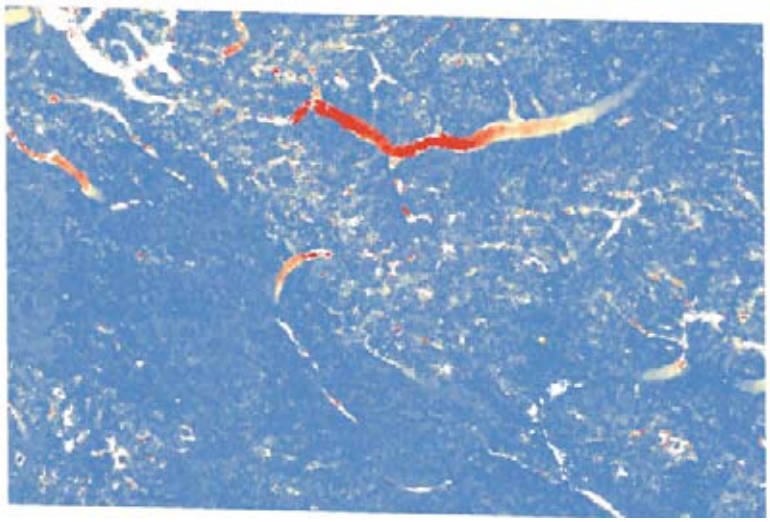

Fig. 6.1: Displacement magnitudes over Baltoro glacier, Karakoram, from participant 1. SAR geometry.

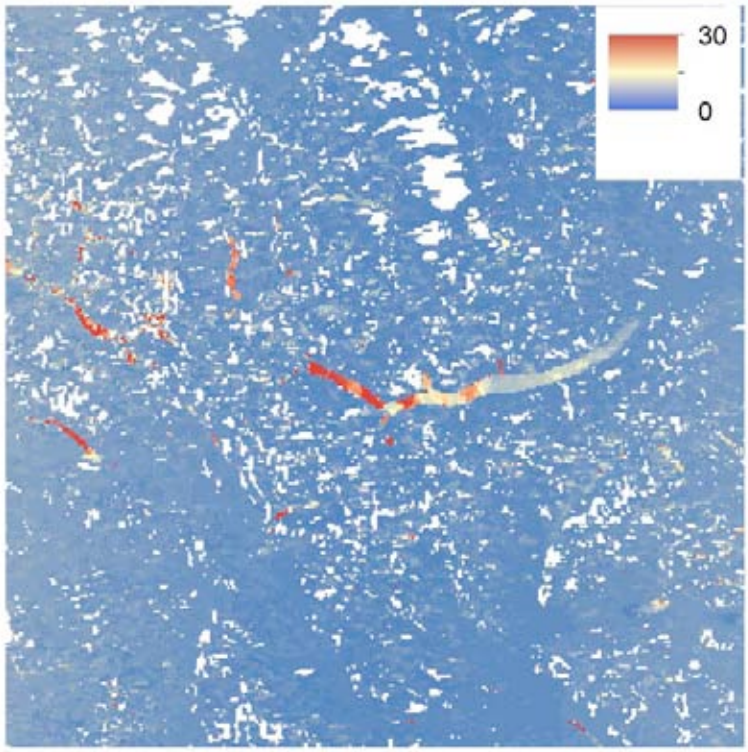

Fig. 6.2: Displacement magnitudes over Baltoro glacier, Karakoram, from participant 2. SAR geometry. Results from participant 3 are virtually the same.

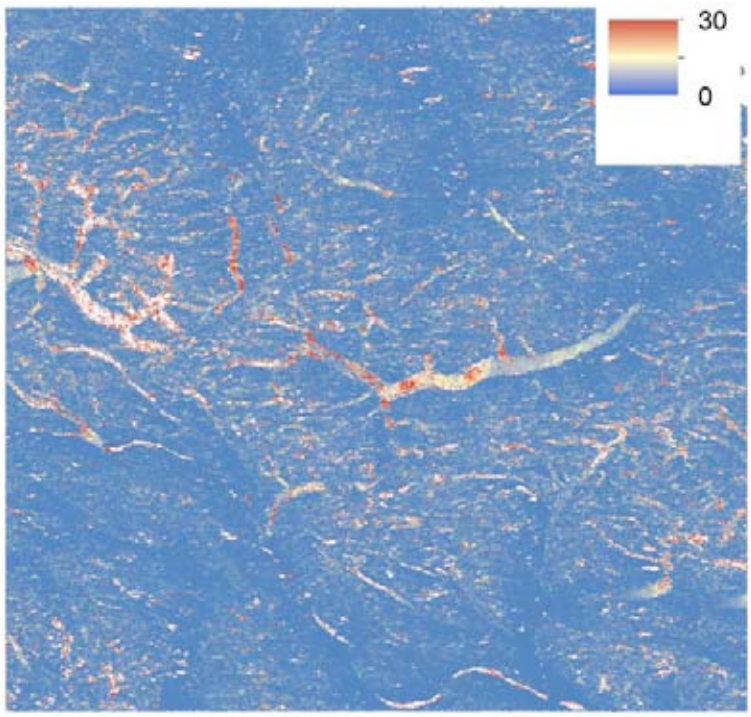

Fig. 6.3: Displacement magnitudes over Baltoro glacier, Karakoram, from participant 4. SAR geometry.

The results for Karakorum indicate that SARscape 4.4 obtains fewer correct matches in the glacierized areas than GAMMA and the GAMMA-modified version do. This is especially visible in upper parts of Baltoro glacier. GAMMA and the GAMMA-modified version perform relatively similarly. The accuracy over stable ground is somewhat lower for the results by participant 2 due to a residual error in the image co-registration. Else, the difference of the two GAMMA results might be governed by the very different window sizes used. This effect stresses the importance of the window size selection, which is so far done by the operator, and thus depending on his/her experience with certain glacier types, and on the time invested in testing different sizes for a test-section of the data before processing the full scene.

Strikingly, the results from participants 2, 3 and 4 show a similar pattern of high and low displacements on Baltoro glacier, whereas the Swansea results shows a continuous smooth displacement field, with decreasing speeds towards the terminus. The Swansea results fit very well to the results obtained by optical matching though the spatial distribution of successful matches is limited in the smaller sections of the glaciers (cf. section 6.4.1). We assume thus that the latter results are more correct. As the results from participants 2, 3 and 4 stem from different softwares, but produce a similar displacement pattern, we further assume that the significant differences of these two solutions to the Swansea method are not necessary due to the algorithms used but perhaps rather due to the different window sizes applied. Larger window sizes (participants 2/3 and 4) tend to average out speed peaks, or tend to include stable ground at the glacier margins, which again reduces the average speed in the window. It should also be noted the cross-correlation tends to prioritize the highest contrasts in the

matching window. If such dominant contrast features within a window are outside the glacier, they tend to dominate the overall displacement, if included in the (large) window, even more than expected from equal weights of the pixels within a window. However, until details of the GAMMA software modification by the participant from Swansea are not published, it will not be possible to definitively attribute the differences on Baltoro glacier to either window sizes or the matching algorithm.

6.3.2 Nordaustlandet, Svalbard

Two repeat pass ALOS PALSAR images acquired on 01/02/2008 and 18/03/2008 with a temporal separation of 46 days were provided as raw data or in Zero Doppler slant-range Single Look Complex (SSC) images in the Gamma format. Four groups participated in the round robin over Nordaustlandet using ALOS PALSAR data, as detailed on Table 6.2. All groups used the intensity cross-correlation as main algorithm. Fringe-visibility, InSAR and Multiple-Aperture InSAR (MAI) were used by two participants in addition to the intensity cross-correlation. It is worth mentioning that the three software products employed for intensity cross-correlation used three different conventions for positive and negative displacements along the slant-range and azimuth directions. This convention has been harmonized for the presentation of the results.

Images of the slant-range and azimuth displacements in pixels are shown in the following figures with the same colour scale, i.e. ±5 pixels in slant-range and ±10 pixels in azimuth for the results of the fringe-visibility and intensity cross-correlation algorithms. The size of the images is different for the various groups, because different matching spacings were employed. Intensity (amplitude) cross-correlation performs better than fringe-visibility with a larger coverage of valid values, in particular along the outlet glaciers. Over the slower moving interior of the ice cap the performance of the two algorithms is similar.

Parti-cipant	Organisation	Algorithm	Software	Matching Window Size	Matching Spacing	Oversamp-ling Factor	Quality Measure
1	UNISTRA	Intensity cross-correlation	GAMMA	128 x 256	5 x 50	2 x 4	SNR
2	ENVEO	Intensity cross-correlation	ENVEO	64 x 64	10 x 20	16 x 16	Correlation Coef
3	GAMMA	Intensity cross-correlation	GAMMA	64 x 192	6 x 36	2 x 2	SNR
4	NORUT	Intensity cross-correlation	NORUT	12 x 54	6 x 29	4 x 4	Contrast measure
5	GAMMA	Fringe-visibility	GAMMA	32 x 32*	6 x 36	2 x 2	SNR
6	UNISTRA	InSAR	GAMMA	N.A.	2 x 8	N.A.	Coherence
7	UNISTRA	MAI	GAMMA	N.A.	6 x 36	N.A.	Coherence

Table 6.2: Round robin participants for Nordaustlandet (search chip interferogram size 32).*

The results of three of the four groups regarding intensity cross-correlation are very similar in terms of coverage with valid information and displacement values. Using a smaller window for running the algorithm of the 4th group is presumably the reason for the higher variability of the retrieved displacement. Nevertheless, the magnitude of the measured displacements is also in this case similar to the other groups.

Slant-range and multiple-aperture interferograms were not unwrapped in order to compute line-of-sight displacement values. Over the outlet glaciers the interferograms are largely decorrelated and phase unwrapping will fail. Over the interior of the ice cap, where coherence is higher, the reason not to unwrap the MAI interferogram is that the performance of this algorithm is in the best case similar to that of intensity cross-correlation but with a smaller areal coverage. In the case of InSAR, the interferogram was not further analysed even over the interior of the ice cap because a Digital Elevation Model of sufficient quality to efficiently remove the topographic component of the interferometric phase is not available.

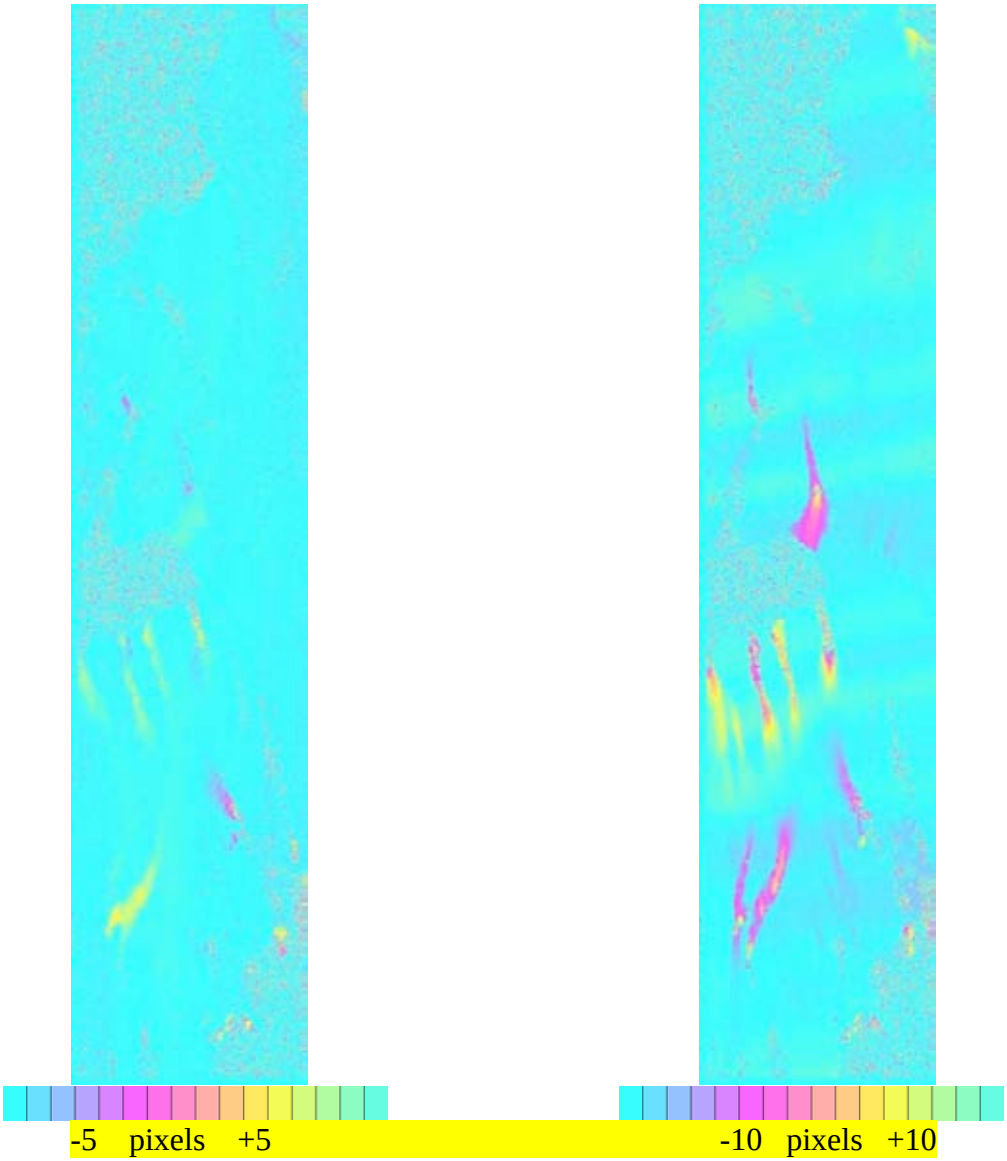

Fig. 6.4: Slant-range and azimuth displacement maps from intensity cross-correlation from ENVEO. Image size is 495 x 2248 pixel.

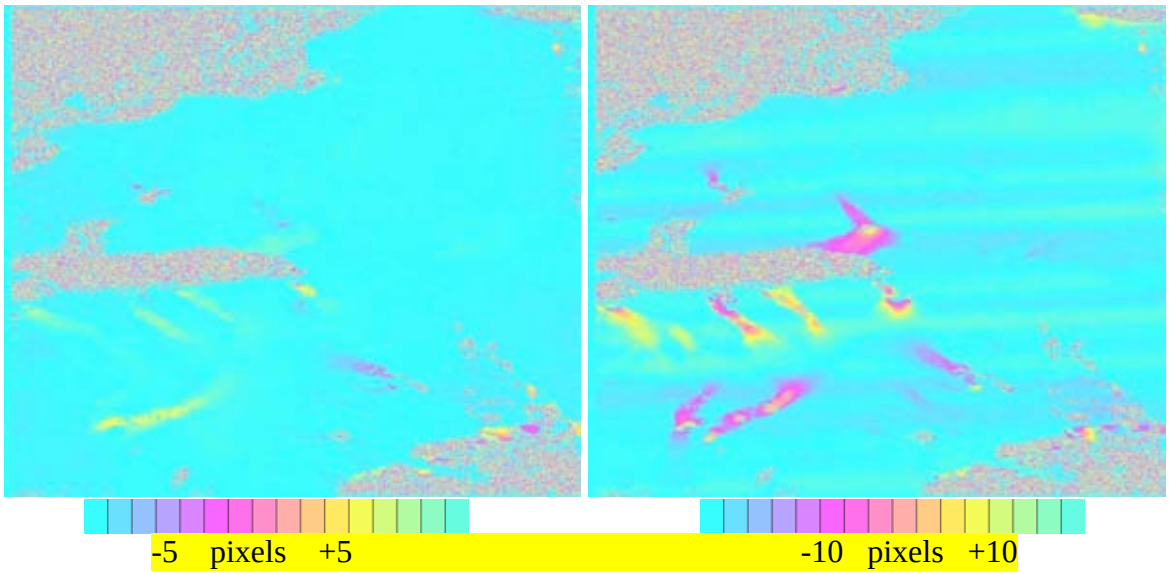

Fig. 6.5: Slant-range and azimuth displacement maps from intensity cross-correlation from UNISTRA. Image size is 893 x 1028 pixel.

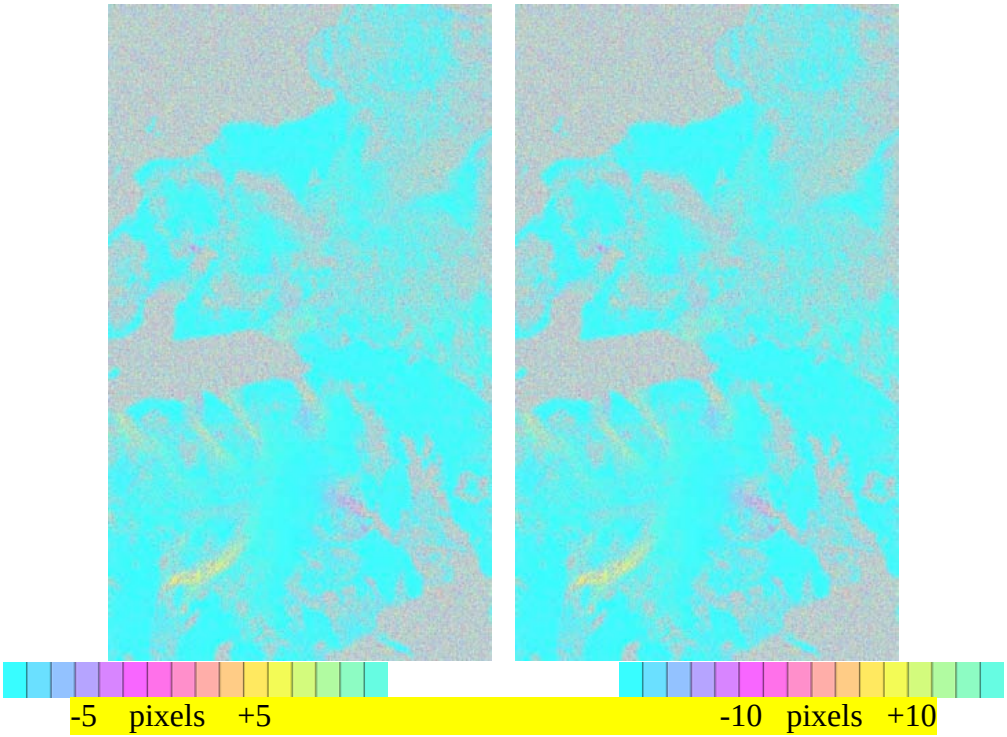

Fig. 6.6: Slant-range and azimuth displacement maps from intensity cross-correlation from NORUT. Image size is 849 x 1458 pixel.

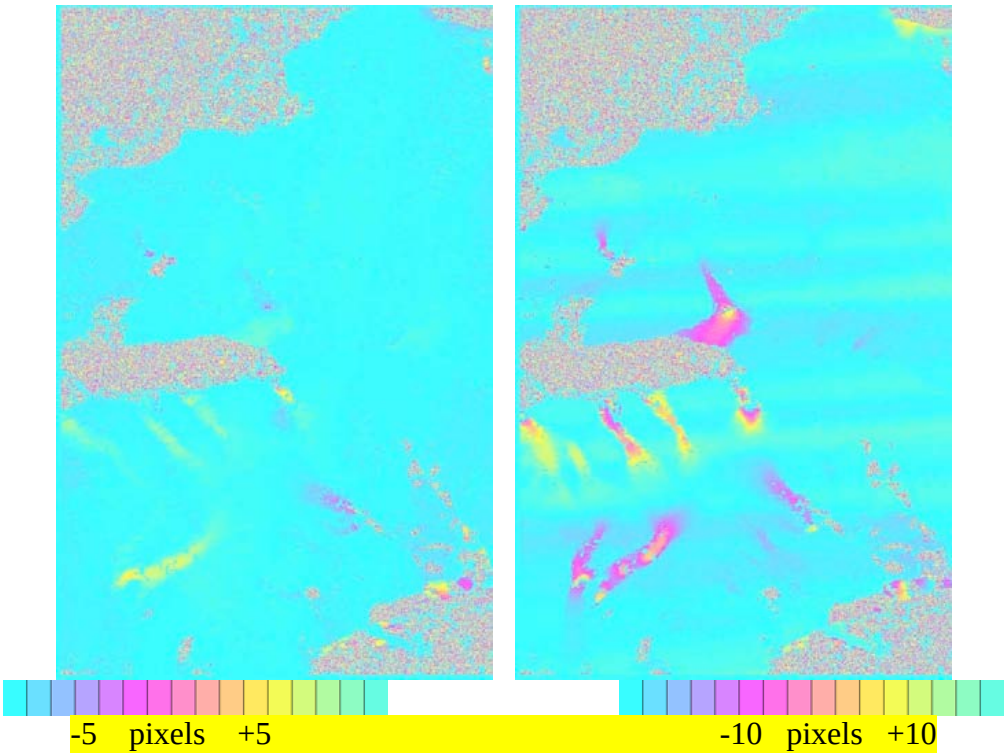

Fig. 6.7: Slant-range and azimuth displacement maps from intensity cross-correlation from GAMMA. Image size is 810 x 1247 pixel.

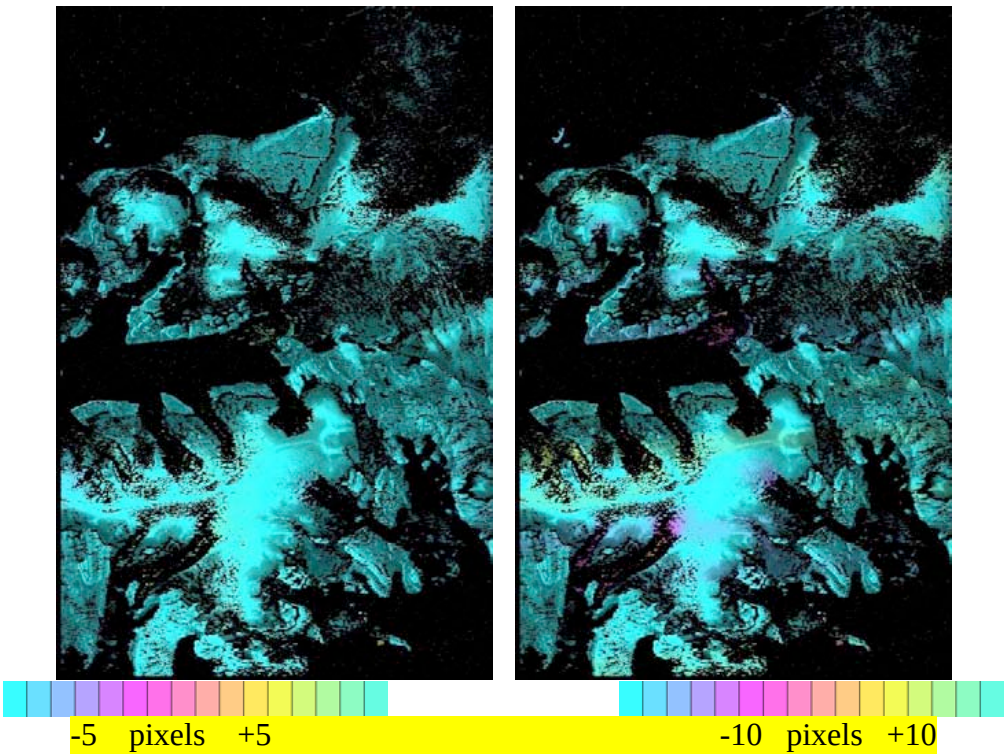

Fig. 6.8: Slant-range and azimuth displacement maps from fringe-visibility from GAMMA. Image size is 810 x 1247 pixel.

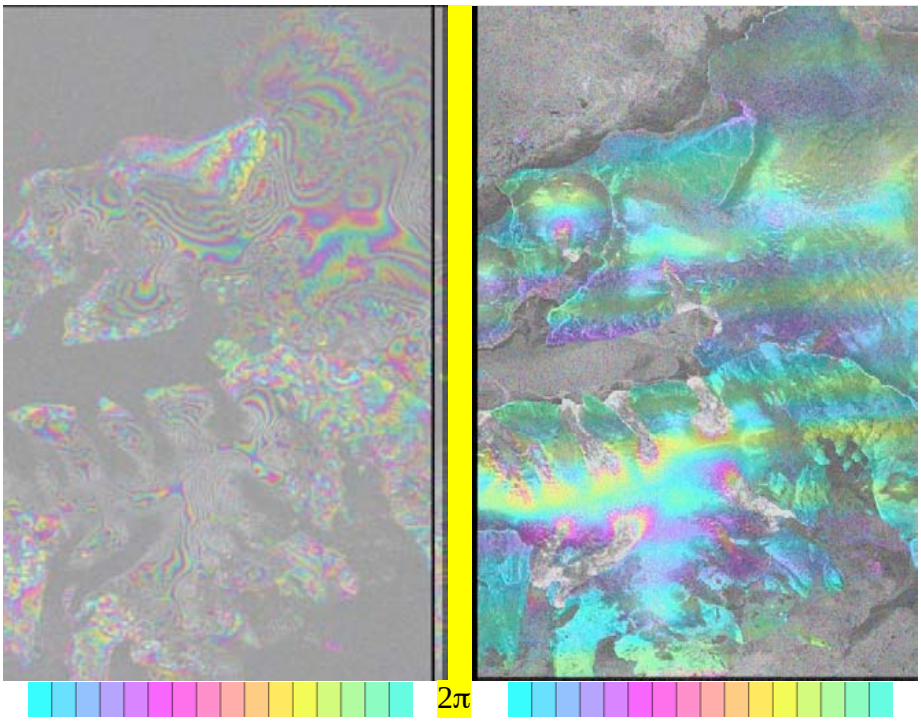

Fig. 6.9: Slant-range and multiple-aperture interferograms from UNISTRA. Image sizes are 2476 x 5579 pixel for InSAR and 825 x 1249 pixel for MAI.

In three out of the four intensity cross-correlation results for Nordaustlandet, all the matching results were provided without filtering. Thus, an evaluation of the post-processing filtering of outliers is not considered.

6.3.3 Breidamerkurjökull, Iceland

Three external results were received for the Breidamerkurjökull test site, they are analysed together with the two results of the internal round robin (Gamma and ENVEO) for this test site. In general, the available five result datasets provide a very good basis for the inter-comparisons as all results were derived by different software packages. Table 6.3 gives an overview of the five datasets (raw displacements without post-processing) for Breidamerkurjökull, together with the most important processing parameters applied by the different groups.

	Gamma	ENVEO	DTU	GFZ	NORUT
Software	GAMMA	IRSL	SUSIE	SARscape	GSAR
Matching window size	128 x 128	128 x 128	128 x 128	64 x 64	44 x 40
Matching window spacing	25 x 25	25 x 25	50 x 50	6 x 6	22 x 20
Oversampling factor	2	16	2	16	4
file info	pixel x	pixel x	pixel x	pixel x	pixel x
	pixel y	pixel y	pixel y	pixel y	pixel y
	disp x	disp x	disp x	disp x	disp x
	disp y	disp y	disp y	disp y	disp y
	CC	CC	SNR	CC	Contrast
output file size (MB)	36.4	16.9	9.36	349	68.4

Table 6.3: Overview of the five datasets over Breidamerkurjökull.

Two repeat pass TerraSAR-X images acquired on 26/08/2008 and 06/09/2008 with a temporal separation of eleven days were provided as Slant range Single look Complex (SSC) images in the DLR TerraSAR-X COSAR format. The following Table 6.4 summarizes the specifications of the provided scenes, Fig. 6.10 illustrates the coverage as well as the local settings.

Sensor	Product	Date	Orbit	Track	Direction	Polarisation	Dimensions
TerraSAR-X	SSC	28.08. 2008	6653	140	descending	HH	17024 / 25424
TerraSAR-X	SSC	06.09. 2008	6820	140	descending	HH	17024 / 25424

Table 6.4: Specifications of the TerraSAR-X scenes used in the round robin.

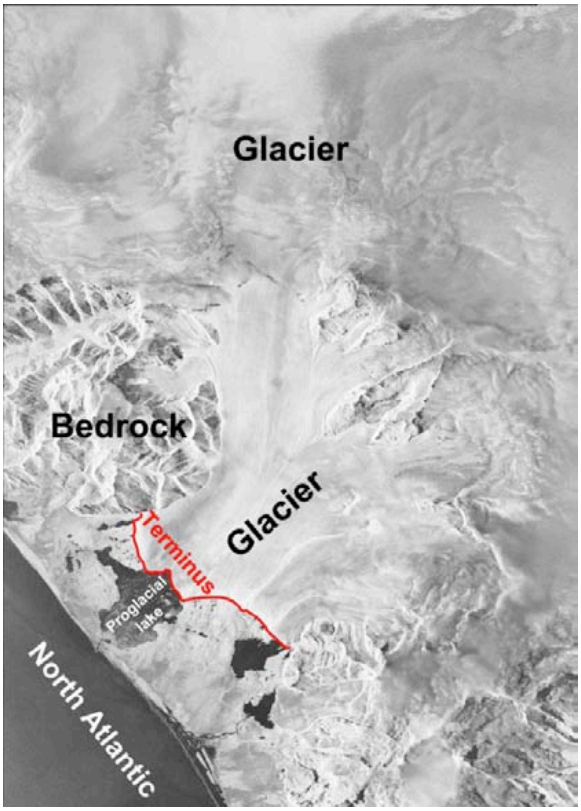

Fig. 6.10: Coverage of the TerraSAR-X scenes provided to the round robin participants.

As a first step, the “raw” output (slant range displacement, azimuth displacement) of the different offset tracking processing chains was analysed. The different groups chose different settings during the offset tracking procedure (matching window size, matching window spacing), resulting in different pixel and x/y sizes. Therefore, accurate co-registration of the datasets was required before an inter-comparison between the results of the different groups could be carried out. This was achieved by referencing all result datasets to the size of the original TerraSAR-X scene (17024 pixels in range, 25424 pixels in azimuth), taking into account the offset introduced by the different starting points for placing the first matching window, as well as the different matching window sizes. In order to allow an inter-

comparison, the range and azimuth displacement was resampled independently using sinc interpolation, to the size of the 10x10 multi-looked TerraSAR-X scene (1702 pixels in range, 2542 pixels in azimuth). All round robin participants delivered the offset tracking results in "pixel displacement" units (in this case range and azimuth shift in pixels/11 days), therefore requiring no adjustment of the pixel values. Based on the range and azimuth shifts we computed 2D velocity in slant range geometry (pixels) for each dataset and used this for the inter-comparisons. Figure 6.11 shows the different results stretched to the same colour scale, super-imposed on the TerraSAR-X amplitude image from 26/08/2008.

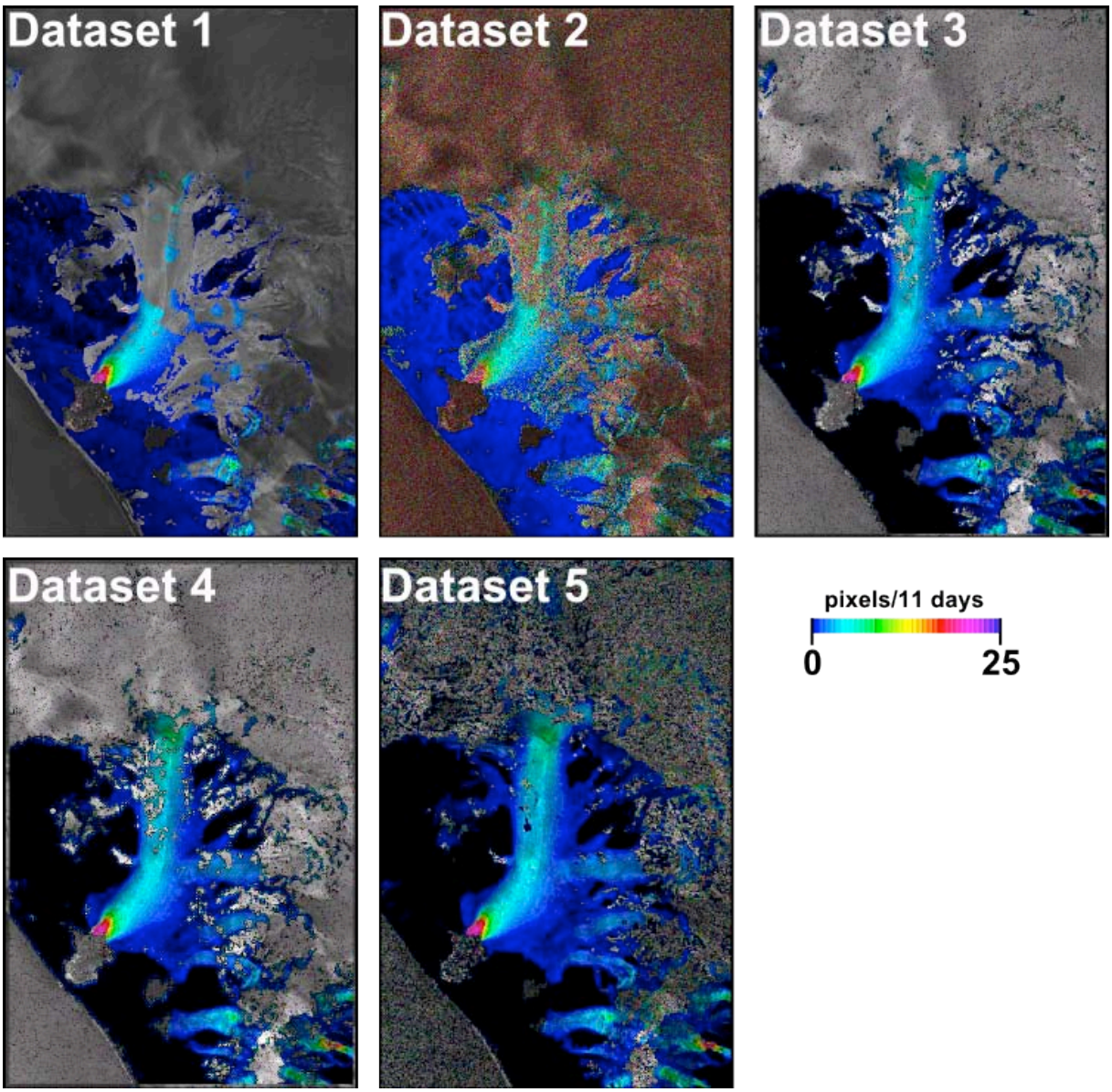

Fig. 6.11: The 2D displacement in slant range geometry stretched to the same colour scale, super-imposed on the TerraSAR-X amplitude image from 26/08/2008.

A qualitative comparison shows that the pattern of ice motion on the glacier is captured quite well by the different datasets; the acceleration of Breidamerkurjökull close to the calving

front as well as the flowband stretching up to the higher reaches of the ice cap can be clearly seen in all results. The datasets have very similar magnitudes, except for dataset 2 which appears to have a slight shift compared to the other results. However, the datasets show large differences in coverage, especially in the upper parts of the glacier.

6.4 Statistical analysis of the results

6.4.1 Comparing results from optical and radar data over Karakoram

As expected from the higher noise level of SAR images, matches of repeat optical data of a similar resolution provide for high-mountain conditions more and more detailed displacements. This becomes especially clear from the small glaciers successfully matched using the optical methods but not using the SAR data, and from problems of SAR-based tracking to represent large velocity gradients, e.g. at glacier margins. Also, layover effects within radar images of mountainous terrain may hide some regions of interest, which in this example (Fig. 6.12) effects the entire northern branch of Baltoro glacier. SAR offset tracking is able to measure displacements better in accumulation areas, where visual contrast in optical images is typically very limited.

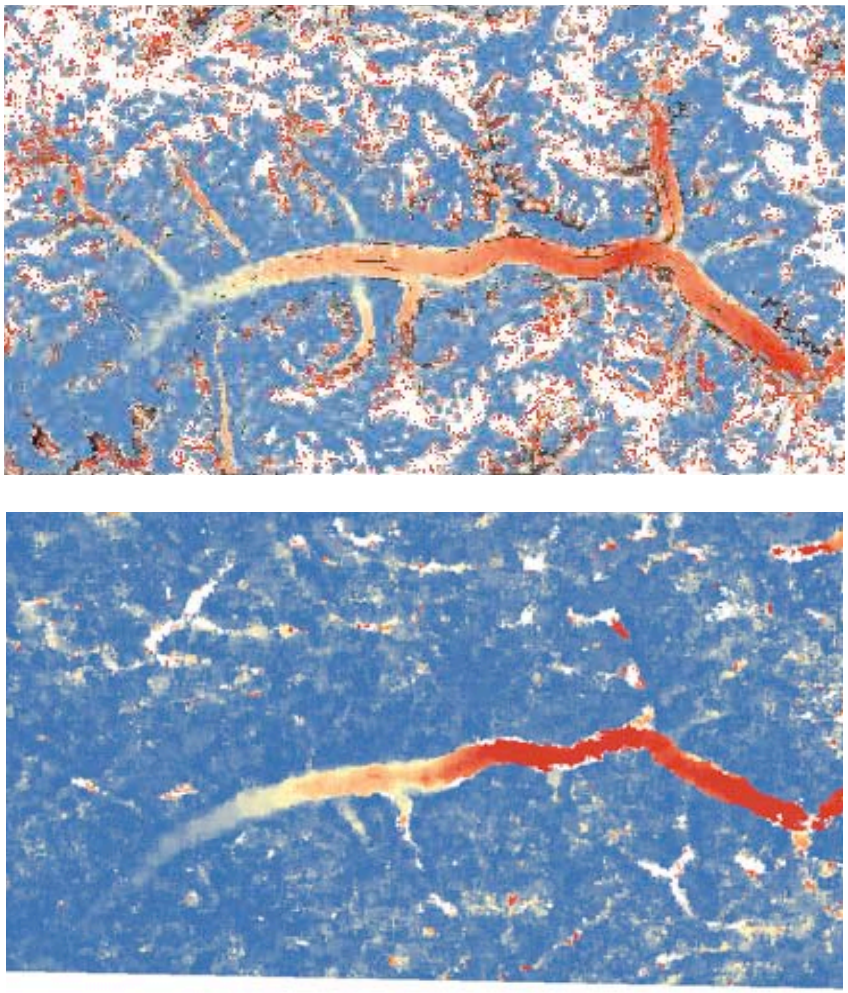

Fig. 6.12: Displacements over Baltoro glacier. Upper panel: from Landsat pan, lower panel: from ENVISAT ASAR. Both from the Swansea method (modified GAMMA).

6.4.2 Nordaustlandet, Svalbard

For a quantitative comparison of the round robin results, a profile over Vestfonna was first extracted (Fig. 6.13). This profile extents well over the fast-flowing outlet glaciers and the more stable interior of the ice cap. Plots of three out of the four groups (results of UNISTRA were computed with the Gamma software and are virtually the same as those provided directly by GAMMA) shows that the participants obtain generally very similar values, but that NORUT has a large noise level.

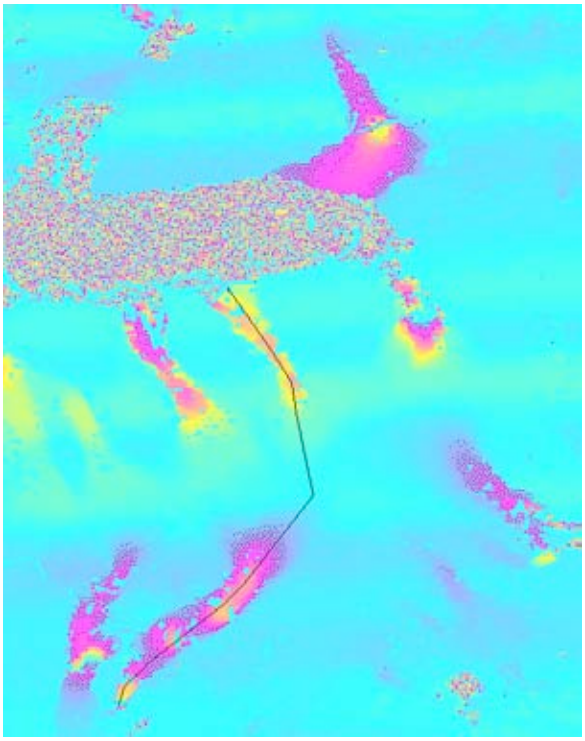

Fig. 6.13: Profile over Vestfonna used for the inter-comparisons, start (pixel 0) is in the upper part of the image.

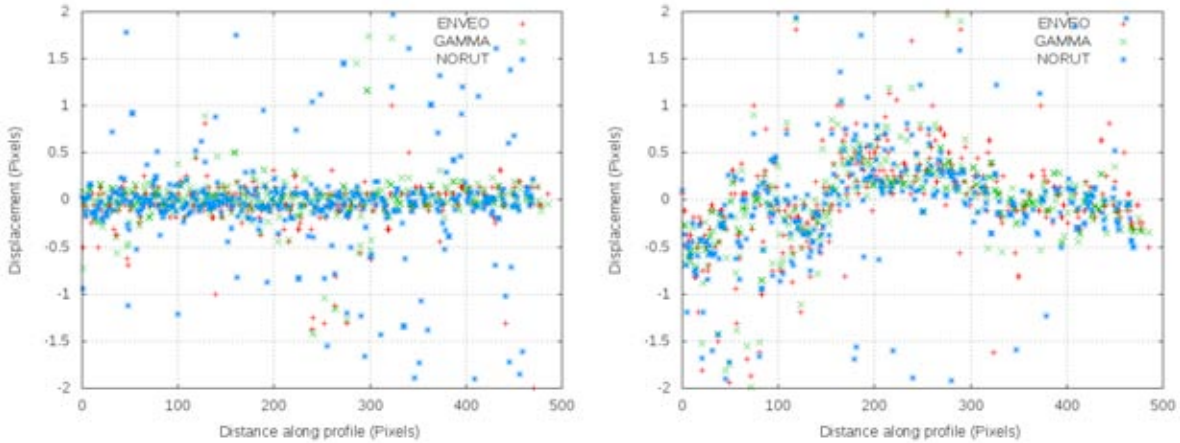

Fig. 6.14: Profile of the round robin results, left slant-range and right azimuth (Enveo: +, Gamma: x and from Norut: ✳).

A further statistical analysis of the retrieval errors is performed for the results of ENVEO and GAMMA using the intensity cross-correlation algorithm. The other three algorithms (fringe-visibility, InSAR and MAI) are not considered further, because already in the qualitative comparison of the previous section their quality was found to be lower than that of the intensity cross-correlation. The intensity cross-correlation results of NORUT are not further considered because their choice to use a smaller matching window size resulted in larger noise level.

Glacier velocities from ALOS PALSAR data are first compared against those from DGPS surveys from Pohjola et al. (2011). Because DGPS data from the geodetic survey campaigns in 2007-2010 are available as horizontal ice surface velocity component, the ALOS PALSAR slant-range and azimuth displacements in pixels were transformed to horizontal ice surface speeds using a ground-range pixel spacing of about 14.95 m (centre of the frame) and an azimuth pixel spacing of 3.14 m. Then, the pixels corresponding to the GPS stakes were considered in the comparison. Results are shown in the following figure. The agreement between DPGS and ALOS PALSAR data is generally satisfactory, considering all the problems related to the validation of glacier velocity data against in-situ measurements explained in section 5.1. The average of the absolute difference between DGPS and ALOS PALSAR results are 9.6 m/yr and 7.6 m/yr in the case of the GAMMA and ENVEO solutions, respectively. Maximum differences are 40.9 m/yr and 25.2 m/yr.

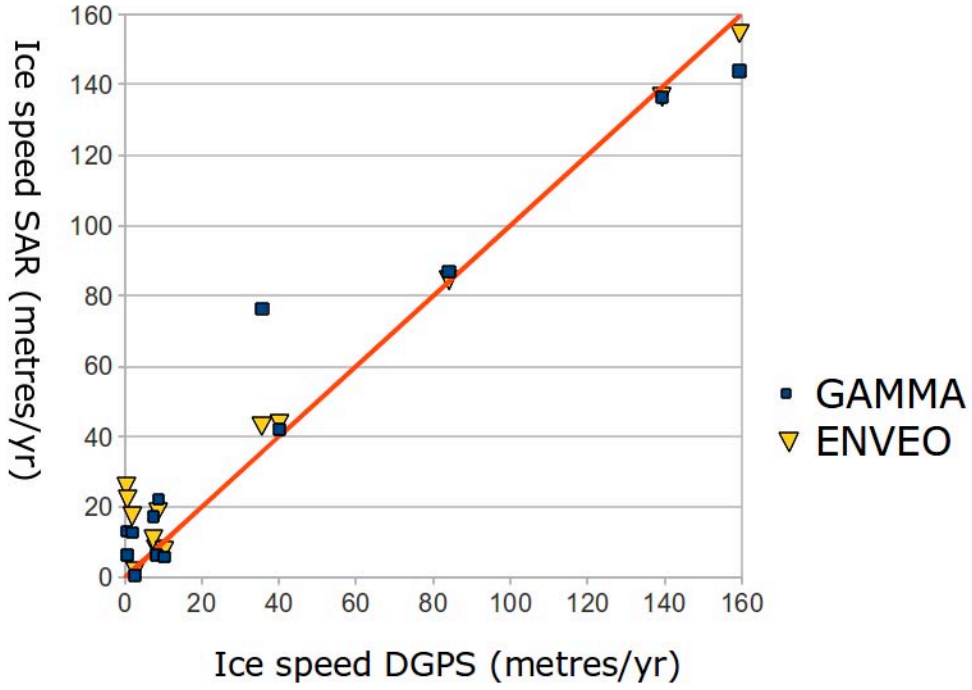

Fig. 6.15: Comparison of InSAR and geodetical (DGPS) ice speeds.

Matching on stable ground in the image set gives indication for the overall co-registration of the repeat images and some general idea of the matching accuracy under the specific image conditions. We extracted the ground-range and azimuth displacements over two areas on stable ground and used a polar plot of all estimates to highlight that there are no systematic co-registration errors for either ENVEO or GAMMA analyses. The standard deviations are 0.69 m in ground-range (i.e. about 1/20th of the range pixel dimension) and 0.52 m in azimuth (i.e. about 1/10th of the azimuth pixel size) for the ENVEO analysis and 0.48 m in ground-range and 0.53 m in azimuth direction for the GAMMA analysis. The spread of the ground-range displacements is larger because ground-range pixel size is larger than azimuth pixel size. These numbers corresponds to total horizontal displacement rates of 6.86 m/yr respectively 5.88 m/yr.

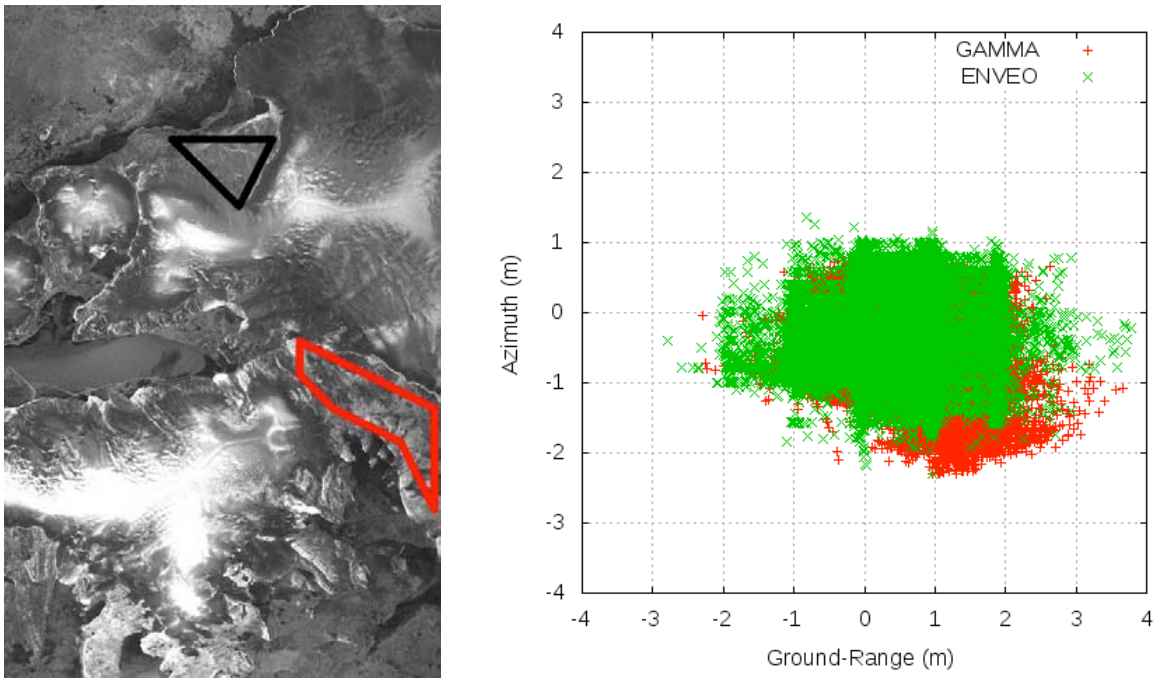

Fig 6.16: Areas on stable ground (blue and red) used for quantitative precision estimates (left) and polar plot of ground-range and azimuth displacements over stable ground (right).

6.4.3 Breidamerkurjökull, Iceland

For a quantitative comparison, we extracted two profiles from the datasets (Fig. 6.17). Profile 1 crosses the glacier tongue close to the terminus where the glacier has the highest flow speeds and a strong gradient towards the lateral area. This profile extents well over bedrock on both sides of the glacier, allowing a reference with non-moving areas. Profile 2 follows the main flowband of Breidamerkurjökull, starting on the stable proglacial Sandur plain stretching well above the equilibrium line into the acculmulation area, where trackable surface features are covered by snow.

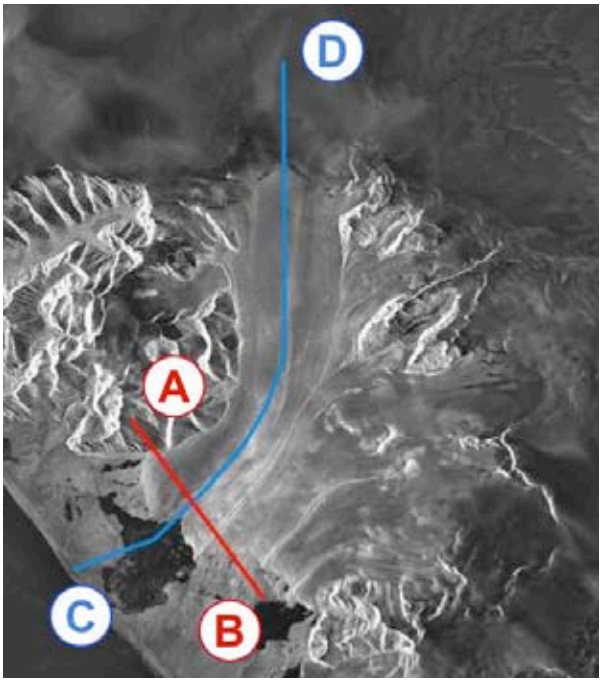

Fig. 6.17: The cross-section and along-flow profiles used for the inter-comparisons.

The cross-section plot (Fig. 6.18) shows that all participants obtain very similar values. Only Dataset 2 has a large noise level. This algorithm fails especially in the fast moving part of the glaciers where the displacement is up to 14.5 pixels. All algorithms perform very well over stable ground. Mean values for the first 40 pixels of the profile (bedrock only) range between 0.108 (Dataset 1) and 0.003 (Dataset 5) pixels. Excluding Dataset 2 we calculate a mean profile of all other datasets and investigate the deviation from this average velocity for every dataset. This comparison shows that all results range between +/- 1 pixel, i. e. below the resolution of the satellite. Largest deviations can be found in the middle part of the profile coinciding with fast-flowing part of the glacier tongue (expect for the outliers in Dataset 1).

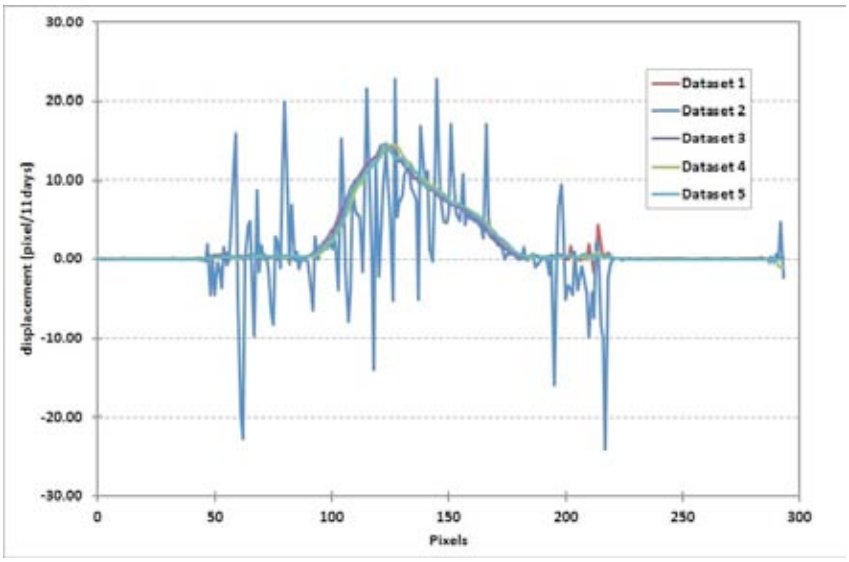

Fig. 6.18: Cross-section profile (A @ 0, B @ 300 in figure 4) of the five results.

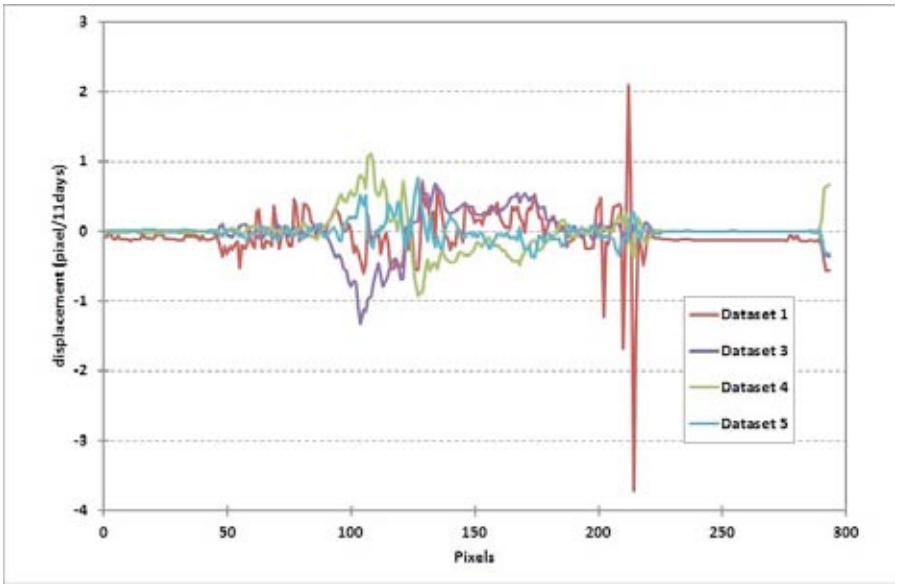

Fig. 6.19: Deviation from mean velocity (excluding Dataset 2); A @ 0, B @ 300 in figure 4.

The centre flowline profile (Fig. 6.20) is generally more difficult to analyse, due to the higher noise level in all datasets. This is caused by the difficult conditions in parts of the profile (e.g. proglacial lake, snow cover) pushing incoherent offset tracking to its limits. In the smooth part of the profile (roughly between 150 and 350 pixels) it can be seen that most results match very well over the fast and slower moving parts. Only Dataset 2 does not capture the highest velocities (up to 20.7 pixels) and seems to produce generally lower values than the other algorithms.

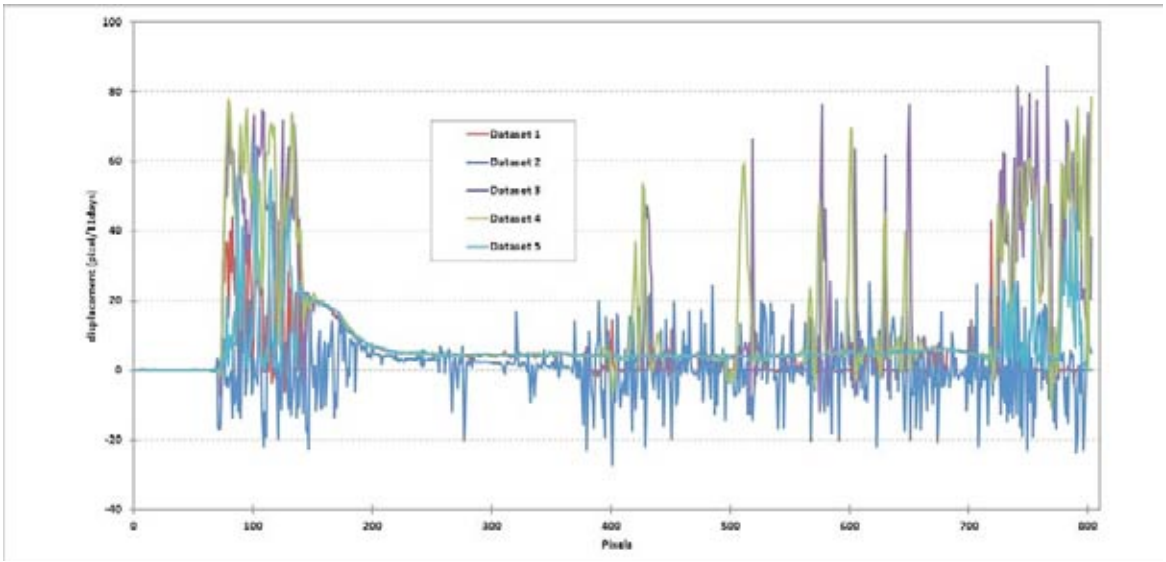

Fig. 6 20: Centre flowline profile (C @ 0, D @ 800 in Fig. 6.17) of the five results.

6.5 Discussion of the investigated algorithms and algorithm selection

Criteria used for algorithm selection are robustness, reliability and accuracy. As demonstrated for the Vest-Austfonna case, intensity cross-correlation performs better compared to the other SAR methods (fringe-visibility, InSAR and MAI), in particular regarding its wider application to different glaciers and SAR data. In the other two considered test-sites interferograms were largely decorrelated with fringes visible only over the non-glaciated areas. In consideration of these obvious results, fringe-visibility and MAI were not even tested with these data.

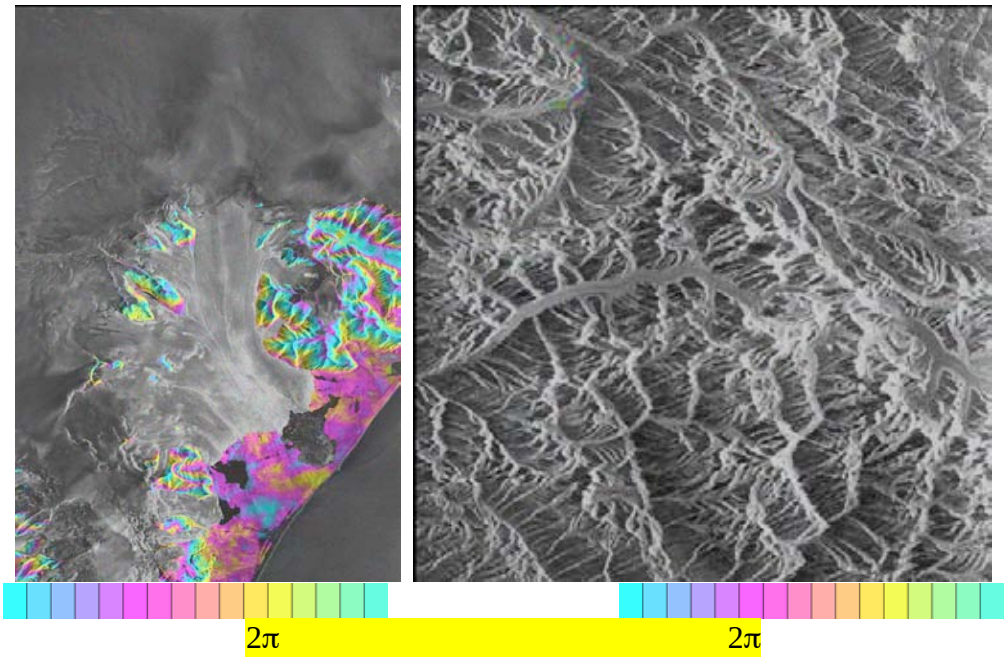

Fig. 6.21: Slant-range interferograms over Breiðamerkurjökull using a summer TerraSar-X image pair separated by 11 days (left) and over Karakoram using an Envisat ASAR image pair separated by ~1 year (right).

In terms of algorithm efficiency, the participants of the round robin were asked to also include computational times. These statistics provide a general overview, but are however biased in terms of variation on computers used and programs applied. While pre- and post-processing are very efficient procedures with processing durations of a few minutes, main processing might be quite computationally intense. For a whole SAR scene processing durations range from about one hour in the best case (Baltoro with a window size of 40x40, 4 pixels spacing and an oversampling factor of 2) to 6 days in the worst case (Breiðamerkurjökull with a window size of 64x64, 6 pixels spacing and an oversampling factor of 16).

From the product validation and algorithm comparison for glacier velocity from SAR data we draw the following **main conclusions**:

- The intensity-cross correlation algorithm is the most robust and efficient algorithm;
- The reliability of the intensity-cross correlation algorithm to return co-registration parameters is as accurate as 1/10th of a SAR image pixel (i.e. about 10 m/yr for ALOS PALSAR and TerraSAR-X data separated by a temporal interval of 46 and 11 day, respectively and about 20 m/yr the ENVISAT ASAR data separated by a temporal interval of 35 days)
- Further investigation is required to automatically tune the matching window size, as in the round robin this step was based on the participants experience
- Further investigation is also required in the post-filtering of the matching outcomes
- The cross-correlation of chips in amplitude SAR images might be implemented in a routine processing of satellite data for ice surface displacement estimation

References

Ahn, Y. and I.M. Howat (2011): Efficient Automated Glacier Surface Velocity Measurement From Repeat Images Using Multi-Image/Multichip and Null Exclusion Feature Tracking. Ieee Transactions on Geoscience and Remote Sensing. 49(8), 2838-2846.

Albert, T.H. (2002): Evaluation of remote sensing techniques for ice-area classification applied to the tropical Quelccaya Ice Cap, Peru. Polar Geography, 26 (3), 210-226.

Andreassen, L.M., F. Paul, A. Kääb and J.E. Hausberg (2008): Landsat-derived glacier inventory for Jotunheimen, Norway, and deduced glacier changes since the 1930s. The Cryosphere, 2, 131-145.

Atwood, D., F. Meyer and A. Arendt (2010): Using L-band SAR coherence to delineate glacier extent. Canadian Journal of Remote Sensing, 36, S186-S195.

Berthier, E., Y. Arnaud, C. Vincent, and F. Rémy (2006): Biases of SRTM in high-mountain areas: Implications for the monitoring of glacier volume changes. Geophysical Research Letters, AGU, 33, L08502.

Berthier, E., Y. Arnaud, R. Kumar, S. Ahmad, P. Wagnon and P. Chevallier (2007): Remote sensing estimates of glacier mass balances in the Himachal Pradesh (Western Himalaya, India). Remote Sensing of Environment, 108, 327-338.

Bishop, M.P. and 16 others (2004): Global Land Ice Measurements from Space (GLIMS): Remote sensing and GIS investigations of the Earth's cryosphere. Geocarto International, 19 (2), 57-85.

Bolch, T., M. F. Buchroithner, A. Kunert and U. Kamp (2007): Automated delineation of debris-covered glaciers based on ASTER data. In: M. A. Gomarasca (Ed.): GeoInformation in Europe (Proc. 27th EARSeL-Symposium, 4.-7.6.07, Bolzano/Bozen, Italy), Millpress, Netherlands, 403-410.

Bolch, T., M. F. Buchroithner, J. Peters, M. Baessler, and S. Bajracharya (2008). Identification of glacier motion and potentially dangerous glacial lakes in the Mt. Everest region/Nepal using spaceborne imagery. Natural Hazards and Earth System Science, 8, 1329-1340.

Bolch, T., B. Menounos and R. Wheate (2010): Landsat-based glacier inventory of western Canada, 1985-2005. Remote Sensing of Environment, 114 (1), 127-137.

Debella-Gilo M. and Kääb A. (2012): Locally adaptive template sizes for matching repeat images of Earth surface mass movements. ISPRS Journal of Photogrammetry and Remote Sensing, 69, 10-28

Dozier, J. (1989): Spectral signature of alpine snow cover from Landsat 5 TM. Remote Sensing of Environment, 28, 9-22.

Duan, Q., S. Sorooshian, and V.K. Gupta (1994): Optimal use of the SCE-UA global optimization method for calibrating watershed models. Journal of Hydrology, 158(3-4), 265-284.

Frey, H. and F. Paul (in press): On the suitability of the SRTM DEM and ASTER GDEM for the compilation of topographic parameters in glacier inventories. International Journal of Applied Earth Observation and Geoinformation.

Frey, H., F. Paul and T. Strozzi (in revision): Compilation of a glacier inventory for the western Himalayas from satellite data: Methods, challenges and results. Remote Sensing of Environment.

Fricker, H. A., A. Borsa, B. Minster, C. Carabajal, K. Quinn, and B. Bills (2005): Assessment of ICESat performance at the Salar de Uyuni, Bolivia. Geophysical Research Letters, 32, L21S06.

Gardelle, J., E. Berthier and Y. Arnaud (2012): Correspondence: Impact of resolution and radar penetration on glacier elevation changes computed from DEM differencing. Journal of Glaciology, 58 (208), 419-422.

Gjermundsen, E.F., R. Mathieu, A. Kääb, T. Chinn, B. Fitzharris and J.O. Hagen (2011): Assessment of multispectral glacier mapping methods and derivation of glacier area changes, 1978-2002, in the central Southern Alps, New Zealand, from ASTER satellite data, field survey and existing inventory data. Journal of Glaciology, 57 (204), 667-683.

Glaciers_cci (2011): Data Access Requirements Document (DARD): Prepared by the Glaciers_cci consortium, 46 pp.

Glaciers_cci (2012a): Product Validation Plan (PVP). Prepared by the Glaciers_cci consortium, 58 pp.

Glaciers_cci (2012b): Algorithm Theoretical Basis Document Version 0 (ATBDv0). Prepared by the Glaciers_cci consortium, 95 pp.

Gruen, A. and D. Akca (2005): Least squares 3D surface and curve matching. ISPRS Journal of Photogrammetry and Remote Sensing, Elsevier Science Bv, 59, 151-174.

Heid, T. and A. Kääb (2012): Evaluation of existing image matching methods for deriving glacier surface displacements globally from optical satellite imagery. Remote Sensing and Environment. 118, 339-355.

Howat, I.M., Smith, B.E., Joughin, I. and Scambos, T.A. (2008): Rates of southeast Greenland ice volume loss from combined ICESat and ASTER observations. Geophysical Research Letters, 35, L17505.

Huggel, C., A. Kääb, W. Haeberli, P. Teysseire and F. Paul (2002): Remote sensing based assessment of hazards from glacier lake outbursts: a case study in the Swiss Alps. Canadian Geotechnical Journal, 39 (2), 316-330.

Kääb, A. (2005): Remote Sensing of Mountain Glaciers and Permafrost Creep. Geographisches Institut der Universit Zürich.

Khvorostovsky, K. (2012): Merging and analysis of elevation time series over Greenland ice sheet from satellite radar altimetry. IEEE Transactions on Geoscience and Remote Sensing, 50 (1), 23-36.

Le Bris, R., F. Paul, H. Frey and T. Bolch (2011): A new satellite-derived glacier inventory for western Alaska. Annals of Glaciology, 52 (59), 135-143.

Miller, P., J. Mills, S. Edwards, P. Bryan, S. Marsh, H. Mitchell and P. Hobbs (2008): A robust surface matching technique for coastal geohazard assessment and management. ISPRS Journal of Photogrammetry and Remote Sensing, Theme Issue: Remote Sensing of the Coastal Ecosystems, 63, 529-542.

Miller, P. E., M. Kunz, J.P. Mills, M.A. King, T. Murray, T.D. James and S.H. Marsh (2009): Assessment of Glacier Volume Change Using ASTER-Based Surface Matching of Historical Photography. IEEE Transactions on Geoscience and Remote Sensing, 47, 1971-1979.

Moholdt, G., J.O. Hagen, T. Eiken and T.V. Schuler (2010a): Geometric changes and mass balance of the Austfonna ice cap, Svalbard. The Cryosphere, 4, 21-34.

Moholdt, G., C. Nuth, J.O. Hagen and J. Kohler (2010b): Recent elevation changes of Svalbard glaciers derived from ICESat laser altimetry. Remote Sensing of Environment, 114 (11), 2756-2767.

Nuth, C. and A. Kääb (2011): Co-registration and bias corrections of satellite elevation data sets for quantifying glacier thickness change. The Cryosphere, 5, 271-29.

Paul, F. (2002): Changes in glacier area in Tyrol, Austria, between 1969 and 1992 derived from Landsat 5 TM and Austrian Glacier Inventory data. International Journal of Remote Sensing, 23 (4), 787-799.

Paul, F. (2008): Calculation of glacier elevation changes with SRTM: is there an elevation-dependent bias? Journal of Glaciology, 54, 945-946.

Paul, F. and A. Kääb (2005): Perspectives on the production of a glacier inventory from multispectral satellite data in the Canadian Arctic: Cumberland Peninsula, Baffin Island. Annals of Glaciology, 42, 59-66.

Paul, F. and L.M. Andreassen (2009): A new glacier inventory for the Svartisen region, Norway, from Landsat ETM+ data: challenges and change assessment. Journal of Glaciology, 55 (192), 607-618.

Paul, F. and J. Hendriks (2010): Optical remote sensing of glaciers. In: Pellikka, P. and Rees, W.G (Eds.)., Remote Sensing of Glaciers - Techniques for Topographic, Spatial and Thematic Mapping of Glaciers. CRC Press, Taylor and Francis Group, Leiden, 137-152.

Paul, F., A. Kääb, M. Maisch, T. Kellenberger and W. Haeberli (2002): The new remote-sensing-derived Swiss glacier inventory: I. Methods. Annals of Glaciology, 34, 355-361.

Paul, F., Huggel, C., Kääb, A. and Kellenberger, T. (2003): Comparison of TM-derived glacier areas with higher resolution data sets. EARSeL Workshop on Remote Sensing of Land Ice and Snow, Bern, 11.-13.3.2002. EARSeL eProceedings, 2, 15-21.

Paul, F., C. Huggel and A. Kääb (2004): Combining satellite multispectral image data and a digital elevation model for mapping of debris-covered glaciers. Remote Sensing of Environment, 89 (4), 510-518.

Paul, F., L.M. Andreassen, S.H. and Winsvold, S.H. (2011a): A new glacier inventory for the Jostedalsbreen region, Norway, from Landsat TM scenes of 2006 and changes since 1966. Annals of Glaciology, 52 (59), 153-162.

Paul, F., H. Frey and R. Le Bris (2011b): A new glacier inventory for the European Alps from Landsat TM scenes of 2003: Challenges and results. Annals of Glaciology, 52 (59), 144-152.

Pohjola, V.A. , P. Christoffersen, L. Kolondra, J.C. Moore , R. Pettersson, M. Schäfer, T. Strozzi, and C.H. Reijmer, 2011. Spatial distribution and change in the surface ice-velocity field of Vestfonna ice cap, Nordaustlandet, Svalbard, 1995– 2010 using geodetic and satellite interferometry data. Geografiska Annaler:, Series A, Physical Geography, 93, 323–335.

Pritchard, H.D., R.J. Arthern, D.G. Vaughan and L.A. Edwards (2009): Extensive dynamic thinning on the margins of the Greenland and Antarctic ice sheets. Nature, 461, 971-975.

Racoviteanu, A.E., M.W. Williams and R.G. Barry (2008): Optical remote sensing of glacier characteristics: a review with focus on the Himalaya. Sensors, 8, 3355-3383.

Racoviteanu, A.E, F. Paul, B. Raup, S.J.S. Khalsa and R. Armstrong (2009): Challenges in glacier mapping from space: recommendations from the Global Land Ice Measurements from Space (GLIMS) initiative. Annals of Glaciology, 50 (53), 53-69.

Raup, B.H., and 11 others (2007): Remote sensing and GIS technology in the Global Land Ice Measurements from Space (GLIMS) Project. Computers and Geosciences, 33, 104-125.

Rinne, E.J., A. Shepherd, S. Palmer, M.R. van den Broeke, A. Muir, J. Ettema and D. Wingham (2011): On the recent elevation changes at the Flade Isblink Ice Cap, northern Greenland. 2011. Journal of Geophysical Research 116: 10.1029/2011JF001972.

Rodriguez, E., C.S. Morris and J.E. Belz (2006): A global assessment of the SRTM performance. Photogrammetric Engineering and Remote Sensing, 72, 249-260.

Schenk, T., B. Csatho, C.J. van der Veen, H. Brecher, Y. Ahn and T. Yoon (2005): Registering imagery to ICESat data for measuring elevation changes on Byrd Glacier, Antarctica. Geophysical Research Letters, 32, L23S05.

Shukla, A., M.K. Arora and R.P. Gupta (2011): Synergistic approach for mapping debris-covered glaciers using optical-thermal remote sensing data with inputs from geomorphometric parameters. Remote Sensing of Environment, 114 (7), 1378-1387.

Strozzi, T., F. Paul and A. Kääb (2010): Glacier mapping with ALOS PALSAR data within the ESA GlobGlacier project. Proceedings of the ESA Living Planet Symposium, 28 June - 2 July 2010, Bergen, Norway, ESA SP-686.

Svoboda, F. and F. Paul (2009): A new glacier inventory on southern Baffin Island, Canada, from ASTER data: I. Applied methods, challenges and solutions. Annals of Glaciology, 50 (53), 11-21.

Toutin, T. (2008): ASTER DEMs for geomatic and geoscientific applications: a review. International Journal of Remote Sensing, 29, 1855-1875.

Tachikawa, T., M. Kaku and A. Iwasaki (2011): ASTER GDEM Version 2 Validation Report. Earth Remote Sensing Data Analysis Centre (ERSDAC).

VanLooy, J. A. (2011): Analysis of elevation changes in relation to surface characteristics for six glaciers in Northern Labrador, Canada using advanced space-borne thermal emission and reflection radiometer imagery. Geocarto International, Geocarto International, Taylor and Francis, 26, 167-181.

Wingham, D.J., A. Ridout, R. Scharroo, R. Arthern and C.K. Shum (1998): Antarctic elevation change from 1992 to 1996. Science, 282, 456-458.

Zwally, H.J., A.C. Brenner, J.A. Major, R.A. Bindschadler and J.G. Marsh (1989): Growth of Greenland Ice-Sheet - Measurement. Science, 246, 1587-1589.

Abbreviations

ALOS	Advanced Land Observing Satellite
ASAR	Advanced Synthetic Aperture Radar
ASTER	Advanced Spaceborne Thermal Emission and Reflection radiometer
CC	Correlation Coefficient
CCI	Climate Change Initiative
CDED	Canadian Digital Elevation Dataset
CGIAR	Consultative Group on International Agricultural Research
CPU	Central Processing Unit
DARD	Data Access Requirements Document
DEM	Digital Elevation Model
DGPS	Differential GPS
dDEM	differential DEM
DLR	German Aerospace Center
DP-RT-RepAlt	DEM sloPe Repeat Track Repeat Altimetry
DS-RT-RepAlt	DEM Subtracting Repeat Track Repeat Altimetry
DTED	Digital Terrain Elevation Data
ELA	Equilibrium Line Altitude
ERS	European Remote Sensing Satellite
ESA	European Space Agency
ETM+	Enhanced Thematic Mapper plus
FCDR	Fundamental Climate Data Record
FoG	Fluctuation of Glaciers
GCOS	Global Climate Observing System
GDB	GLIMS Data Base
GDEM	Global DEM
GIC	Glaciers and Icecaps
GIS	Geographic Information System
GLAS	Geoscience Laser Altimeter System
GLIMS	Global Land Ice Measurements from Space
GLS	Global Land Survey
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GV	Glacier Velocities Product
ICESat	Ice, Cloud, and Elevation Satellite
ID	IDentification number
IGOS	Integrated Global Observing Strategy
InSAR	Interferometric SAR

L1T	Level 1 T (terrain corrected)
LDCM	Landsat Data Continuity Mission
LIDAR	LIght Detection And Ranging
LOS	Line of Sight
MAI	Multiple Aperature InSAR
NASA	National Aeronautic and Space Administration
NDSI	Normalized Difference Snow Index
PALSAR	Phased Array type L-band SAR
PSD	Product Specifications Document
PVP	Product Validation Plan
RA XO	
RMSE	Root Mean Square Error
RP-RT-RepAlt	Rectangular Plane Repeat Track Repeat Altimetry
RR	Round Robin
SAR	Synthetic Aperture Radar
SD	Standard Deviation
SEC	Surface Elevation Change
SLC	Scan Line Corrector
SNR	Signal to Noise Ratio
SoW	Statement of Work
SPOT	System Pour l'Observation de la Terre
SRTM	Shuttle Radar Topography Mission
SSC	Slant range Single look Complex
SWIR	Short Wave InfraRed
TM	Thematic Mapper
TP-RT-RepAlt	Triangular Plane Repeat Track Repeat Altimetry
URD	User Requirements Document
USGS	United States Geological Survey
UTM	Universal Transverse Mercator
WGI	World Glacier Inventory
WGMS	World Glacier Monitoring Service