Uncertainty Characterization Report version 2 (UCRv2)

glaciers cci

Prepared by: Glaciers_cci consortium

Contract:	4000101778/10/I-AM
Name:	Glaciers_cci-D6.2-UCRv2
Version:	0.6
Date:	30.06. 2013

Contact:
Frank Paul
Department of Geography
University of Zurich
Winterthurerstr. 190
CH-8057 Zurich
frank.paul@geo.uzh.ch

Technical Officer:
Stephen Plummer
ESA ESRIN

Document status sheet

Version	Date	Changes	Approval
0.1	31.08. 2012	Draft ToC according to the CCI Project Guidelines	
0.3	30.03. 2013	Revised v1 based on Feedback from TO	
0.4	30.03. 2013	Feedback from TO integrated	
0.5	04.05. 2013	Extended version of v1	
0.6	30.06. 2013	Feedback from TO integrated	F. Paul

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), Thomas Nagler, Kilian Scharrer (Enveo), Tazio Strozzi (Gamma), Andreas Kääb, Christopher Nuth (GUIO), Andrew Shepherd, Francesca Ticconi (SEEL)

Glaciers_cci Technical Officer at ESA:
Stephen Plummer

Contents

1. Introduction

According to the CCI project guidelines (ESA, 2010), each CCI project should provide a separate document describing the product uncertainties as deliverable 6.1 Uncertainty Characterization Document (UCD). This is the second version of this document (UCRv2) for the Glaciers_cci project. As a common frame of reference, it repeats at first section 6.1 of the project guidelines in Ch. 2. Afterwards we describe for each of the four products glacier area, elevation changes from altimetry and DEM differencing, and velocity the sources of error and uncertainties along with the methods to quantify them. This information is taken from the PVP (Glaciers_cci, 2012a) and the PVASR (Glaciers_cci, 2012b), as they had to be introduced there to understand the selection of the validation sites and the set-up of the round robin. In this second version of the document we have added information on how the accuracy is finally determined considering the mostly very limited possibilities for proper product validation.

2. Definition of terms

2.1 Describing error and uncertainty

A measurement is a set of operations having the object of determining the value of a quantity. Following BIPM (2008) it is helpful to define the term measurand as

- **Measurand**: particular quantity subject to measurement

so that the phrases 'true value of a quantity' and value of the measurand are synonymous. Very few instruments directly measure the measurand. Generally an instrument reports the effect of a quantity from which the magnitude of the measurand is estimated. As an example, an instrument sensitive to infrared light might be used to measure the temperature of an object. The process of measurement is inexact, so that difference between a measured value and the value of the measurand is called the error. Traditionally (e.g. Beers, 1975) the word 'error' has also meant a numerical value that estimates the variability of the error if a measurement is repeated (i.e. a width of the distribution of possible errors). This dual meaning of "error" can lead to confusion or ambiguity. To separate these meanings and avoid confusion the BIPM (2008) definitions are used, i.e.

- **Error (of measurement)**: result of a measurement minus a true value of the measurand
- **Uncertainty (of measurement)**: is a parameter, associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand.

Except in a few cases the "true" value of the error is not known, and the magnitude of the error is hypothetical. An error is viewed as having a random component and a systematic component.

Following BIPM (2008) the definitions of these terms are:

- **Random error:** result of a measurement minus the mean that would result from an infinite number of measurements of the same measurand carried out under repeatable conditions,
- **Systematic error:** mean that would result from an infinite number of measurements of the same measurand carried out under repeatable conditions minus the true value of the measurand.

In general terms the random error is variable from measurement to measurement, whereas the systematic error is the same for each measurement. Although it is not possible to compensate for the random error, its effect on uncertainty in our estimate of the measurand can usually be reduced by averaging over a number of independent repeat observations.

The statistical distribution of random error can be described by a probability density function (pdf) of which the **expected value** (i.e., the average over the pdf) is zero. As the random error

often arises from the addition of many effects the central limit theorem suggests that a Gaussian distribution is a good representation of this pdf. Therefore the random uncertainty value commonly adopted for a single observation is equal to the one-sigma standard deviation that would be obtained from repeated measurements of the same quantity under the same conditions. If N repeated uncorrelated observations are available, the random uncertainty is the one-sigma standard deviation multiplied by a factor of $1/\sqrt{N}$ (under the Gaussian assumption). The smallest possible change in value that can be observed can be taken as 1⁄2 the uncertainty. This value can also be used as the detection limit of the instrument.

The total uncertainty attributed is the combination of this random uncertainty and systematic uncertainty. Often a correction can be applied to compensate for the systematic effects. It is assumed that correction is done such that, after correction, the expected value of the error arising from a systematic effect is zero. A systematic uncertainty remains, however, characterized by the uncertainty in the correction. There are many reasons why a measurement is uncertain. For example, error components in satellite remote sensing may include terms such as

- instrument noise,
- error arising from simplifications in radiative transfer,
- calibration error,
- geolocation/interpolation error,
- error arising from the uncertainty in parameters used to derive the measurement.

Measurement here is used to include satellite retrievals (estimates by some process of inversion) of measurands, although by some strict usage of "measurement", it is typically radiance that a sensor on a satellite actually measures.

An **Uncertainty budget** is a list of random and systematic errors with estimates of the uncertainty they contribute to the measurement (preferably with information about how component uncertainties combine). Standard methods of error propagation (e.g. Hughes and Hase, 2010) are used to transform uncertainties into measurement units. The total uncertainty is the total combined accounting for any correlation between component errors.

In some cases the measurement process returns a vector of measurands. The error between the components of the measurand may not be independent so is represented by an uncertainty covariance matrix of which each element i,j is defined by the expectation value $<\varepsilon_i\varepsilon_j>$ of the product of the respective errors ε_i of the ith measurand. If the measurands are independent then the off-diagonal terms are zero and the uncertainty on each measurand is given by the square-root of the corresponding diagonal element. For vector measurements, the uncertainty budget is a list of random and systematic errors with estimates of their associated uncertainty covariance matrices.

Two qualitative terms not defined in BIPM (2008) but commonly used to describe a measurement (e.g. Beers, 1957, Hughes and Hase, 2010) are precision and accuracy defined here as:

- **precision**: a measurement with a small random uncertainty is said to have high precision
- **accuracy**: a measurement with a small systematic uncertainty is said to have high accuracy

2.2 Validation of Measurements

Validation is the assessment of a measurement and the uncertainty attributed to it. This is principally achieved by external validation, i.e. comparison of a measurement to an independent measurement and assessment of their consistency relative to their estimated uncertainties. This independent estimate of the measurand is termed the validation value. The discrepancy is then defined as

- **discrepancy**: the difference between the measurement and the validation value

A small average discrepancy with respect to the root-sum-square of the measurement and validation value uncertainties is indicative of an accurate measurement, but could also result from a fortuitous cancellation of error terms.

For a small number of measurements it is possible to report individual discrepancies. However, for the large number of measurements typical of satellite remote sensing validation involves statistically characterising the discrepancies. There are often regimes of instrument behaviour for which uncertainties can be expected to differ, so it is usual to characterize discrepancies for the minimum number of regimes of consistent instrument behaviour. The choice of regimes could come from a cluster analysis of discrepancy (if the difference in regimes causes differences in systematic error), but more commonly comes from knowledge of the measurement process.

The statistical characterization of the discrepancies within a regime is made through three **quality parameters**. Consider the set of n measurements $\{x_1 \pm \delta x_1,\ x_2 \pm \delta x_2,\ x_3 \pm \delta x_3,\ \ldots\ x_n \pm \delta x_n\}$ of some quantity together with the set of validation values $\{v_1 \pm \delta v_1,\ v_2 \pm \delta v_2,\ v_3 \pm \delta v_3,\ \ldots\ v_n \pm \delta v_n\}$ made of the same quantity. The quality parameters are then:

- **Bias b**: the mean value of the discrepancy, i.e.:

$$b = [\ \Sigma_{i=1}^{n} (x_i - v_i)\ ] / n$$

- **Chi-squared** χ^2: the goodness of fit between the actual and estimated uncertainties of measurement and validation values, defined by:

$$\chi^2 = [\ \Sigma_{i=1}^{n} (x_i - v_i)^2 / (\delta x_i^2 + \delta v_i^2)\ ] / n$$

- **Stability s**: the change in bias with time defined as:

$$s = [\ b(t+\Delta t) - b(t)] / \Delta t$$

The expectation value of the bias is the sum of the residual systematic errors in the measurement and the validation value. The bias can only be attributed to the measurement if the residual systematic error in the validation value is known a priori. In an ideal case the bias would be zero.

The expected value for χ^2 is unity. A value lower than this indicates the uncertainties attributed to the measurements or the validation values or both are too high. A value greater than unity

indicates the uncertainties attributed to the measurements or the validation values or both are too low.

In the ideal case the stability would be zero over any timescale. In remote sensing the stability can display periodicity related to factors such as instrument drift or solar illumination of the satellite – both over an orbit and seasonally. It is suggested that the stability is estimated at the same temporal scale that any trends in the data are calculated.

It may be that the quality parameters are independent of the measurement magnitude and conditions of measurement and apply at all locations and times. In that case the three quality values adequately characterize the quality of measurement. More commonly, the quality values vary so a **validation table** is used to summarise the bias, χ^2 and stability for regimes of consistent instrument behaviour.

In some case **internal validation** can be used to check reported uncertainty. Consider the situation where an instrument measures the same quantity under conditions where the reported uncertainty does not vary. Then the variability of the measurements should agree with the reported random uncertainty.

2.3 Comparing Measurements with a Model

Further understanding can be achieved through comparison of measurements with model output. In this approach, a model is sampled to give model values at the same place and time as the measurement values. The same three quality parameters can be calculated. However these caveats apply:

- the model error may not be reported and may have to be assumed,
- the bias cannot be attributed to the model or measurements without reference to additional information

An estimate of interpolation uncertainty must be included if the model reports results at different times and location from the measurements so that the model results are interpolated to the measurement location.

If the model is at a coarser resolution than the measurements, an approach could be to compare the model value with a (weighted) average of the measurements. The fact that the systematic uncertainty is correlated needs to be accounted for if this approach is taken.

The statistical comparison of model and measurement data must account for bias due to sampling. For example a monthly time series comparison between model output and averaged measurements may show bias due to conditions, such as cloud coverage, under which measurements are not possible.

3. Glacier area

3.1 Sources of errors and uncertainties

3.1.1 Impact of the algorithm used for glacier classification

A wide range of methods was and still is applied to map glaciers from optical images (e.g. to 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 ATBDv1 (Glaciers_cci, 2012c). 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 existing algorithms we here exclude manual delineation as this is an important method for improvement of product quality (e.g. adding debris covered parts) and is 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 remaining 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).

The key classification step when applying one of the band ratio methods is the (manual) selection of a threshold value to convert the ratio image in a binary glacier map. A potential additional threshold has to be selected if the TM1 equivalent band is used for improved mapping in cast shadow (cf. Paul and Kääb, 2005). Under otherwise perfect mapping conditions, these two threshold values determine the accuracy of the product. If wrongly selected, too large or too small glacier areas result and the workload required for manual corrections can increase substantially. The main rule for threshold selection is thus the minimization of the workload for post-processing and this mainly concerns glacier parts in cast shadow as debris cover cannot be mapped with this method anyway. As a second step, it has to be considered that the application of a median filter (to reduce noise) also alters glacier extent. And finally, it has to be noted that all glacier outlines are visually controlled and corrected against the satellite image or other available datasets where required. In this regard, the above error sources (threshold, median filter) have more of a theoretical nature.

3.1.2 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 intro-

duced 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.

3.1.3 Post-classification issues

After a raw glacier map has been created, 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). 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 can be derived, the individual glacier entities. This step requires a co-registered DEM to derive drainage divides and digitally intersect them with the corrected 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; Rastner et al., 2012). In particular the division of ice caps into hydrologic catchments does often not make much sense and can thus be discussed controversial. In this regard the drainage divide issue is a methodological problem (result changes with the purpose) rather than a technical one.

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. In addition, while 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).

The fully 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 when snow conditions are satisfactory. This step has thus to be done with great care to meet the accuracy specifications for the glacier area product (better than 5%).

3.1.4 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., 2012). For example, in the regions with voids in the SRTM DEM caldera like structures were visible in the hillshade of the DEM, pointing to a systematic underestimation of elevation in these regions. In consequence, geolocation shifts of about 5 pixels (150 m) or more were found by Frey et al. (2012) for data void regions compared to an independent dataset (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. There is also the possibility to inform the USGS customer service in the case of detected errors.

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, non-systematic 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 do not use vector data as available from independent sources (e.g. national mapping agencies) for product validation.

3.2 Methodology to determine uncertainties

3.2.1 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 3.2.2). In regard to reference data, the major problem is that they 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 ‘reference dataset’ (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.

3.2.2 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 3.2.1 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.

3.2.3 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 considered, quantitative measures can be applied to assess product accuracy. They include 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 including a further tributary). In such a case the area difference alone has little meaning.

Another 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 3.1 we provide an overview on the accuracy assessments that can be performed and the details to assess them. The calculation of the quantitative accuracy measures is based on calculations of the glacier area as implemented in the GIS with subsequent statistical analysis of the derived values (e.g. mean, standard deviation). 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 3.1: Overview of the different possibilities to assess product accuracy for glacier area.

3.3 Accuracy to be reported

Given the range of possibilities for accuracy assessment and error characterization described above, the question arises which of these measures will be reported along with a dataset? At the first level of this decision one has to distinguish between (A) the mandatory entries in the GLIMS database and (B) the accuracy reported in a paper. In regard to (A) it is required to provide with each glacier outline the global uncertainty in x and y direction (that is related to geographic location and can be found in the metadata of each satellite scene) as well as the local (within-image) uncertainty (see Table 3.2 in the PSD). The latter can in principle even be applied to individual segments of a glacier outline line as mapping accuracy of clean-ice parts can be very different from manually edited debris-covered parts. However, practically it is both very difficult and time consuming to determine accuracy in such a detailed way. Hence, this information is in general only provided for all glaciers in a scene in the same way (e.g. ±1 pixel for the global and ±2 pixels for the local uncertainty).

The GLIMS database also has a link to associated publications that can be used to describe other uncertainty measures (from Table 3.1). Depending on the availability of an appropriate reference dataset and in view of the project results achieved so far, we recommend a tiered strategy for accuracy assessment:

1. At the first level, glacier outlines can be compared to an independent dataset of a likely better quality. The latter means it has been derived from a higher-resolution dataset with the same or even better mapping conditions (e.g. less seasonal snow) in a year close to the acquisition date by a well experienced analyst (indeed, these constraints are difficult to match). Area differences should be calculated and the following information provided with each study:
(1a) mean area values for both samples, number of glaciers in the sample, absolute differences (accuracy) and standard deviation;
(1b) an overlay of glacier outlines as derived from both datasets;
(1c) a description and reasoning for the differences in text form.

2. When a reference dataset is not available, an independent multiple digitization experiment should be performed (minimum five different glaciers three times). This will provide important information on the analysts' accuracy for the manually digitized parts: The following information should be provided:
(2a) mean area values and standard deviation, sample size and a short characterization of the selected glaciers (debris-covered?, in shadow?)
(2b) an overlay of outlines obtained by the multiple digitizations

(2c) for clean ice glaciers (requiring no manual correction): a comparison of the automatically derived area values to the manually digitized ones (differences in percent),
(2d) a discussion of the results.

3. When multiple manual digitising cannot be performed for whatever reason, a buffer of ± a half pixel should be applied to all glacier outlines and a minimum and maximum total glacier area should be calculated from this buffered extent.
(3a) The range of the area differences should be calculated and provided as a relative measure (in percent).
(3b) The uncertainty range should be discussed in view of the glacier characteristics (e.g. are clean-ice glaciers dominant or should a ±1 pixel buffer be applied?) and mapping conditions.
(3c) An overlay of the derived glacier outlines in a region with challenging mapping conditions should be shown and discussed.

4. At an absolute minimum, results from other studies should be considered (e.g. Paul et al., 2013) and the accuracy measures determined in these studies should be applied to the dataset considering the mapping conditions.
(4a) An uncertainty value should be selected from the literature, justified for the current study, and applied to the sample.
(4b) As (3c).

Due to the rapid glacier change with time and the wide range of possible outline interpretations by different analysts, it is not recommended to use datasets from another epoch or digitized by another expert for validation. Using them for change assessment also requires careful evaluation, in particular when the accuracy is not given. Of particular importance is reporting of potential errors that might be much larger than those discussed above: This is required when seasonal snow or clouds/fog are part of the scene and hide glacier perimeters partly. In such cases it must also be described how these challenges were solved (multi-temporal images?, exclusion from the sample?) to derive accurate values.

The above strategy of quantitative accuracy assessment will also be followed by Glaciers_cci for the generated glacier area product. In the case of a accuracy evaluation of an already existing product, the following qualitative measures will be applied:
(i) overlay of the outlines with existing satellite images (shapefiles and geotiff in the GIS or kml files in Google Earth) and analysis of the differences focusing on debris cover, shadow, clouds, water and potential seasonal snow
(ii) consideration of potential rapid glacier change and analysis of existing alternative satellite data to improve the quality
(iii) contact the responsible GLIMS Regional Centre for further steps to be performed.

4. Elevation change (Altimetry)

4.1 Sources of errors and uncertainties

4.1.1 External conditions influencing product accuracy

The accuracy and precision of ICESat elevation data have been thoroughly documented in Shuman et al. (2006). In regard to the *dh/dt* error, it 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. For the slope correction, a DEM is used. The difference between the laser altimetry measured elevations and the DEM elevations at each altimetry footprint location, as determined from bilinear interpolation, is computed. As a consequence, 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 the slope. In addition, the error associated with the bilinear interpolation has to be considered.

4.1.2 Impact of the algorithm used

The selected DEM subtracting method (DS-RT-RepAlt), described in the ATBDv2 (Glaciers_cci, 2013), computes the difference between the laser altimetry measured elevations and the DEM elevations at each altimetry footprint location as determined from bilinear interpolation. Elevation differences are then grouped within the selected area, e.g. a pre-defined grid, and within different time seasons. Linear trends are then fitted to the data to estimate average elevation change rates, using a first order polynomial. Thus, the error introduced by the algorithm itself is due to the assumption that the elevation change is supposed to vary linearly as a function of time.

4.2 Methodology used to determine uncertainties

4.2.1 External validation

The uncertainty associated with the estimated elevation trend can be assessed by means of a validation process based on better reference remote sensing data sets, like airborne lidar measurements, which need to coincide temporally with the satellite altimeter overpasses. Due to this large constraint, a relative comparison of results from different algorithms or different sensors or sources can provide an alternative way to assess the uncertainty.

In both cases, 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 datasets to compare.

When the two different sensors, radar versus laser, are used in the validation process, a number of challenges will occur, mainly due to the fact that the two datasets need to be collected during the same time/period. In fact, the laser altimeter data are temporally sparser compared to radar and in addition they are only available during the cloud free days of the operation periods. Then, laser and radar measurements will not always refer to the same location, but to two locations close to each other. Finally, if the snow surface is dry, laser and radar measure a different surface elevation due to the penetration of the radar beam, whereas in case of a snowfall or rapid densification of the upper snowpack due to liquid water, the laser measurement will be misleading due to its insensitivity to the density and moisture of the snowpack. As a consequence, the RMS value of the difference between radar and laser product can be interpreted as the maximum uncertainty of both instruments.

4.2.2 Quantitative measures for accuracy assessment

The selected DEM subtraction method (DS-RT-RepAlt), as mentioned above, is based on fitting a first order polynomial to the elevation differences obtained using bilinear interpolated DEM elevations at each altimetry footprint location, which are then grouped within the selected area, e.g. a pre-defined grid, and within different time seasons:

$$\Delta h(t) = b_0 + b_1 t \tag{4.1}$$

The best estimation of the coefficients b_0 and b_1 is obtained minimising the sum square error:

$$\min_{b_0,b_1} \sum_{i=1}^{M} \left[\Delta h(\bar{t}_i) - \overline{\Delta h}_i\right]^2 = \min_{b_0,b_1} \sum_{i=1}^{M} \left[b_0 + b_1 \bar{t}_i - \overline{\Delta h}_i\right]^2 \tag{4.2}$$

where M is the total number of operational period, or epoch, $\bar{t}_i$ represents the averaged time of the *i-th* operational period, or epoch, and $\overline{\Delta h}_i(x_c, y_c)$ is the average difference relative to such time in the grid cell centred around (x_c, y_c).

If the cost function $f(b_0, b_1)$ is defined as follow:

$$f(b_0,b_1) = \sum_{i=1}^{M} \left[b_0 + b_1 \bar{t}_i - \overline{\Delta h}_i\right]^2 \tag{4.3}$$

minimising the eq. (5.21) corresponds to solve the following system of equations:

$$\begin{cases} \dfrac{\partial f}{\partial b_0} = 0 \\ \dfrac{\partial f}{\partial b_1} = 0 \end{cases} \tag{4.4}$$

whose solutions are:

$$\hat{b}_0 = \frac{\sum_{i=1}^{M} \overline{\Delta h_i}}{M} - \hat{b}_1 \frac{\sum_{i=1}^{M} \bar{t}_i}{M} \tag{4.5}$$

$$\hat{b}_1 = \frac{\frac{\sum_{i=1}^{M} \bar{t}_i \overline{\Delta h_i}}{M} - \frac{\sum_{i=1}^{M} \bar{t}_i}{M} \frac{\sum_{i=1}^{M} \overline{\Delta h_i}}{M}}{\frac{\sum_{i=1}^{M} \bar{t}_i^{\,2}}{M} - \left(\frac{\sum_{i=1}^{M} \bar{t}_i}{M} \right)^2} \tag{4.6}$$

Once the coefficients of the linear equations are determined, the uncertainty of the estimation is assessed by computing the confidence interval, corresponding to a given significance level (α), of the estimated slope coefficient, which represents the elevation trend. The significance level α represents the probability of the error committed in rejecting the assumption that the true elevation trend falls within the interval around the estimated elevation trend.

In order to understand if $\hat{b}_0$ and $\hat{b}_1$ provide a good estimation of the coefficients b_0 and b_1 of the linear equation (4.1), it is necessary to consider the average difference, $\overline{\Delta h_i}$, and the two estimators $\hat{b}_0$ and $\hat{b}_1$ as random variables. In particular, it is necessary to analyse how the latter distribute around their respectively exact values, b_0 and b_1.

It is demonstrable that (Draper and Smith, 1981):

1) $\mathrm{E}\{\hat{b}_0\} = b_0$, that is $\hat{b}_0$ is an unbiased estimator of b_0;
2) $\mathrm{E}\{\hat{b}_1\} = b_1$, that is $\hat{b}_1$ is an unbiased estimator of b_1;
3) $\mathrm{var}\{\hat{b}_0\} = \frac{\sum_{i=1}^{M} \bar{t}_i^{\,2}}{M \sum_{i=1}^{M} \left(\bar{t}_i - E\{\bar{t}_i\}\right)^2} \sigma^2$, where σ^2 is the variance of $\overline{\Delta h_i}$;
4) $\mathrm{var}\{\hat{b}_1\} = \frac{\sigma^2}{\sum_{i=1}^{M} \left(\bar{t}_i - E\{\bar{t}_i\}\right)^2}$, where σ^2 is the variance of $\overline{\Delta h_i}$.

The method of the confidence interval is introduced in this module for the determination of the error associated with the estimates.

In regard to the angular coefficient of the linear interpolation, the method consists of determining the interval around $\hat{b}_1$ that is likely to contain the true parameter value, i.e. b_1, with a cer-

tain probability (called confidence level). As an equation, the following expression defines the confidence bounds:

$$P\left(\left|\hat{b}_1 - b_1\right| \le \delta\right) = P\left(\hat{b}_1 - \delta \le b_1 \le \hat{b}_1 + \delta\right) = 1 - \alpha \tag{4.7}$$

where δ is half of the interval extent, 1 - α is the confidence level, that is the probability of producing an interval containing b_1, and α is the significance level, that is the probability that the true parameter value, b_1, is outside the confidence interval (or, in other words, it is the probability of the error committed in rejecting the assumption that the true parameter value, b_1, falls within the confidence interval).

As the average differences $\overline{\Delta h}_i$ are normally distributed (random variable), it follows that also $\hat{b}_1$ distributes normally, because it is a linear combination of the random variables $\overline{\Delta h}_i$ (see equation, 4.6). Hence, the random variable

$$Z = \frac{\hat{b}_1 - b_1}{\sqrt{\dfrac{\sigma^2}{\sum_{i=1}^{M}\left(\bar{t}_i - E\{\bar{t}_i\}\right)^2}}} = \frac{\hat{b}_1 - b_1}{\left[\mathrm{var}\left(\hat{b}_1\right)\right]^{\frac{1}{2}}} \tag{4.8}$$

has a normal distribution with mean equal to 0 and variance equal to 1. If the variance σ^2 is known, then it is possible to use the Sheppard tables (Sheppard, 1903; Sheppard, 1939) to determine the confidence interval, that is to determine the value of $z_{1-\alpha/2}$ that satisfies the following expression:

$$\begin{aligned} P\left(-z_{1-\alpha/2} \le Z \le z_{1-\alpha/2}\right) &= P\left(-z_{1-\alpha/2} \le \frac{\hat{b}_1 - b_1}{\left[\mathrm{var}\left(\hat{b}_1\right)\right]^{\frac{1}{2}}} \le z_{1-\alpha/2}\right) = \\ &= P\left(\hat{b}_1 - z_{1-\alpha/2}\left[\mathrm{var}\left(\hat{b}_1\right)\right]^{\frac{1}{2}} \le b_1 \le \hat{b}_1 + z_{1-\alpha/2}\left[\mathrm{var}\left(\hat{b}_1\right)\right]^{\frac{1}{2}}\right) = 1 - \alpha \end{aligned} \tag{4.9}$$

From the above equation, the half extent of the confidence interval is given by:

$$\delta = z_{1-\alpha/2}\left[\mathrm{var}\left(\hat{b}_1\right)\right]^{\frac{1}{2}} = z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{\sum_{i=1}^{M}\left(\bar{t}_i - E\{\bar{t}_i\}\right)^2}} \tag{4.10}$$

In most of the case, the variance σ^2 is unknown and it is often estimated by means of the following expression adapted to the symbology defined and used above:

$$s^2 = \frac{1}{M-2}\sum_{i=1}^{M}\left[\overline{\Delta h}_i - \left(\hat{b}_0 + \hat{b}_1\bar{t}_i\right)\right]^2 \tag{4.11}$$

When s^2 is used to replace the variance σ^2, $\hat{b}_1$ does not present anymore a normal distribution. However, the variable:

$$t = \frac{\hat{b}_1 - b_1}{\sqrt{\dfrac{s^2}{\sum_{i=1}^{M}\left(\bar{t}_i - E\{\bar{t}_i\}\right)^2}}} \tag{4.12}$$

assumes a *t*-Student distribution with $(M-2)$ degrees of freedom (the 2 degrees of freedom subtracted to the number M of couples $\left(\bar{t}_i, \overline{\Delta h}_i\right)$ is due to the number of degrees of freedom on which the estimate s^2 is based).
Indicating with $t_{1-\alpha/2}$ the upper critical value of the *t*-Student distribution with M-2 degrees of freedom which identifies the tail of the distribution with a subtended area equal to the probability of $\alpha/2$, the following expression allows the determination of the half extent of the confidence interval:

$$\begin{aligned} P\left(-t_{1-\alpha/2} \le t \le t_{1-\alpha/2}\right) &= P\left(-t_{1-\alpha/2} \le \frac{\hat{b}_1 - b_1}{\sqrt{\dfrac{s^2}{\sum_{i=1}^{M}\left(\bar{t}_i - E\{\bar{t}_i\}\right)^2}}} \le t_{1-\alpha/2}\right) = \\ &= P\left(\hat{b}_1 - t_{1-\alpha/2}\sqrt{\frac{s^2}{\sum_{i=1}^{M}\left(\bar{t}_i - E\{\bar{t}_i\}\right)^2}} \le b_1 \le \hat{b}_1 + t_{1-\alpha/2}\sqrt{\frac{s^2}{\sum_{i=1}^{M}\left(\bar{t}_i - E\{\bar{t}_i\}\right)^2}}\right) = 1-\alpha \end{aligned} \tag{4.13}$$

which is given by:

$$\delta = t_{1-\alpha/2}\sqrt{\frac{s^2}{\sum_{i=1}^{M}\left(\bar{t}_i - E\{\bar{t}_i\}\right)^2}} = t_{1-\alpha/2}\, s_{\hat{b}_1} \tag{4.14}$$

where

$$s_{\hat{b}_1} = \sqrt{\frac{s^2}{\sum_{i=1}^{M}\left(\bar{t}_i - E\{\bar{t}_i\}\right)^2}} \tag{4.15}$$

denotes the standard error of the estimate.
Summarising, the confidence interval of the true elevation trend, b_1, corresponding to a confidence level of 1-α, can be written as follows:

$$\left[\hat{b}_1 - t_{1-\alpha/2}s_{\hat{b}_1}, \hat{b}_1 + t_{1-\alpha/2}s_{\hat{b}_1}\right] \tag{4.16}$$

In the above equation, $t_{1-\alpha/2}$ is the $100\left(1-\frac{\alpha}{2}\right)$ percentage point of a *t*-Student distribution with $(M-2)$ degree of freedom.

When the confidence interval of the elevation trend is determined, the trend error is set to be the greatest value between the lower and upper bound of the confidence interval, that is:

$$uncert = \max\left(\left|\hat{b}_1 - \left(\hat{b}_1 - t_{1-\alpha/2}s_{\hat{b}_1}\right)\right|, \left|\hat{b}_1 - \left(\hat{b}_1 + t_{1-\alpha/2}s_{\hat{b}_1}\right)\right|\right) \tag{4.17}$$

5. Elevation change (DEM differencing)

5.1 Influences on product accuracy

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. This requires to mask glaciers, (hydro-power) lakes, and also pro-glacial areas that are subject to frequent change. 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.

5.2 Methods for accuracy determination

5.2.1 Validation

The validation involves verification of a satellite DEM using a high accuracy DEM when available (e.g. from laser scanning that is increasingly available for glaciers also in polar regions, or airborne photogrammetry). The process begins with co-registration of the datasets using stable terrain, and then resampling of one dataset to another. 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 (e.g. radar wave penetration into snow/firn within the SRTM DEM). In most cases, however, the DEMs will temporally not be consistent and the analysis can only be made over stable terrain (e.g. 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 as it depends on the site topography and the sensors/procedures used for DEM generation, and therefore cannot be universally standardized such as the co-registration pre-processing step.

Alternate components of the 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, 2012). 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 affect the topographic parameters, especially when incorporating them into the glacier database (i.e. GLIMS).

5.2.2 Uncertainty

Uncertainty assessment 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 validation, except that there may not be one DEM that is clearly more accurate than the others. The two multi-temporal DEMs (no. 1 and no. 2) are first co-registered to remove a potential systematic linear bias (horizontal and vertical) (co-registration vector 2-1). Random errors are estimated through statistics (i.e. RMSE) of the stable terrain. The bias is more complicated to detect as it is dependent upon the acquisition techniques of the DEMs (i.e. radar penetration on glaciers using interferometric DEMs) and the quality/resolution of the multi-temporal datasets. To this avail, a third elevation dataset (no. 3) is then co-registered to both the other two DEMs (co-registration vectors 1-3 and 3-2). Elevation data set no. 3 may be elevation profiles acquired by satellite laser altimetry (i.e. ICESat) or simply another DEM. This process returns three co-registration vectors (2-1, 1-3, 3-2) 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, the un-removed linear vertical bias can be added to the error budget for elevation change. The resulting parameters of this uncertainty assessment are an estimate of the random error and an un-removed linear vertical systematic bias.

5.3 Accuracy to be reported

The accuracy to be reported depends on the input data available for DEM differencing. Three constellations can be discriminated:

(1) Two DEMs/elevation data sets of different times and comparable accuracy (typical case for product generation):
The two elevation data sets are co-registered using stable ground. Together with the co-registration parameters the statistical accuracy for these is provided. As the co-registration uses analytical solutions, namely sinusoidal fits (for horizontal offsets) and linear fits (for vertical offsets), the fit accuracies are directly output as statistical standard errors of the fitted offsets. In addition, the RMSE of the elevation differences for stable ground after co-registration is provided.

(2) Three DEMs/elevation data sets of different times:
In addition to the standard errors of offset parameters for all three possible co-registrations, and the RMSEs of stable ground after co-registration (as of case 1), the sum of offset vectors is computed and reported.

(3) One elevation data set typical for product generation and one of high accuracy (validation case):
In principle, the accuracy is reported as in case 1, with the difference that this accuracy can mainly be attributed to the less accurate DEM and characterizes thus this DEM. For sufficiently temporally consistent elevation data sets (i.e. no significant glacier elevation changes to be expected) on-glacier elevation differences can be treated similarly to off-glacier ones, and the co-registration accuracy be reported as in case 1. Additionally, only stable ground co-registration can be performed and the resulting on-glacier elevation differences analysed and offsets be reported (vertical bias, RMS, standard deviation, trend with elevation and its statistical error).

For all three cases, expert observations on remaining elevation differences after co-registration can be added in a comment field, if assessed by an experienced operator. Such observations include for example higher-order errors such as those from satellite jitter or unsolved orbit deformations, or imperfect co-registrations. The uncertainties given will be valid for the product, despite such higher-order errors. However, associated observations given as comment might help users to better interpret or even improve results together with own data. It should thus be ensured that this type of comment becomes visible at a prominent place or in a prominent form when the product meta-data are displayed. Even, in important cases a production system might include a process that requires acknowledgement of the comment before download. A similar process is in place for download of ASTER data and derived products.

6. Velocity

6.1 Sources of errors and uncertainties

6.1.1 Impact of the algorithm for uncertainty characterisation

As a result of the round-robin experiments, intensity cross-correlation was selected as the algorithm for ice surface velocity estimation using both optical and SAR satellite images. Criteria used for algorithm selection were robustness, reliability and accuracy.

The implementation of the algorithm with regard to the matching window size and the oversampling factor has a direct consequence on the precision of the estimates and on the computational time. The former also depends on the target under observation: For SAR sensors estimates with very large window sizes (e.g. 512x512 pixels) are generally more precise for large struc-tures, but this is not applicable for narrow (e.g. < 500 m) glaciers. For optical sensors, typical window sizes are around 10-30 pixels. But still, the same rule applies that large windows are in principle able to produce better accuracy for large structures. A trade-off between these parameters is therefore necessary in the implementation of the algorithm (Debella-Gilo and Kääb, 2012). Also the post-processing of the matching outcomes (i.e. filter) is a critical processing step. Here a trade-off is necessary in terms of number of estimates versus level of confidence, or between the number of mismatches kept and correct matches lost through the filter process.

The implementation of the cross-correlation algorithm and of the post-processing step have a direct impact on the noise of the resulting estimates. A further element which is also affected the noise is the presence of intensity features to be tracked. This depends in particular on the surface conditions over the glaciers and are discussed in the next section.

For SAR sensors, excluding noise, the image offsets in the slant-range and azimuth directions are related to stereo offsets, the different satellite orbit configurations of the two SAR images, the displacement occurring between the acquisition time interval of the image pair, and ionospheric effects. In order to accurately determine the displacement, all the other terms must be carefully characterised. Orbital and stereo offsets are discussed in this section as part of the internal uncertainty characterisation, ionospheric effects are considered external factors and discussed in the next session.

The orbital offsets are determined by fitting a bilinear polynomial function to offset fields computed globally from the SAR images assuming no displacement for most parts of the image. Residual errors on stable ground might be used to inspect the results against systematic residual offsets. Stereo offsets *DR* are relevant for the range offset-field:

$$DR = z * B / (R * \sin(\theta))$$

where z is the height, B the baseline, R the height of the SAR platforms above the Earth's surface, and θ the incidence angle.

In the case of an ALOS PALSAR image pair with a baseline of 700 m and for a height difference of 700 m:

$$DR = 700 * 700 / (879762.8985*\sin(35)) = 0.97 \text{ m}$$

For an ALOS PALSAR slant-range pixel of 9.37 m, a stereo offset of 0.97 m corresponds to 1/10th of a pixel.

The following Fig. 6.1 shows the slant-range offsets distribution over the Austfonna ice cap as a function of height for a region without outlet glaciers. In order to avoid stereo offsets, the two SAR images can be co-registered with consideration of the topography, a procedure which is becoming standard in particular for the very-high resolution X-band data. Optical velocity products from Glaciers_cci are based on orthorectified nadir images, so that above distortions are negligible. Effects from errors in the DEMs used for orthoprojection are treated below.

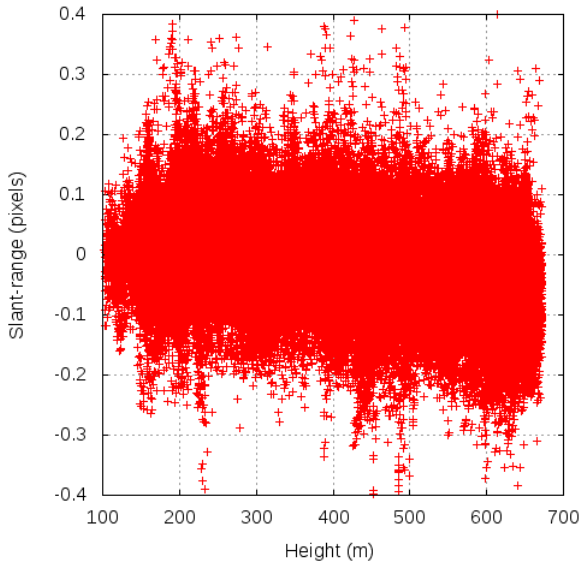

Fig. 6.1: Visualisation of the slant-range offsets distribution over the Austfonna ice cap.

6.1.2 External conditions influencing product uncertainty

Surface conditions over the glaciers have a direct impact on the quality of the matches. In general, the algorithm performs best if there are clear intensity features to be tracked with regard to the size of the employed window. For optical sensors, the presence of clouds or snow cover will mask underlying structures and reduce precision. In the case of SAR sensors, wet snow will drastically change the signal and the employed wavelength has an impact on the penetration depth within ice and snow, allowing in some cases (L-band) the retrieval of information over ice caps where other sensors (X-band) do not provide information.

As a further important remark, valid for optical and SAR sensors, it should be noted that image cross-correlation strictly provides displacements for the time period between the acquisitions used. Thus, the glacier velocity product is the mean velocity over the observation period, and

does not take any velocity variations between the image acquisitions into account. This fact is of particular importance when analysing time series of glacier velocities.

In the case of SAR sensors, ionospheric conditions during the acquisition of the images also have an impact on the accuracy of the results (see Fig. 6.2). The free electron density in the ionosphere varies with the activity of the Sun, the Earth magnetic field and atmospheric parameters, with higher concentrations and stronger spatial variations in Polar regions. The free electrons interact with electromagnetic waves as a dispersive medium, with inverse effects on the phase and group velocities and stronger effects at lower frequencies. Electron density fluctuations result in variations in the interferometric range phase. Furthermore, delay phase ramps across the synthetic aperture cause significant azimuth positional shifts ("azimuth streaking"). ionospheric streaks detected on the azimuth offset maps might be high-pass filtered along the range direction (Wegmüller et al., 2006). It is in any case suggested careful inspection of the azimuth offset field to highlight possible ionospheric effects.

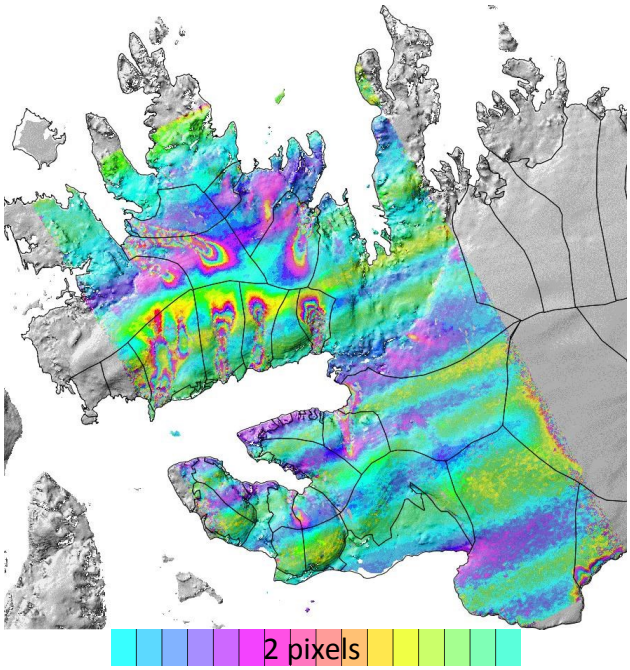

Fig 6.2: Azimuth offset-field over Vestfonna in the case of an ALOS PALSAR image pair. Azimuth streaks show up along with the motion of the outlet glaciers.

The most important external conditions influencing uncertainty for displacements from optical sensors are (more details are given in the PVASR, see Glaciers_cci, 2012b):

- Vertical error components in the DEMs used for orthoprojection will translate in horizontal displacement errors. This effect becomes typically negligible when data from the same orbit are used for tracking, similar to the SAR case. For data from different orbits the effect typi-cally becomes visible in stable ground offsets that thus can serve as estimates for such ef-fects. DEM errors propagated into orthoimages cannot be undone or cancelled else as they are not of analytical nature.

- Higher-order distortions (e.g. jitter) from errors in the provided or modelled attitude angles will lead to an according error pattern in the displacements. Depending on their nature (e.g. for ASTER) and provided that sufficiently many and distributed stable-ground offsets are available, they can be modelled statistically and the glacier displacements corrected accord-ingly (Nuth and Kääb, 2011)
- Ground and illumination conditions (provided that the data are taken under day-light and clear-sky conditions at all) can complicate the matching process. Optical tracking methods rely on corresponding visual contrast in both images that defines offsets in both horizontal dimensions in a unique way. Optical matching problems occur therefore, for instance, under lack of contrast (e.g. fresh snow), self-similar features (e.g. seracs or ogives), or contrast that defines only one offset dimension (e.g. longitudinal moraines or flow stripes without trans-verse or contrast variations).

(Remark: the latter two external factors apply similarly for optical stereo DEMs).

6.2 Methods for accuracy determination

6.2.1 Background

Glacier motion retrieval from space-borne sensors is characterised by some inherent methodological drawbacks, which complicate the validation of glacier displacements from space against independent data with equal or better resolution, accuracy and precision. The main reasons are:

(1) **Coincident observation of EO and validation data**

Glacier motion often follows diurnal to seasonal cycles and year-to-year variations, among others as a consequence of the varying sub-glacial hydrology, therefore glacier motion is highly variable temporally, at scales from hours to seasons and years. A strict validation of glacier velocity products would therefore require simultaneous acquisitions of product generation and validation data, which can only be achieved by continuous ground measurements.

(2) **Adjusting spatial scales**

Glacier displacement measurements from repeat images require image windows to be compared, i.e. the motion of feature ensembles rather than single features is estimated. Therefore, the derived displacement is not representative for a certain finite point, but rather for an area. Further, this representativeness is not a strict analytical function of the real displacement field, but a statistical relation of it, its gradients, image features and contrast, as well as the tracking algorithm and its implementation.

(3) **Observation of different velocity components**

Depending on the sensor (SAR, optical) and applied method, different components of the true 3D velocity are observed. For example, SAR interferometry is sensitive to the Line-Of-Sight (LOS) velocity component of the ice, but it is in-sensitive to along track motion. Image Cross correlation techniques using SAR data measure displacement in LOS and along track direction, but are also sensitive to net elevation changes (e.g. melting, vertical velocity components). In addition, the movement of the ice particles on a glacier does not follow the surface. In general, there is submergence in the accumulation area and emergence in the ablation area. Additionally, in summer, changes of the ice surface motion of glaciers are a combination of ice displace-

ment and surface melt. In the field, ice velocity is measured with stakes at various depths or with continuous GPS on the surface. In order to validate and/or compare the products from various methods, the transformation to the same velocity component is a pre-requisite.

As a summary, accuracy determination 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), for instance 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 are estimated by dividing the displacements through the time period between the acquisitions used. Thus, the product is a mean velocity over the observation period, and does not take into account velocity variations in-between. This fact is of particular importance when analysing time series of glacier velocities.

6.2.2 Validation

In principle, two methods of validation can be applied:

- 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.

6.2.3 Uncertainty

Determination of uncertainty is possible from a large number of methods:

- 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 is 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.

6.3 Accuracy to be reported

Summarizing the above possibilities of validation and uncertainty determination, the accuracy of offsets tracked between two optical or SAR images will be reported as

- A comment from an experienced operator based on visual inspection of the resulting displacements (overall impression, sensor effects, etc.)
- The correlation coefficient or signal-to-noise ratio or both, depending on the algorithm implementation, for each displacement.
- Similarity between original measurements and low-pass filtered ones for each displacement, given as deviations in x- and y-offsets, or vector magnitude and direction.
- Statistical measures for stable ground matches (mean, standard deviation, RMS)

For validation, offsets between two images can in rare cases be compared to temporally consistent matches based on images with higher resolution, or to ground-based measurements of displacements. In these cases, in addition to the above accuracy measures the following numbers will be reported:

- Deviations between product-type displacements and validation displacements for x- and y-offsets or vector magnitude and direction for each location with both results available.
- Summary of deviations (mean, standard deviation, RMS, min, max)

As a very general statement, the accuracy of individual glacier displacement measurements from repeat satellite optical data using offset tracking is on the order of one pixel. For a one year Landsat panchromatic image pair this corresponds to an accuracy of 15 m /year. For SAR sensors, we estimate the reliability of the cross-correlation algorithm to return co-registration parameters as accurate as 1/10th of an image pixel. This corresponds for the ALOS PALSAR and TerraSAR-X data separated by a temporal interval of 46 respectively 11 days to an accuracy of about 10 m/yr and for the ENVISAT ASAR data separated by a temporal interval of 35 days to an accuracy of about 20 m/yr.

8. References

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.

Beers, Y. (1957): Introduction to the theory of error, Addison-Wesley, Massachusetts, 1957.

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

BIPM (2008): Guide to the Expression of Uncertainty in Measurement (GUM), Bureau International des Poids et Mesures. (http://www.bipm.org/en/publications/guides/gum.html)

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., Menounos, B. and Wheate, R.D. (2010): Landsat-based inventory of glaciers in western Canada, 1985 - 2005. Remote Sensing of Environment, 114, 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.

Draper, N. R. and Smith, H. (1981). Applied Regression Analysis, John Wiley & Sons, Inc.

Frey, H. and Paul, F. (2012): 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, 18, 480-490.

Frey, H., F. Paul and T. Strozzi (2012): Compilation of a glacier inventory for the western Himalayas from satellite data: Methods, challenges and results. Remote Sensing of Environment, 124, 832-843.

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., Mathieu, R., Kääb, A., Chinn, T., Fitzharris, B. and Hagen, J.O. (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): Product Validation and Algorithm Selection Report (PVASR): Prepared by the Glaciers_cci consortium, 113 pp.

Glaciers_cci (2013): Algorithm Theoretical Basis Document Version 2 (ATBDv2). Prepared by the Glaciers_cci consortium, 81 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.

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.

Hughes, I.G., and T.P.A. Hase, Measurements and their uncertainties, Oxford University Press, Oxford, 2010.

Kääb, A. (2005): Combination of SRTM3 and repeat ASTER data for deriving alpine glacier flow velocities in the Bhutan Himalaya. Remote Sensing of Environment, 94, 463-474.

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

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.

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

Paul, F. (2008): Calculation of glacier elevation changes with SRTM: Is there an elevation dependent bias? Journal of Glaciology, 55 (188), 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 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. (2011): 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.

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., Paul, F., Raup, B., Khalsa, S.J.S. and Armstrong, R. (2009): Challenges and recommendations in mapping of glacier parameters from space: results of the 2008 Global Land Ice Measurements from Space (GLIMS) workshop, Boulder, Colorado, USA. Annals of Glaciology, 50, (53), 53-69.

Rastner, P., T. Bolch, N. Mölg, H. Machguth, R. Le Bris and F. Paul (2012): The first complete inventory of the local glaciers and ice caps on Greenland. The Cryosphere, 6, 1483-1495.

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.

Sheppard, W. F. (1903), New Tables of the Probability Integral. Biometrika, 2, 174-190 (available at http://biomet.oxfordjournals.org/content/2/2/174.full.pdf).

Sheppard, W. F.(1939), The Probability Integral. Cambridge, University Press.

Shuman, C. A., Zwally, H. J., Schutz, B. E., Brenner, A. C., DiMarzio, J. P., Suchdeo, V. P., and Fricker, H. A. (2006). ICESat Antarctic elevation data: Preliminary precision and accuracy assessment. Geophysical Research Letters, 33, L07501.

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.

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

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.

Abbreviations

ASTER	Advanced Spaceborne Thermal Emission and Reflection radiometer
CC	Cross-Correlation
CCI	Climate Change Initiative
DARD	Data Access Requirements Document
DEM	Digital Elevation Model
ESA	European Space Agency
GCM	General Circulation Model / Global Climate Model
GCOS	Global Climate Observing System
GDEM	Global DEM (from ASTER)
GLIMS	Global Land Ice Measurements from Space
ICESat	Ice, Cloud, and Elevation Satellite
L1T	Level 1 T (terrain corrected)
NASA	National Aeronautic and Space Administration
NDSI	Normalized Difference Snow Index
PALSAR	Phased Array type L-band SAR
RMSE	Root Mean Square Error
SAR	Synthetic Aperture Radar
SPOT	System Pour l'Observation de la Terre
SRTM	Shuttle Radar Topography Mission
TM	Thematic Mapper
USGS	United States Geological Survey