Article:

Paul, F, Bolch, T, Briggs, K et al. (8 more authors) (2017) Error sources and guidelines for quality assessment of glacier area, elevation change, and velocity products derived from satellite data in the Glaciers_cci project. Remote Sensing of Environment, 203. pp. 256-275. ISSN 0034-4257

https://doi.org/10.1016/j.rse.2017.08.038

Error sources and guidelines for quality assessment of glacier area, elevation change, and velocity products derived from satellite data in the Glaciers_cci project

Frank Paul[1]*, Tobias Bolch*[1]*, Kate Briggs*[2]*, Andreas K*ää*b*[3]*, Malcolm McMillan*[2]*, Robert McNabb*[3]*, Thomas Nagler*[4]*, Christopher Nuth*[3]*, Philipp Rastner*[1]*, Tazio Strozzi*[5]*, Jan Wuite*[4]

[1] Department of Geography, University of Zurich, Winterthurerstr. 190, 8057 Zurich, Switzerland

[2] Centre for Polar Observation and Modelling, University of Leeds, Leeds, LS2 9JT, UK

[3] Institute of Geosciences, University of Oslo, P.O. box 1047, 0316 Oslo, Norway

[4] ENVEO IT GmbH, Technikerstrasse 21a, 6020 Innsbruck, Austria

[5] GAMMA Remote Sensing, Worbstr. 225, 3073 Gümligen, Switzerland

Abstract

Satellite data provide a large range of information on glacier dynamics and changes. Results are often reported, provided and used without consideration of measurement accuracy (difference to a true value) and precision (variability of independent assessments). Whereas accuracy might be difficult to determine due to the limited availability of appropriate reference data and the complimentary nature of satellite measurements, precision can be obtained from a large range of measures with a variable effort for determination. This study provides a systematic overview on the factors influencing accuracy and precision of glacier area, elevation change (from altimetry and DEM differencing), and velocity products derived from satellite data, along with measures for calculating them. A tiered list of recommendations is provided (sorted for effort from Level 0 to 3) as a guide for analysts to apply what is possible given the datasets used and available to them. The more simple measures to describe product quality (Levels 0 and 1) can often easily be applied and should thus always be reported. Medium efforts (Level 2) require additional work but provide a more realistic assessment of product precision. Real accuracy assessment (Level 3) requires independent and coincidently acquired reference data with high accuracy. However, these are rarely available and their transformation into an unbiased source of information is challenging. This overview is based on the experiences and

lessons learned in the ESA project Glaciers_cci rather than a review of the literature.

1. Introduction

The wide range of freely available satellite data (e.g. Pope et al., 2014) allows deriving numerous glacier-related products (Malenovsky et al., 2012) using, in most cases, well-established algorithms (Paul et al., 2015). These products (e.g., glacier outlines, flow velocities, volume changes, snow facies, surface topography) provide baseline information about glacier distribution (inventories) and changes in length, area and volume/mass, thus informing about the state of the cryosphere, regional trends of water resources, glacier dynamics and impacts of climate change (e.g. Vaughan et al., 2013).

In general, the satellite-derived products are complimentary to ground measurements that provide information on glacier fluctuations (length and mass) only for a small sample (about 1000) of the estimated 200 000 glaciers (Pfeffer et al., 2014), albeit for a much longer period (centuries) and so far at a higher temporal resolution (Zemp et al., 2015). The main asset of satellite data is to obtain a regionally more complete picture of glacier changes and the spatio-temporal extension of the information available from the ground network. The project Glaciers_cci is one of several projects from the ESA climate change initiative (CCI) that is analysing the Essential Climate Variable (ECV) ‘Glaciers’ using a suite of satellite data (Hollmann et al. 2013). Table 1 provides an overview on the three main products (glacier outlines, elevation changes, flow velocity) generated in Glaciers_cci along with some general characteristics of their determination.

Their digital combination and joint assessment, for example to determine the contribution of glaciers to global sea level rise, requires a large computational effort and several assumptions for unmeasured regions (Gardner et al., 2013). We do not discuss here the uncertainties related to such combined datasets or follow-up applications, e.g. a missing temporal match of glacier outlines and elevation change data. However, all measurements have uncertainties and these need to be available for error propagation. Unfortunately, they are not always reported and the reliability of a dataset is thus difficult to assess. Moreover, uncertainties might be locally variable and different (sometimes incomparable) measures have been used in the literature. In part this is due to the complimentary nature of field-based measurements, which is limiting their use as reference data for validation, as location, sampling interval and cell-size (point data versus averages per grid cell) might not match.

Table 1: Satellite-derived glacier products (EC-ALT/DEM: elevation change from altimetry / DEM differencing), typical freely available sensors or datasets, auxiliary datasets (GO: glacier outlines, DEM: digital elevation model) and their purpose, processing methods and output format.

Product	Input	Sensors or Datasets	Auxiliary Datasets	Purpose of Auxiliary data	Processing	Output
Outlines	Optical image	Landsat, Sentinel 2, ASTER, SPOT	DEM, high-res. optical	Divides, topographic parameters	Ratio image with threshold	Vector (polygon)
EC-ALT	Laser altimeter	ICESat	GO, DEM	Mask, slope	Filtering and differences	Vector (point)
	Radar altimeter	Cryosat 2	GO	Mask		Vector (point)
EC-DEM	Optical DEM	GDEM, SPIRIT	GO	Mask	Co-registration & subtraction	Raster
	Radar DEM	SRTM C/X, TanDEM-X	GO	Mask		Raster
Velocity	Optical image	Landsat, Sentinel 2, ASTER	GO	Mask	Offset-tracking	Vector (point)
	Radar image	Palsar, Sentinel 1, TerraSAR-X	GO, DEM	Mask, geocoding, flow conversion	Offset-tracking (InSAR)	Vector (point)

In the following, we use the term accuracy (error) as a measure of the difference between a true value (obtained from independent reference data) and the measured value, or its mean in case several measurements are available. In the latter case the term trueness (representing the systematic error) would be more correct (Menditto et al., 2007). The resulting difference is named bias and in general corrected by subtraction from all measurements. In the absence of reference data, the accuracy of a measurement cannot be determined. However, several measures exist where the deviation from zero is tested (e.g. flow velocities off glaciers) or two similar datasets are compared (e.g. elevation differences over stable ground). The related deviations from zero are also named bias and are in general corrected. The term precision (uncertainty), on the other hand, is representing the variability of measurements around a mean value (also known as random error). Assuming the individual measurements are independent, this variability has a normal distribution characterized by its mean value (to be used for accuracy or bias assessment) and its standard deviation (STD) is representing its precision (Menditto et al., 2007). Some background regarding error propagation can be found in

Merchant et al. (2017).

A key issue when deriving changes or trends from a series of measurements is knowledge about its significance, i.e. whether the change is larger than the precision of the derived product (assuming a potentially detected error or bias is corrected). For glacier outlines, the determination of accuracy is challenged by suitable reference data, as these have to be obtained (weather not interfering) at about the same time (within a week) from a sensor of higher accuracy. It is widely assumed that the latter is fulfilled when its spatial resolution is higher, but this is not generally correct, for example due to sometimes missing image contrast in high-resolution pan-chromatic images (Paul et al., 2013). On the other hand, several internal methods are available for determination of precision and accordingly different measures for uncertainty assessment of glacier products are proposed in the literature and are more or less frequently applied in the respective studies. In contrast to glacier outlines, the elevation change and velocity products are already based on at least two independent input datasets or multiple measurements taken at different times. This allows their direct comparison and a first estimate of bias and uncertainties in regions that should not have changed (so-called stable terrain). In general, neither of the two datasets is 'perfect' (i.e. can serve as a reference for the other) and the derived differences are thus a relative rather than an absolute accuracy measure (i.e. providing bias). Table 2 gives an overview on the initial problems, typical post-processing issues and possibilities of correcting them for the products listed in Table 1.

Table 2: Overview of initial problems, resulting issues for post-processing, methods of editing and some internal accuracy measures for the four products.

Product	Initial problems	Post-processing issues	Editing	Internal accuracy
Outlines	Clouds, seasonal snow, debris, water, shadow	Corrections by the analyst	Manual (on-screen) digitizing	Buffer method, multiple digitization
EC-ALT	Clouds (optical), footprint size, sampling	Terrain slope and roughness, radar penetration	Statistical filtering, bias corrections	Model fit accuracy
EC-DEM	Co-registration, data voids	Outliers, radar penetration, effects of DEM resolution	Outlier filtering, void filling, interpolation	Difference over stable ground
Velocity	Lack of contrast, wet snow / ice, ionospehric effects, radar shadow	DEM errors, data voids, outliers	Outlier filtering, multi-temporal data merging	Correlation coefficient, stable ground velocity

Besides these direct impacts on product accuracy and precision, there are also indirect influences. They are related to auxiliary datasets used for processing (e.g. the quality of the DEM used for orthorectification) and sensor specific ones (e.g. differences in spatial resolution) that impact differently on the generated products. Product specific differences can be found for the (frequency-dependent) radar penetration into snow and ice: whereas they must be carefully considered when deriving elevation changes from at least one SAR component, they are neglected when computing flow velocities as these are assumed to be very similar at the surface and the penetration depth.

Whereas most of the methods provide quantitative information that can be included in the product meta-data, there is a wide range of (external) factors influencing product accuracy that can only be determined in a qualitative sense. These can be related to differences in the interpretation of a glacier as an entity, such as the consideration of steep accumulation areas, attached snow fields, dead ice and rock glaciers, or location of drainage divides derived from different DEMs (Bhambri and Bolch, 2009; Le Bris et al., 2011; Pfeffer et al., 2014; Nagai et al., 2016). Further issues are handling of clouds in glacier mapping from optical sensors, consideration of ionospheric effects for velocity from SAR sensors (Strozzi et al., 2008; Nagler et al., 2015), and handling of data voids or artefacts in DEMs used to calculate elevation changes (Kääb, 2008; Le Bris and Paul, 2015; Wang and Kääb, 2015).

We provide a systematic overview on the determination of product accuracy and precision for each of the four products (A) glacier area (outlines), elevation changes from (B) altimetry and (C) DEM differencing, and (D) velocity from space borne optical sensors and Synthetic Aperture Radar (SAR) using offset tracking (see Tables 1 and 2). For each product we shortly summarize the processing lines before potential error sources and methods of their determination are presented. For all products we close with a tiered list of recommendations that is sorted for workload and data availability. Selected examples illustrate how the different measures vary for the same dataset.

2. Glacier outlines

2.1 Processing line

Glacier outlines are mostly derived from automated classification of optical satellite images (10-30 m spatial resolution) using pixel or object-based classification. This step is followed by manual editing to correct misclassification in regions with water, debris-cover, shadow, and clouds (e.g. Racoviteanu et al., 2009). The automated mapping utilizes the very low reflectance of ice and snow in the shortwave-infrared (SWIR) compared to the visible (VIS) or near infrared (NIR). A threshold applied to the related band ratio (e.g. red/SWIR) already provides a very accurate (pixel sharp) map of 'clean' ice (e.g. Hall et al., 1988; Paul et al., 2002). The scene-specific selection of a threshold value is an optimization process where lower values include more ice in shadow, but at the same time the mapping of bare rock in shadow creates more noise. In most regions this balance is leading to a clearly deined threshold value (Paul et al., 2015). For noise reduction, a median or more correctly majority filter (3 by 3 kernel) is often applied to the classified glacier map. This filter is very effective in removing isolated pixels and filling small gaps with limited changes of the glacier outline.

Unfortunately, most glaciers are not 'clean' but covered to a variable degree by debris so that - depending on its percentage of coverage per image pixel - the ice underneath can either be mapped or not. To some extent this also applies to clouds that can be sufficiently thin (cirrus, fog) to map the glaciers underneath. Ice and snow in shadow are normally precisely mapped (e.g. Paul et al. 2016), but due to atmospheric conditions and/or low solar elevation (creating deep shadows), the method can also fail. There are workarounds such as using the green or blue band instead of the red or NIR for the band ratio, but these have other shortcomings (e.g. they map all water as glaciers). Hence, visual control of all glacier outlines and related manual corrections are required for creating accurate glacier outlines. Alternatively or when a SWIR band is not available (such as for panchromatic imagery from very high-resolution sensors or aerial photography), complete manual digitization can or has to be applied. The main goal of the editing is always to create complete outlines as – in contrast to the widely accepted data voids in elevation change and velocity products – incomplete outlines are not accepted. This creates special challenges and often requires implementing workarounds. Accordingly, the list of issues described in the following for glacier outlines is longer than for the other products.

2.2 Factors influencing product accuracy

2.2.1 Scene conditions and interpretation rules

Selection of the best scene for glacier mapping is also an optimization process. One has to balance between cloud cover, snow conditions and shadowing. For example, late in autumn cloud and snow conditions are better but shadows are getting increasingly large, hiding glaciers. More seasonal snow (hiding the glacier perimeter) makes the mapping increasingly vague and result in an overestimation of glacier area. Depending on the region, it might be possible to overcome the cloud problem by combining scenes from a different date where clouds might have different locations (Fig. 1). For remaining clouds in the accumulation area time is not critical as changes in this region are generally small. This allows using either scenes from other years or copying the outlines from an already existing dataset such as the Randolph Glacier Inventory (RGI; Pfeffer et al., 2014).

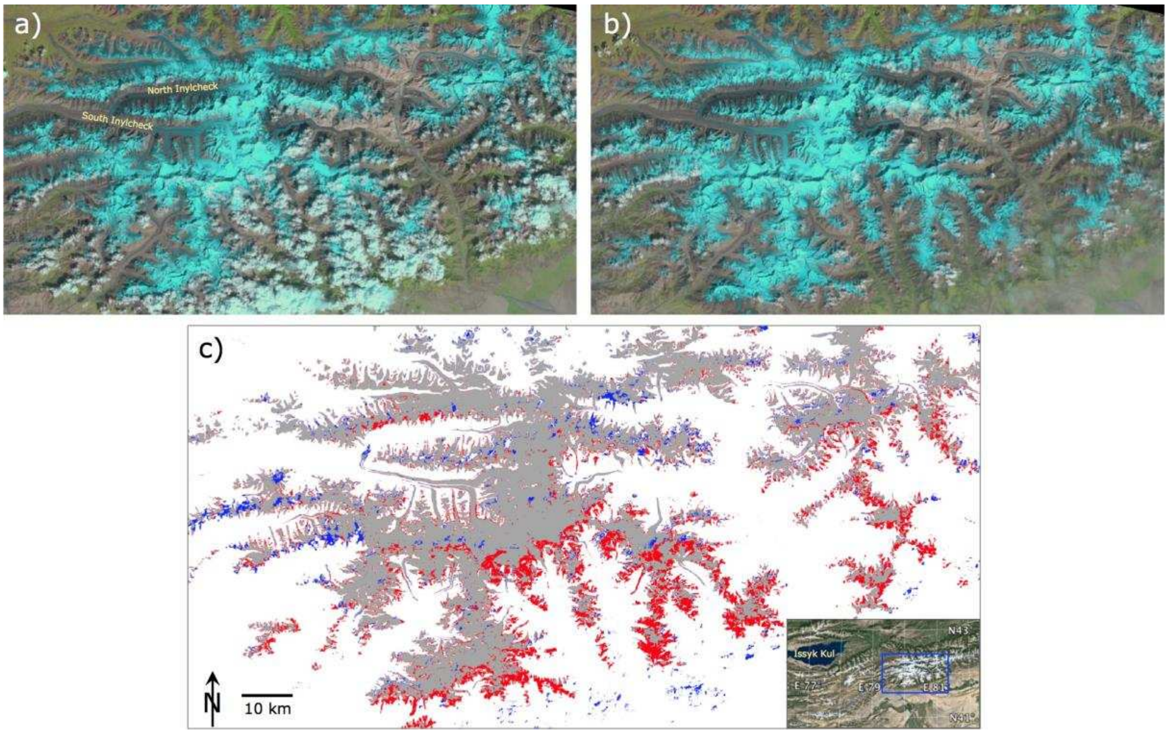

Fig. 1: The two false colour Landsat images (path-row: 147-031) in the top row cover the region around North and South Inylcheck Glacier in the central Tien Shan (see blue square in inset map for

location) and show clouds (white) at different locations (ice and snow in shades of blue-green). They were acquired on a) 21.08.2006 and b) 24.08.2007. c) The digital combination of the classified glacier maps (2006: grey/blue, 2007: grey/red) allows creating a near complete glacier coverage. Inset map: screen shot from Google Earth, Landsat images: USGS/NASA.

Seasonal snow is also a very critical factor that can only be resolved by using the best scenes for glacier mapping (even if clouds are present). Methods for exploiting time-stacks of satellite images to synthesize optimal mapping conditions have also been proposed, though (Winsvold et al., 2016). Seasonal snow is a particular problem in maritime regions, the tropics, and very high mountain ranges and one might have to wait several years before an appropriate scene is available (Paul et al., 2011). Whereas some seasonal snow can be identified from its irregular shape and removed during manual editing, this is challenging for larger regions and might not always work (Fig. 2). Moreover, it is often nearly impossible to differentiate between seasonal and perennial snow, even at high spatial resolution. Including the latter in a glacier inventory or not is also a matter of the interpretation rules.

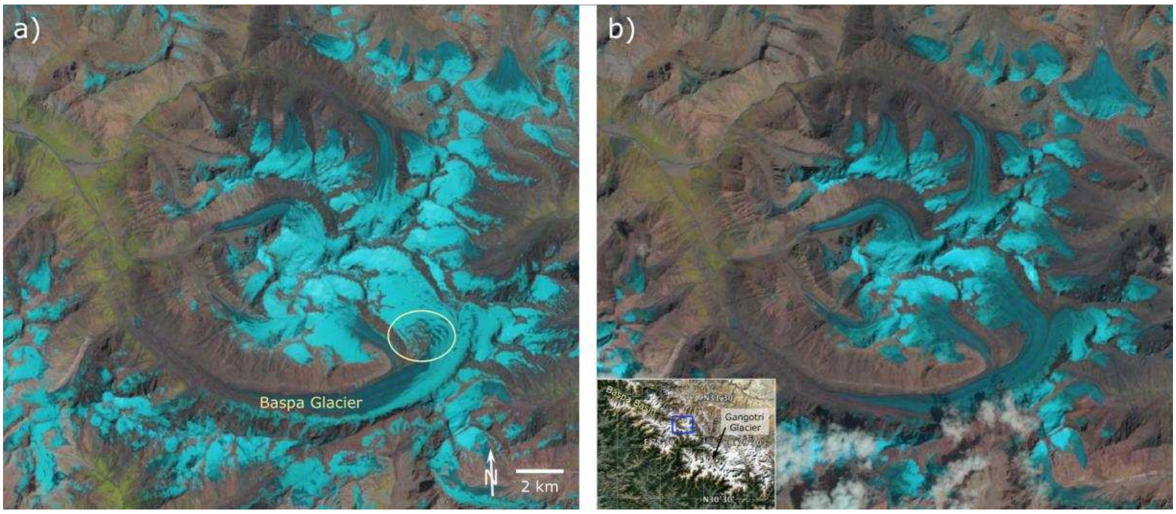

Fig. 2: The region around Baspa Glacier at the headwater of the Baspa river basin (see blue square in inset map for location) as seen on two false colour Landsat images (path-row: 146-038) acquired on a) 20. Aug. 2014 and b) 10. Sep. 2016. Although a) looks usable for glacier mapping at first sight, it suffers from abundant seasonal snow (circle) and avalanche cones hiding glacier parameters. In b) snow outside of glaciers has largely disappeared and glacier mapping is much more easy. However,

some clouds are now hiding some of the glaciers and need to be mapped by other scenes (see Fig. 1). Inset map: screen shot from Google Earth, Landsat images: USGS/NASA.

Similarly, what belongs to a glacier might be defined differently. Although a long list of rules has been defined by the Global Land Ice Measurements from Space (GLIMS) initiative (Raup and Khalsa 2007) to achieve some consistency in interpretation, other definitions have been applied and challenges remain. For example, Nuimura et al. (2015) have neglected ice at steep slopes and distinguishing debris-covered glaciers from rock glaciers or ice-cored moraines (only visible in very high resolution images) is a key challenge in cold and dry high-mountain environments from both remote sensing and field surveys (e.g. Berthling, 2011; Frey et al., 2012; Janke et al., 2015; Østrem, 1971). Figure 3 is illustrating the complexity of periglacial landforms with two examples, showing also the difficulties in identifying a clear glacier outline. Hence, glacier area differences might be large without outlines being wrong and related change assessment with datasets created by other analysts requires some caution (Nagai et al. 2016).

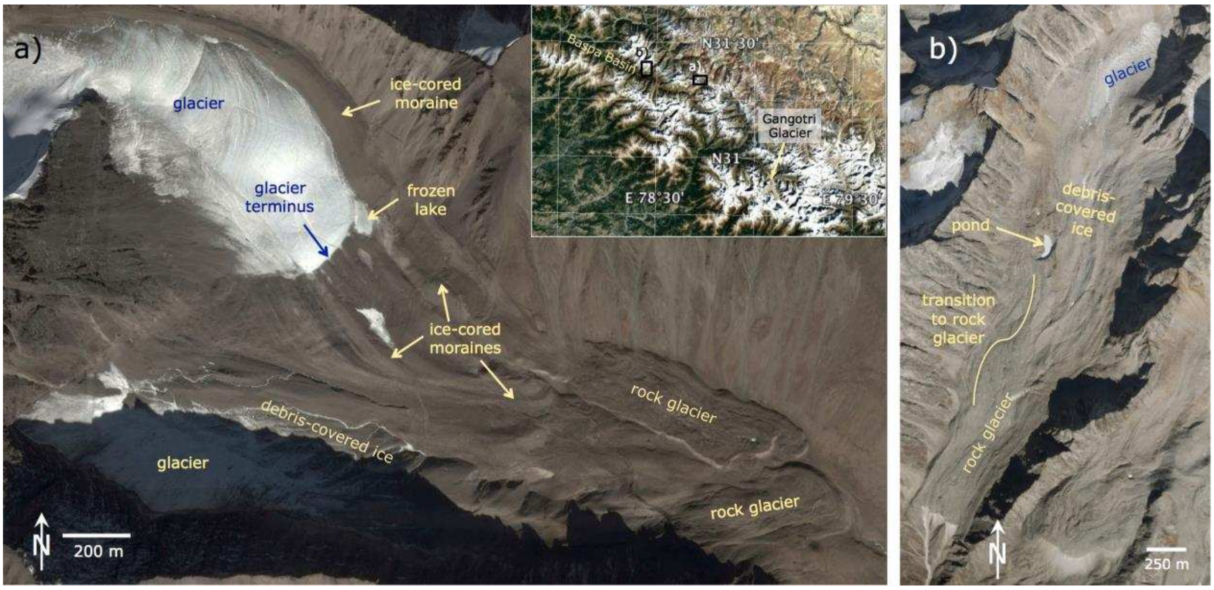

Fig. 3: a) Glaciers, debris-covered ice, rock glaciers, ice-cored moraines and other periglacial features in a small catchment of the Baspa basin (see inset for location). In this region the glacier terminus is clearly defined, but the other marked periglacial landforms containing ice are based on subjective interpretation. b) A small cirque glacier (upper right) that continuously evolves into a

debris-covered glacier and a rock glacier with its steep front in the lower left (there is a further rock glacier to the right). In this case several possibilities exist to assign a glacier terminus (indicated by the transition zone). Images and inset map: Screen shots from Google Earth, (C) 2017 CNES / Airbus.

2.2.2 Sensor characteristics: Spatial / spectral resolution and Landsat 7 striping

Characteristics of the source data (spatial resolution, spectral range, ETM+ striping) also impact on the quality of the resulting glacier outlines. As the boundary of real glaciers is curved rather than rectangular, any resampling of the original outline into a grid with a spatial resolution coarser than about 1 cm (typical size of ice grains), results in a generalization and thus in a change of the true area. The related change of area with pixel size was analysed in a theoretical experiment by Paul et al. (2003) for grid cell sizes of common satellite sensors (e.g., 5, 10, 15, 20, 30 m). Whereas this study did not found a systematic trend of area differences with glacier size, the standard deviation of the area differences strongly increased towards smaller glaciers.

On the downside of a higher spatial resolution is automated mapping. As glaciers are often slightly dirty along their perimeter and/or are covered by narrow medial moraines, mapping them with a higher spatial resolution will exclude these features, as the percentage of coverage with non-ice information within a 10 m pixel is higher. A corresponding 30 m pixel (covering nine 10 m pixels) might still be mapped as (clean) glacier ice if more than half of its area is ice. This results in somewhat larger glacier extents being mapped by lower resolution sensors. For example, 5% larger extents were mapped with Landsat OLI 30 m bands compared to 10 m Sentinel 2 MSI bands (Paul et al., 2016). The resulting higher workload for manual corrections has to be considered before working at the higher spatial resolution (this requires resampling of the Sentinel 2 / Landsat 8 SWIR bands from 20 to 10 / 30 to 15 m). On the positive side: The higher resolution considerably improves the visibility of debris-covered glacier parts, resulting in a more accurate outline after manual editing, at least when image contrast is sufficient. In the case of panchromatic imagery a reduced contrast between dirty ice and bare rock might also cause problems in identifying the boundary.

The spectral range of a sensor is important, as automated mapping cannot be applied without a SWIR band (often the case for aerial photography or high-resolution sensors). The required manual digitization is prone to subjective interpretation, generalization and reduced consistency. This has in particular to be taken into account for the manual delineation of debris-covered glacier parts, as their correct interpretation is even more challenging (Fig. 3b). To reduce the regions requiring manual intervention we recommended using automated mapping first and then focus on the remaining manual editing.

The striping of Landsat 7 ETM+ scenes that is present since 2003 due to a failure of the scan-line corrector (SLC-off scenes) causes data loss and is difficult to overcome. Whereas it might be possible to add missing parts of the outline by hand without introducing too high errors, this becomes increasingly difficult towards smaller glaciers and wider stripes near the image boundaries. As the stripes are in general at different places in other scenes, it might be possible to overcome the data loss by mosaicking scenes from different dates as for partial cloud cover (e.g. Rastner et al., 2012). However, users will always prefer glacier outlines from one date over multi-temporal composites.

2.2.3 Auxiliary data: DEMs and projection

The use of out-dated and coarse resolution DEMs (90 m) to orthorectify current satellite scenes with 10 or 15 m spatial resolution in steep, high-mountain topography with rapidly changing glacier surfaces introduces deformations and geo-location errors of the true (ortho-projected) glacier shape (Kääb et al., 2016). Whereas the impact of shape deformations on glacier area is likely small (<1%), geo-location errors have no direct impact on glacier area. However, they challenge the combination with other geocoded datasets (see below) and make ground-based validation nearly impossible. Accordingly, geo-location errors should be included in the error budget when different geocoded datasets are digitally combined (e.g. to calculate length changes). Hall et al. (2003) presented a detailed study on related uncertainties. As geolocation errors are sometimes considered when calculating glacier area uncertainties, we include them here for completeness.

Uncertainty in glacier area is also introduced when separating glacier complexes with DEM-derived drainage divides into individual glaciers, as the location of the divide defines the glacier area. However, the total area of the glacier complex (all originally connected glaciers) remains the same and is not affected by the positional uncertainty. At mountain crests, a shift of the drainage divides by 2 or 3 image pixels can easily introduce hundreds of sliver polygons that have to be assigned back to the glacier they belong to (e.g. Kienholz et al. 2013). This is tedious work when it has to be done repeatedly for large samples of glaciers, e.g. over entire mountain ranges. Without this correction, geolocation errors cause indeed errors in the derived glacier areas.

Scenes from Landsat and Sentinel 2 are provided in UTM projection with WGS1984 datum. For a scene-by-scene processing and later merging across different UTM zones, the formerly rectangular outlines are slightly rotated. This has an impact on visual appearance and on glacier area for ±1 UTM zone. If ±2 zones are merged, glacier area changes already by a few per cent, as UTM is conservative for angles rather than area. We thus recommend processing all scenes in their respective UTM zones or merge all scenes using a metric equal-area projection (e.g. Rastner et al., 2012).

2.2.4 Algorithm application

Algorithm intercomparison experiments (e.g. Paul et al., 2015; Raup et al., 2014) revealed that the method applied to map glaciers (clean ice and snow) causes only minor differences in glacier area. From simple band ratios to the NDSI (normalized difference snow index) using raw DNs or top of atmosphere reflectance, the outlines are generally on top of each other and deviations are only visible at the level of individual pixels. The only region where results slightly differ is for partial debris cover and ice in shadow, as the manually selected threshold value is most sensitive here (see Paul et al., 2015). As debris has to be manually corrected anyway, it is recommended to select a threshold that is optimized for best mapping results in shadow. This might require using an additional threshold on a band in the blue (or green) part of the spectrum, as the contrast between ice/snow and bare rock in shadow is often higher here (e.g. Raup et al., 2007). In some regions bare rock in shadow can be very bright due to surrounding snow in sunlight creating diffuse scattering (e.g. nunataks in an ice field). In this case it might be difficult to include dark ice in shadow and at the same time exclude bright rock in shadow. A solution for this is the application of two different thresholds and later merging of the results. This also worked when thin clouds or fog require two thresholds (e.g. Le Bris et al., 2011).

The band combination selected for glacier mapping also impacts on misclassification. For example, red/SWIR ratios include larger areas of wrongly mapped lakes compared to NIR/SWIR whereas the latter might include vegetation in shadow. Regions with water and vegetation can partly be excluded by using additional methods in the processing line (e.g. NDVI/NDWI), but parts might remain for removal in the post-processing stage. More difficult can be the detection and removal of surfaces covered by ice (lakes, sea ice, ice bergs) that are correctly classified as ice but are obviously not glaciers. Accurate removal of these ice features from the glacier map requires careful checking with the original (contrast-enhanced) satellite image in the background and some experience (or a previous inventory). Vice versa, lakes on a glacier might be excluded by the mapping, but need to be included again. Object-based classification can be used to identify these context-related differences automatically and correct the result accordingly (e.g. Rastner et al., 2014).

A further impact on glacier size during glacier mapping is introduced by applying a majority filter to the binary glacier map for noise removal. Whereas this filter is very effective in reducing noise by eliminating isolated (snow) pixels and closing gaps in shadow or debris cover (e.g. Paul et al., 2003), the filter also impacts on the extent of small glaciers. If they are elongated and only comprise a few pixels, they might even be completely deleted by the filter. It has thus to be carefully evaluated by the analyst if the application of such a filter is a good idea or not. If snow conditions are poor (many isolated snow fields) and glaciers are comparably large, applying such a filter is recommended.

2.2.5 Post-processing and editing

Post-processing is required to remove and correct obvious misclassification (debris, clouds, scan-line gaps, water surfaces, ice bergs, etc.) and create a high-quality glacier map that can be used for change assessment. One can distinguish two levels of corrections, the easier ones that have to be removed (e.g. lakes, rivers, sea ice, clouds) and the more complex ones that have to be added or re-digitized (debris, shadow, calving termini). In particular debris cover is prone to differences in interpretation (Fig. 3) resulting in potentially large area differences (Paul et al., 2013). These can reach 50% of the total area or even more and have to be corrected to obtain product accuracy better than 5% (according to GCOS 2006). In average, the maybe 10 to 20% uncertainty in the derived area for debris-covered glaciers has to be considered when at another place the correction of individual pixels is discussed.

Moreover, the separation from rock glaciers and other periglacial features is difficult (e.g. Janke et al., 2015) even when using very high-resolution images (Fig. 3). Different opinions exist on their inclusion or exclusion in glacier inventories (e.g., Bown et al., 2008; Frey et al., 2012), but at least they should be marked in the attribute table to easily exclude them from change assessment. Their response to temperature increase is different and they can basically only advance or down-waste at their current extent (Müller et al., 2016). We recommend using coherence images from SAR data (Atwood et al. 2010, Frey et al. 2012), high-resolution images in Google Earth (or from Sentinel 2), and former glacier inventories to guide decisions on boundaries of debris-covered glaciers. For consistency with previous inventories it might be required to include attached perennial ice and snowfields (Lambrecht and Kuhn, 2007; Paul et al., 2011) but mapped glacier extents will be too large then. Along with ice-covered steep mountain flanks that might be included or not, glacier extents including perennial snow fields can easily be 30% larger or smaller. Hence, the dominant sources of uncertainty and error for glacier outlines are clouds, seasonal snow, debris cover and shadow.

2.3 Determination of accuracy and precision

From the two methods applied to generate glacier outlines (automated / manual) and the different error sources influencing accuracy and precision, it is clear that different measures are required to determine them. These include qualitative (e.g. overlay of outline) as well as quantitative (e.g. mean difference and standard deviation) measures. A third group is uncertainty that can only be described but not assessed and needs to be provided as meta-information (e.g. the definition of a glacier and handling of attached snow fields). Unfortunately, missing reference data often hampers real product validation. For example, the sometimes used higher-resolution datasets can have different snow, cloud or shadow conditions when they are not acquired at roughly about the same time, the required manual delineation has uncertainties in its own, and the generally missing SWIR band leads to a different interpretation of the images (e.g. Paul et al., 2013). Other issues of high-resolution satellite data are their limited spatial coverage, high-costs and problems in getting an accurately orthorectified product from the comparably coarse resolution DEMs. In consequence, reference datasets are often used for cross-comparison rather than validation. Table 3 is providing an overview on the different measures to determine precision and accuracy of glacier outlines. They are discussed in the following sections in more detail.

Table 3: Overview of the measures to determine accuracy and precision of glacier outlines (GO). The level refers to section 3.3. GO-4 is only listed for completeness but it is not a measure of accuracy. All differences and standard deviations should be calculated in relation to the total area.

Nr.	Name	Level	Application	Measures	Section
GO-1	Outline overlay	L0	Manual editing, cross-comparison, interpretation differences, visualisation	Descriptive text	2.3.1
GO-2	Literature value	L0	Assume accuracy will be as good	Per cent	2.3.2
GO-3	Buffer method	L1	Buffer outline by 1/2 or 1 pixel, calculate min and max area, assume normal distribution	STD	2.3.2
GO-4	Geolocation	n/a	RMS error of satellite orthorectification	STD	2.3.2
GO-5	Shape deformation	n/a	Pixel shift due to DEM errors (area difference)	Mean	2.3.3
GO-6	Multiple digitizing	L2	Determine analysts precision (area variability)	Mean, STD	2.3.3
GO-7	Area difference	L3	Use of HR reference data for accuracy	Mean (STD)	2.3.4
GO-8	Outline distance	L3	Horizontal distance to HR reference data	Mean, STD	2.3.4
GO-9	Field-based DGPS	L3	Only outline parts, horizontal distance	Mean, STD	2.3.4

2.3.1 Qualitative methods: Overlay of outlines

The overlay of outlines (GO-1 in Table 3) is a mandatory step in determining product accuracy despite its qualitative nature. The method is used to: (a) correct the automatically derived glacier outlines (on-screen digitizing), (b) comparison to higher resolution datasets, (c) determination of differences in interpretation, and (d) visualisation of glacier change. Hence, this method is used to improve product accuracy a priori (a and b) and to communicate interpretation rules, potential shortcomings of the input dataset (e.g. snow cover), and usage restrictions of the dataset (Pfeffer et al., 2014). It is of key importance that outline overlay is performed on the original satellite image to identify regions of misclassification and subsequently correct these, as clouds, seasonal snow, debris, shadow and water can have a large impact on the mapped glacier area (see above). Practically, clouds are best identified in SWIR/NIR/red RGB composites, water in NIR, red, green, and debris or shadow in red/green/blue (natural colours). An example image in a related publication should focus on a worst-case region to correctly inform about the interpretation of these challenging regions by the analyst.

2.3.2 Quantitative methods I: Statistical extrapolation

In the absence of appropriate reference data, the following two methods are frequently used to determine precision: taking values from the literature that have investigated precision in more detail (e.g. Paul et al., 2013, Pfeffer et al., 2014) and applying it to the own dataset (GO-2), and the buffer method (GO-3) that expands and shrinks the outline of each glacier by an uncertainty value from the literature (e.g. ±1/2 or 1 pixel; Granshaw and Fountain, 2006; Bolch et al., 2010). Both methods have their shortcomings, e.g. GO-2 would require consideration of the size dependence (precision improves towards larger glaciers), and GO-3 is likely variable along the perimeter of a glacier (e.g. smaller buffer for clean ice, larger for debris-covered parts). Additionally, GO-3 should only be applied to glacier complexes (before intersection with drainage divides), to not provide any values where glaciers join. Whereas GO-2 is mostly applied as is (using some value between 3 and 5%), GO-3 is providing minimum and maximum values for each glacier that can be converted to a standard deviation (STD) when a normal distribution is assumed for the differences. The STD is then used as one component of the precision of the outline.

Further terms that are often but wrongly considered in the error budget are uncertainties related to (GO-4) geolocation, which is derived from the error of ground control points (GCPs) provided with the satellite data. Geolocation has no impact on the obtained glacier area as outlines are just shifted and should thus not be applied. The only exception is when quantities are directly derived from the digital intersection of outlines, such as glacier length changes (cf. Hall et al., 2003). The deformation of the outline by DEM errors (GO-5) propagating into the orthorectification is another issue. This indeed impacts on the glacier area but has so far never been assessed. It would require a comparison with an outline created at the same date, but using a ‘near perfect’ DEM (photogrammetrically derived) with a much higher spatial resolution than the satellite data.

2.3.3 Quantitative methods II: Analysts precision

As described above, manual correction of glacier outlines is required in most regions and the related corrections introduce uncertainty as they are based on subjective interpretation and generalization. It is thus not possible to repeat a manual digitization consistently. This variability can be used as a measure of uncertainty, given the analyst performs independent, multiple digitisations of a set of glaciers (GO-6). From the experience of a former study with more than 15 participants (Paul et al., 2013) we recommend that the analysts precision be obtained from such a multiple digitization experiment whenever manual digitization has to be performed to correct glacier outlines. The sample should consist of about 5-10 glaciers of different size and challenges (clean, debris, shadow, attached snow fields) that are representative for the manually digitized glacier sample. Each glacier should at least be digitized three times without checking the previous outlines (e.g. with one day between each round). For each glacier the resulting mean area and the STD should be calculated. Plotting the latter vs. glacier size will likely show an increase of the STD towards smaller glaciers (e.g. Fischer et al., 2014). A regression through the data points might provide an equation that can be used for size-class specific up-scaling to the full dataset (Pfeffer et al. 2014).

2.3.4 Quantitative methods III: Comparison to reference data

In the case an appropriate reference dataset is available (same date, higher resolution, same analyst) a one-to-one comparison of glacier extents can be performed (GO-7) to estimate accuracy of the derived glacier extents. Assuming that the outlines for the reference dataset are digitised manually, it is recommended to digitize them independently at least three times and use the mean area as the reference value. The relative area difference of the lower resolution area to the reference value provides the accuracy for an individual glacier. If extents of several glaciers are available as a reference, a mean difference and STD of the accuracy can be calculated. Due to the normal distribution of extent over and underestimations, mean differences are often close to the reference data. The more interesting value is thus the STD that can be seen as an estimation of the variability of the biases. However, multiple reference datasets are seldom available and for small samples it would be better to provide the range of differences (or a histogram).

It is also possible to calculate the mean distance of outlines (GO-8) but this requires some special software (Raup et al., 2014) and an extra-effort that is in general not taken as the simple overlay of outlines provides similar results (Paul et al., 2013). Both studies along with some others revealed that outlines are located within one (clean ice) or two (debris-covered ice) pixels if measured perpendicular to the direction of the outline. Application of this method has thus provided the values commonly applied to the buffer method (GO-3).

Finally, it is possible to obtain outlines of a glacier from field-based DGPS surveys (GO-9). These might only include a part of the outline as walking around a glacier can be difficult in its steep upper region (bergschrund, avalanches, etc.). However, for small ice caps it might be well possible to walk around their perimeter (at the time of satellite overpass) to obtain such a reference dataset. It might even be more precise than accurately orthorectified aerial photography, but its compilation is compromised by the large effort to obtain it and thus the rare availability. In the case such a dataset is available, the same calculations as described under GO-7 and GO-8 can be performed.

2.3.5 Examples

For two glaciers in the Austrian Alps we have applied some of the above methods to obtain how the uncertainty changes with the method applied (Table 4). In Fig. 4 some of these measures (GO-1, 3, 6 and 7) are illustrated. The values reveal that the often applied 3% precision for both glaciers gives a reasonable estimate for the larger one (Gurgler Ferner), but is likely too small for the smaller one (Hinterer Guslarferner). This assumes that the values obtained from the two other methods (GO-3 and GO-6) are more realistic, as they consider the size dependence better. The buffer method (GO-3) gives somewhat higher values than the multiple digitizing (GO-6), i.e. a lower precision, but this result for only one glacier should not be over-interpreted. Comparison with the reference data (the mean value of a multiple digitizing) gives an accuracy of -2.9% for the area derived automatically from TM. Considering the uncertainty of the manual digitization for this glacier, one can say that manual delineation of clean glacier ice is as good as automated mapping.

Table 4: Values of precision for two glaciers of different size. Precision is given as 67% of the min/max value. For GO-7 the column 'Glacier 1' gives the variability of the digitizing using the high-resolution image and the last column gives the resulting accuracy of the area derived by Landsat.

			Area min/mean/max/difference [km^2]		**Precision [%]**
Nr.	**Name**	**Measure**	**Glacier 1**	**Glacier 2**	**Gl1 / Gl2**
GO-2	Literature value	±3%	0.507/0.531/0.555/0.024	8.536/8.936/9.336/0.40	±3 / ±3
GO-3	Buffer method	±1/2 pixel	0.463/0.531/0.601/0.069	8.455/8.936/9.411/0.48	±8.7 / ±3.6
GO-6	Multiple digitizing	STD	0.511/0.560/0.610/0.05	8.56/8.92/9.40/0.36 to 0.48	±6.1 / ±2.9
GO-7	Reference area	Difference	0.540/0.547/0.556/0.008	n/a	-2.9 / n/a

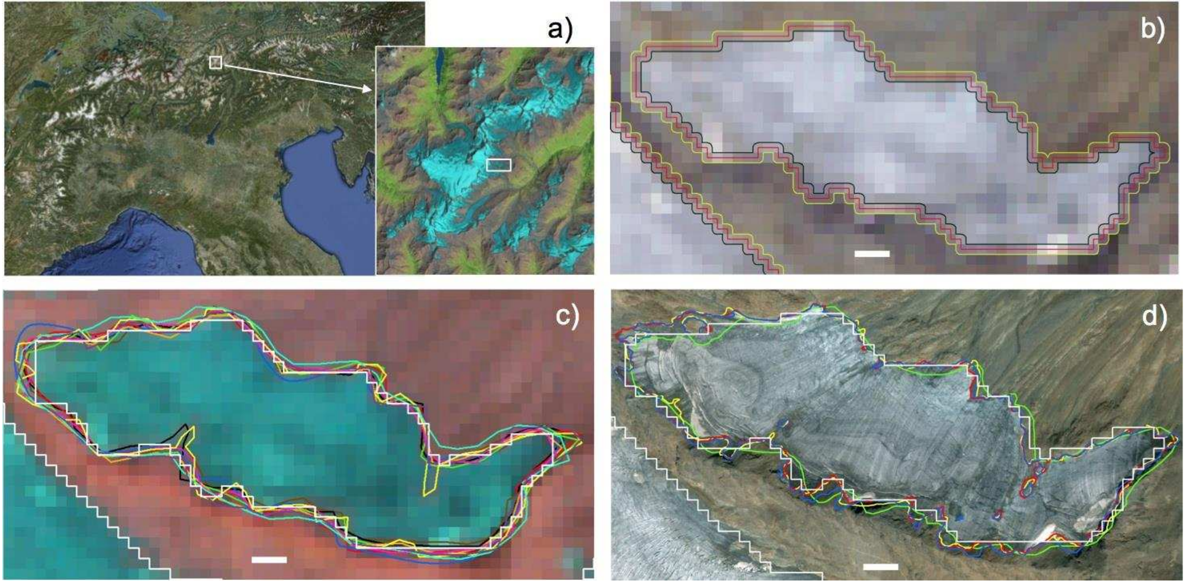

Fig. 4: Illustration of three methods used to determine uncertainty for glacier outlines. a) Location of the study glaciers in Austria (the main image is a screenshot from Google Earth), b) buffer method GO-3 (±1/2 pixel) illustrated for the smaller glacier, c) multiple digitizing (GO-6) for the glacier in b), and d) comparison to a reference area (GO-7) for the glacier in b). Panels b) and c) are based on 30 m Landsat images whereas d) is from Quickbird (screenshot from Google Earth). The white bar measures 100 m, North is up.

2.4 Recommended strategy

The above possibilities for assessment of product accuracy and precision vary in regard to the required effort and data availability. In general, the more simple methods only provide precision rather than accuracy. For tiered system presented below we recommend applying the lowest level in any case and the higher levels as possible. Abbreviations of the glacier outline (GO) number refer to Table 4.

Level 0

Overlay of outlines (GO-1) on the satellite image used to produce them is performed in any case for the internal manual editing in the post-processing stage (clouds, water, debris, shadow). It should also become a standard in a publication to illustrate external factors (snow/cloud conditions and interpretation rules). Whereas this qualitative method does not provide any measure of accuracy or precision, it reveals potential sources for deviations and has thus to be considered in the discussion.

In the absence of any further estimates specific to the dataset, a value describing precision should be selected from the literature (GO-2), justified for the current study (considering histograms of clean vs. debris covered and large vs. small glaciers), and applied to the sample, at best size class specific.

Level 1

The buffer method (GO-3) provides a minimum/maximum estimate of precision that scales with glacier size. Its overall value will thus vary with the size distribution of the selected sample and is thus more specific to the dataset under investigation than GO-2. It should be used instead of GO-2 when possible. A size-class specific calculation is recommended rather than just applying one mean value.

Level 2

The likely best method to determine precision of a dataset generated by one analyst is the multiple digitising of glacier outlines (GO-6). This gives the most realistic (analyst-specific) estimate for the provided dataset. Despite its higher workload, it is recommended using this method instead of GO-2 or GO-3. As for level 1, a size dependent regression should be used for up-scaling to the entire dataset.

In case several analysts have created the outlines, it is recommended that all analysts digitise a couple of glaciers (at least 3, better 5 to 10 of different size) independently after rules for interpretation have been settled. This would provide a measure for the consistency in interpretation and should be reported along with the results (mean and STD) for Level 2a

Level 3

This level requires the use of an appropriate reference dataset for accuracy assessment (GO-7). As the glacier outlines from the reference dataset are likely digitised manually, it is recommended to also apply GO-6 to determine its precision. It is well possible that its precision is within the accuracy of the test dataset (e.g. Paul et al., 2013). If possible, outlines from several glaciers with different characteristics (size, debris, shadow) should be used for accuracy assessment. To also have an estimate of precision, the measures of Level 2 should be applied additionally. The related overlay of outlines is most welcome in a publication.

If the required software exists, a mean horizontal distance between the outlines can be calculated and reported (GO-8). An estimation based on an overlay of outlines can also be used. If possible, the differences should be calculated separately for outline segments representing debris-covered and clean ice.

If ground-based reference data like dGPS are available (GO-9), the calculations described under Level 3a (complete outline) and 3b (segments) should be computed.

3. Elevation Change (altimetry)

3.1 Processing lines

Rates of surface elevation change over glaciers and ice caps that are sufficiently large and flat can be computed using repeat measurements of surface elevation from satellite altimeters such as on CryoSat-2 (e.g., Gray et al., 2015; Trantow and Herzfeld, 2016), EnviSat (e.g., Rinne et al., 2011a and b) and ICESat (e.g., Moholdt et al., 2010; Bolch et al., 2013) or in combination with a DEM (e.g., Kääb et al., 2012; Neckel et al., 2013). The three altimeters differ by the size of their footprint, beam wavelength / frequency (laser and radar) and measurement principle. These properties impact differently on the uncertainties of the derived product (e.g., radar penetration into snow and ice vs. impact of clouds and atmospheric scattering on laser). Moreover, due to the non-exact repeats of the satellite tracks, several methods have been developed to separate the effects of elevation change in space and in time (e.g., cross-over, across-track, plane-fitting, DEM reference for ICESat) (e.g. Moholdt et al., 2010), all with different impacts on product uncertainty. Due to the small footprint of the altimeter on ICESat (about 70 m), it has also been applied to detect elevation changes over comparably small mountain glaciers (e.g., Bolch et al., 2013; Gardner et al., 2013; Treichler and Kääb, 2016).

All altimeters measure surface elevation by converting the time delay between the pulse transmission and the surface echo return to a distance and then subtracting it from the well-known elevation of the sensor above a reference ellipsoid. The now decommissioned ICESat had 18 observation campaigns of about 35 days duration between 2003 and 2009 (Wang et al., 2011). Cryosat-2 has been providing data since 2010 and, at the time of writing, is still in operation. ICESat's reported single-shot accuracy of 0.15 m over gently sloping terrain (Shuman et al., 2006) was confirmed in subsequent studies (e.g. Treichler and Kääb, 2016). Whereas clouds limit data availability from ICESat, the measurement principle has no issues with surface penetration or missing optical contrast over homogenous (snow) surfaces. In consequence, ICESat data are frequently used for validation (accuracy assessment) of DEMs in different regions of the world or as a reference to register DEMs (e.g. Nuth and Kääb, 2011; Gonzales et al., 2010; Gruber et al. 2012; Pieczonka and Bolch, 2015; Treichler and Kääb, 2016 and references therein). Most uncertainties (for instance apart from geolocation, clouds, terrain roughness) are introduced by the methods used for the further processing of the raw data (filtering, spatial aggregation, plane fitting) rather than by the measurement itself.

In the following we shortly describe the CryoSat-2 processing in Glaciers_cci as ICESat processing has been described in detail before (e.g. Wang et al. 2011). The CryoSat-2 altimeter operates in Synthetic Aperture Radar Interferometric (SARIn) mode and has also been applied over regions of complex topography, such as mountain glaciers and ice caps. This novel mode allows precise location of the returned echo in the across-track plane and addresses some of the limitations associated with conventional pulse-limited radar altimeters. To compute linear rates of elevation change, CryoSat-2 records are grouped into grid cells, and then the various contributions to elevation fluctuations within each grid cell are solved for using the following model:

$$z(x, y, t, h) = \bar{z} + a_0 x + a_1 y + a_2 x^2 + a_3 y^2 + a_4 x y + a_5 h + a_6 t$$

Elevation (z) is modelled as a quadratic function of surface terrain (x, y), a time-invariant function of the satellite heading (h, assigned a value of 0 or 1 depending upon whether it was acquired on an ascending or descending pass), and a linear function of time (t). Further details relating to the model are given in McMillan et al. (2014; 2016). Following analysis from previous radar altimeter missions (Wingham et al., 1998; Davis et al., 2005), a backscatter correction is applied based upon the local covariance between elevation and backscatter (McMillan et al., 2014). The correction is computed for each grid cell (Davis et al., 2005; Flament and Rémy, 2012). Grid cells where the elevation rate solution is poorly constrained are then removed, based upon statistical thresholds from the model fit. These include thresholds of the Root-Mean-Square of the residuals, the elevation trend magnitude, the slope magnitude (as derived from the model fit), and the number of measurements that ultimately constrained the solution. The processing line is thus aiming at removing most of the outliers to reduce uncertainties, but the specific settings for the filters vary and thus impact on the result.

3.2 Factors influencing product accuracy

For Cryosat 2, the principle factors affecting the accuracy of measured rates of surface elevation change are (1) temporal fluctuations in the altimeter range due to variations in snowpack properties, and (2) limitations in the model's capacity to correctly partition the elevation fluctuation within each grid cell. In the case of the former, temporal variations in snowpack liquid water content, density and roughness can alter the depth distribution of the backscattered energy and impact upon radar altimeter elevation measurements (Scott et al., 2006; Gray et al., 2015). As a result, changes in snowpack properties, for example driven by anomalous melt events (Nilsson et al., 2015; McMillan et al., 2016), can introduce artificial elevation changes. To mitigate these effects, a backscatter correction is implemented which is designed to account for correlated fluctuations in elevation and power during the observation period. Alternatively, a re-tracking algorithm, which aims to reduce sensitivity to the volume echo, can be used (Davis et al., 1997; Helm et al., 2014; Nilsson et al., 2016). However, the latter may be more sensitive to short term snowfall fluctuations. Formally determining the uncertainty associated with this correction is, however, challenging and further research into understanding the radar wave interaction with the snowpack is ongoing. Until then, it is recommended to conduct additional independent evaluation using external data sources to confirm data accuracy.

The second principal factor affecting elevation rate uncertainty is due to the capability of the prescribed model of elevation change to fit the altimeter elevation measurements. Specifically, any deviation of the ice surface, and its evolution, away from the functional form of the model will introduce uncertainty into the model fit. As a result, rates of elevation change tend to become less certain in areas of complex topography or where non-linear rates of elevation change persist. This is reflected in the confidence associated with the parameters retrieved from the model fit and is discussed in more detail in Section 3.3.

Key sources of uncertainty for ICESat are (3) instrument related errors such as elevation biases between campaigns ("intercampaign biases", Urban et al., 2012), the range error due to degrading elevation precision (Borsa et al., 2014) or effects from geolocation errors, (4) uncertainty caused by the atmosphere such as saturation of the waveform or multiple peaks of the return beam (e.g. caused by reflections from clouds) and atmospheric propagation effects, i.e. the attenuation introduced by the scattering of water droplets and aerosols, and the multiple scattering phenomenon (Duda et al., 2001), and (5) uncertainties caused by the topography such as changes of terrain roughness and slope within the footprints, biases and spatio-temporal inconsistencies of the measurements, and the DEM, if used for differencing of the altimetric surface heights (Kääb et al., 2012; Treichler and Kääb, 2016). We do not discuss here uncertainties related to the spatial extrapolation of the point measurements to the entire glacier area or the spatio-temporal representativeness of footprint locations. An overview on the impacts of various techniques on the derived elevation changes is given by Kääb (2008).

3.3 Accuracy determination

In Table 5 we provide a sorted overview on measures to determine accuracy and precision for the elevation change from altimetry product that are described in the indicated sections in more detail. Due to the different nature of the altimeters and their data sampling strategy, some measures only apply to one of the sensors (e.g. ALT-3 and 4 for ICESat and ALT-5 to Cryosat 2). We do not provide an example for altimetry here as ICESat is used itself as a reference dataset and even more precise validation data for the same measurement points are rare.

Table 5: Overview of the measures to determine accuracy and precision of glacier elevation changes from altimetry (ALT)). The level refers to section 4.3. All mean values and standard deviations (STD) are expressed in absolute units.

Nr.	Name	Level	Measure	Format	Section
ALT-1	Instrument errors	L0	Provide the release/version used	Text	3.3.1
ALT-2	Topography	L1	List source data (DEM, glacier mask) and (slope) thresholds used, list old and new number of valid point counts	Text	3.3.2
ALT-3	Atmosphere	L1	List criteria and thresholds used, describe impact on point count	Text	3.3.3
ALT-4	Interpolation method	L2	one campaign trends or plane fitting residual, double differencing to reference DEM	Mean, STD	3.3.4
ALT-5	Model-fit accuracy	L2	1 Sigma uncertainty for each grid cell	Mean, STD	3.3.5
ALT-6	Reference data	L3	Difference (gives accuracy and precision)	Mean, STD	3.3.6

3.3.1 Instrument errors (ICESat)

Three individual lasers on ICESat were used in the different measurement campaigns and inter-campaign biases have been detected and related to the transmit energy and pulse shape as the individual instruments evolve. This particular error resulted in inter-campaign bias variations which were related to products that determined the range mixing a centroid for the transmit pulse and

Gaussian for the return pulse (Borsa et al., 2014). Corrections for these biases have been applied in updated versions of the datasets (Release 34) and for those products that were affected (i.e. GLAH06, GLAH14 products used centroid peaks for both the transmit and return pulses, so corrections do not apply). Biases through time and degrading elevation precision have also been detected from some of the lasers due to declining instrument transmit energy (Fricker et al., 2005; Borsa et al., 2014). Corrections for these bias trends approach the order of 1-2 cm per year, are not necessarily universal for each campaign rather varying in space and time (Borsa et al., 2014). Key requirements for the user are to work with the latest release of the data, to provide the release number, and to consider the potential effects of declining transmit energies on elevation change trends being calculated.

3.3.2 Topography (ICESat)

With increasing small-scale surface roughness and sloping terrain, the reflected pulse is spread more and its signal-to-noise ratio is reduced (i.e. the uncertainty is increased; e.g. Hilbert and Schmulius, 2012). To reduce the impact of this uncertainty, points are removed by statistical filtering. For example, slope derived from a DEM may be used to identify points located on slopes higher than a certain threshold that are to be excluded (Kääb et al., 2012; Treichler and Kääb, 2016). The threshold values used should be reported.

3.3.3 Atmospheric effects (ICESat)

Clouds and atmospheric effects (reflection/absorption, scattering, turbulence) impact on the form and intensity of the received signal (Fricker et al., 2005). They have a high spatio-temporal variability and thus need to be considered separately for each analysis. This resulted in the application of different statistical filters that exclude data points not meeting the prescribed criteria. As an uncertainty measure, the criteria applied to the raw dataset should be provided (e.g. Sørensen et al., 2011).

3.3.4 Interpolation method (ICESat)

The range of methods for accounting for the spatial offset in the repeat ICESat tracks when deriving elevation change rates have different associated uncertainties and methods for uncertainty estimation.

Following the three methods presented by Moholdt et al. (2010), precision can be determined from (a) elevation trends at cross-over points obtained within the same campaign (assuming changes are small within ~35 days), (b) doing the same but for neighbouring repeat tracks, and (c) using residuals of the plane-fitting method. When values from different campaigns are compared, the seasonality of the changes (e.g. snow fall during winter) needs to be considered by only selecting values from the same season. Method (b) requires a DEM to correct for slope and elevation related differences between two tracks. The precision to be reported is the STD of the differences measured by each method.

A second type of method is typically applied over mountain glaciers – double differencing (Kääb et al., 2012). ICESat elevations are differenced to a reference DEM (topographic normalisation) and elevation trends are estimated from the differences to the reference DEM. Thus, errors and uncertainty in the DEM propagate into derived elevation change products. The spatio-temporal consistency of the reference DEM turned out to be particularly important, and spatially variable biases and DEM elevation from different times, which is typical for DEMs composed from different sources, degrade the ICESat-derived products substantially (Treichler and Kääb, 2016).

3.3.5 Model-fit accuracy (CryoSat-2)

The elevation rate of change uncertainty is estimated at each grid cell using the 1-sigma uncertainty associated with this parameter from the model fit. This provides a measure of the extent to which our prescribed model fits the CryoSat-2 observations. In consequence, this term accounts for both departures from the prescribed model and for uncorrelated measurement errors, such as those produced by radar speckle and retracker imprecision.

3.3.6 Reference data (CryoSat-2 and ICESat)

The accuracy of elevation change rates from both sensors may be further evaluated through comparison with rates calculated from an alternative dataset. The requirements of such elevation rates are that they are coincident in both space and time, and are highly accurate. Elevation rates calculate from NASA’s IceBridge ATM data have commonly been used for this purpose, with the mean difference between elevation rates at coincident grid cells given as the measure for evaluation (McMillan et al., 2014; 2016; Wouters et al., 2015). For ICESat also DEMs from laserscaning and photogrammetry, and ground measurements have been used for comparison (Kropacek et al., 2014; Kääb et al., 2012; Treichler and Kääb, 2016).

3.4 Recommended Strategy

Level 0

It is always required to provide the release version of the dataset used for the calculations to be clear which kind of corrections have already been applied. These might also be shortly listed in the metadata and/or publication related to the dataset.

Level 1

Also the list of criteria and thresholds (statistical filters) used to compensate for topographic and atmospheric influences should always be given for the study region. It should also be described how the selection changed the sample count and if biases regarding their representativeness have to be expected due to the selection.

Level 2

Depending on the method applied to obtain elevation trends from ICESat, the related numbers should be calculated and provided in the metadata. As they can be calculated automatically their retrieval should be implemented in the processing line.

Level 2

For Cryosat 2 we recommend estimating the elevation rate of change uncertainty for each grid cell using the 1-sigma uncertainty associated with this parameter from the model fit as outlined in Section 4.2.1.

Level 3

If possible, the elevation rate of change should be evaluated through a comparison with coincident elevation rates calculated from an external data source, for example, IceBridge ATM data, as outlined in Section 4.2.2.

Level 4

Finally, thresholds for the selection of points from ALT-2 and 3 should be varied within reasonable limits and the impacts on the elevation change rates should be provided. Although the impact might be small compared to other effects and the processing might be demanding, we think this step is important to reveal that the very critical decisions taken for ALT-2 and 3 are insensitive to the overall outcome of a study.

4. Elevation Change (DEM differencing)

4.1 Processing line

Determination of glacier elevation changes derived from differencing of digital elevation models (dDEM) require (at least) two DEMs acquired at different times (Peipe et al., 1978; Reinhardt and Rentsch, 1986). The DEMs are typically generated from (a) satellite optical stereo images (i.e., ASTER, SPOT, Pléiades, WorldView), (b) Satellite Radar Interferometry (i.e., SRTM, TanDEM-X, ERS-1/2), and (c) aerial photogrammetry or laser scanning. Voids (data gaps) in optical imagery tend to occur in the accumulation area of glaciers due to a largely featureless surface or in regions of shadow. These voids can bias elevation change estimations, and several approaches for void handling are described in the literature (e.g., Kääb, 2008; Melkonian et al., 2013; Le Bris and Paul, 2015). They include, among others, interpolation of raw elevation values before differencing, interpolation of elevation changes to fill voids, and fitting of some function dh(z) to fill in gaps. Further challenges may arise with sensor arrays such as ASTER, due to platform shaking during acquisition ("jitter"; e.g., Ayoub et al., 2008), or due to shortening of steep terrain with back-looking sensors. For DEMs from InSAR, penetration of microwaves into snow/ice is highly variable, depending on the frequency of the microwaves and the snow conditions at acquisition (e.g. Dehecq et al., 2016). Biases introduced due to signal penetration can potentially be modelled and corrected, for example through comparison to elevation measurements acquired from the same time period using different frequencies or methods.

Before differencing, DEMs have to be checked for differences in their geoid and potentially re-projected to the same one. Afterwards they can be co-registered in x, y, and z to reduce biases caused by mis-alignment, a process that requires a glacier mask to ensure that only stable, off-glacier terrain is considered in the co-registration routine (Nuth and Kääb, 2011). Once the DEMs are co-registered, they can be differenced, and outliers can be detected and removed. The accuracy of the DEM differences can be estimated through calculating mean values of changes in pixels over stable (non-glacier) terrain. Importantly, all regional and global DEMs such as ASTER GDEM, SRTM, TanDEM-X IDEM, ArcticDEM, national DEMs, etc., are composed of individual raw DEMs and individual spatio-temporal biases are thus combined in such mosaics in a complex way that typically cannot be decomposed anymore (e.g., Nuth and Kääb, 2011; Treichler and Kääb, 2016).

4.2 Factors influencing product accuracy

4.2.1 Source data and pre-processing

The accuracy of glacier elevation changes derived from DEM differencing (dDEM) is influenced primarily by the accuracies, precision, and resolution of the individual DEMs that are differenced. These accuracies are dependent on the acquisition technique used – photogrammetric principles applied to optical images (i.e., aerial photos, ASTER, SPOT), interferometric techniques on repeat radar images (i.e., SRTM, ERS-1/2, TanDEM-X), or laser ranging (LiDAR DEMs), as well as the environmental conditions at the time of acquisition.

DEMs derived from optical stereo photogrammetry and LiDAR point clouds require cloud- and fog-free conditions and daytime, which can limit the temporal availability of DEMs and impact locally on their quality (e.g. in case of frequent orographic clouds). In addition, the largely featureless, low-contrast nature of the accumulation areas of many glaciers can limit the ability of photogrammetric techniques to reliably determine elevations in these areas, potentially leading to data gaps (voids). Accuracy may also decrease due to inaccurate determination of the satellite position and attitude, which introduces biases into altitude estimations. However, recent developments have helped to reduce these uncertainties in the pre-processing stage, reducing the overall certainty of DEM products derived from, for example, ASTER imagery (Girod et al., 2016). In general, the accuracy and resolution of DEM products derived from satellite-borne stereo optical photogrammetry has increased with time (i.e., SPOT and Pléiades are more accurate and have higher spatial and radiometric resolution than ASTER). In addition, DEMs generated from aerial photographs tend to have higher accuracy and resolution than those from satellite imagery. With DEMs that have recently been generated from very high-resolution satellite sensors such as Pléiades, Quickbird or WorldView, the gap in resolution and quality has been reduced (Shean et al., 2016) and first successful applications for volume change determination over comparably small glaciers were performed (e.g. Berthier et al., 2014; Holzer et al., 2015; Kronenberg et al., 2016).

DEMs derived from radar interferometry do not have the daytime or cloud- and fog-free restrictions that optical DEMs do. Whereas optical images portray the surface of glaciers and snow, however, radar signals penetrate ice and dry snow to varying depths dependent on snow and ice properties (i.e., moisture content and purity), as well as the properties of the signal itself (e.g., Rignot et al., 2001; Shugar et al., 2010). With simultaneously-acquired data of different frequency (i.e., SRTM C-band and X-band data), it is possible to estimate and correct for penetration effects locally, though these approaches are limited in extent and not universally applicable (Gardelle et al., 2012; Melkonian et al., 2014). Accuracy of radar interferometric DEMs is also dependent on precise knowledge of satellite orbital parameters, which tends to be lacking in earlier interferometric missions. Despite this, radar signals tend to be quite sensitive to small changes in topography, and so the overall accuracy of most radar interferometric DEMs is high (typically <15 m, as high as 2.5 m; e.g., Joughin et al., 1996; Dehecq et al. 2016). A good strategy to avoid the above issues is the comparison of DEMs from sensors with the same wavelength, e.g. the SRTM and TanDEM-X X bands (e.g. Neckel et al., 2013;

Rankl and Braun, 2016).

To ensure that the elevations being compared correspond to the same spatial location, DEMs must first be adjusted to the same vertical reference (geoid or ellipsoid) and then be co-registered. This co-registration can be accomplished manually (e.g., VanLooy, 2011), or through automated algorithms to reduce elevation residuals (e.g., Berthier et al., 2007; Nuth and Kääb, 2011). A comparison of four different methods for DEM co-registration (Paul et al., 2015) found that three automated solutions (e.g., Gruen and Akca, 2005; Berthier et al., 2007; Nuth and Kääb, 2011) performed similarly in terms of accuracy after co-registration, but with different efficiencies. In addition, different software packages have different routines for importing the same file format, which has implications for the pixel definition (pixel centre vs. corner), leading to co-registration errors if inconsistent.

Resampling of DEMs to lower resolutions, a necessary step when comparing DEMs of differing resolutions, can also reduce accuracies in the final product. A related study by Jörg and Zemp (2014) has shown that although the two DEMs were very accurately co-registered, systematic and random method- and scale-dependent errors still occurred. Well-documented elevation biases of up to 12 m km^{-1} have been described in SRTM data (Berthier et al., 2006; Schiefer et al., 2007; Paul, 2008). As noted by Paul (2008), these effects are most likely related to resampling of elevation data, introduced because of the curvature of high-elevation terrain, and not because of elevation per se. Further studies have extended these findings (e.g., Gardelle et al., 2012) to correct elevation biases using the maximum terrain curvature, and implemented in other studies using the SRTM data (e.g., Willis et al., 2012; Gardelle et al., 2013; Melkonian et al., 2013, 2014).

Finally, detection of significant elevation changes over glaciers depends on the time separation between DEMs, as well as characteristics of the glaciers in question. Fast-changing glaciers such as tidewater glaciers or surging glaciers will potentially show significant changes in a single year, while smaller alpine glaciers will tend to require more time between acquisition dates to show significant change, typically a decade (e.g. Zemp et al. 2013).

4.2.2 Post-processing and editing

One of the largest sources of uncertainty occurring in post-processing is the handling of voids in the source DEMs. In any region with voids, the dDEM product will have voids. In general, voids in DEM differencing products have been handled in one of four ways: (1) interpolating elevation values in the source DEMs before differencing (e.g., Kääb, 2008); (2) differencing the source DEMs, then interpolating elevation change values over the void areas (e.g., Kääb, 2008; Melkonian et al., 2013); and (3) utilizing the relationship between elevation change and elevation to estimate elevation change as a function of altitude, then applying this function to unsurveyed areas (e.g., Bolch et al., 2013; Kohler et al., 2007; Kääb, 2008; Kronenberg et al., 2016).

Each of these methods have their advantages and disadvantages. Kääb (2008) compared approaches (1) and (2), finding a mean difference in elevation changes of 1 ±12 m RMS between the two approaches. Generally, method (2) is likely a better approach, given that elevation changes over glaciers tend to be more self-similar in nearby regions than does elevation itself. Rather than interpolating values, other studies have filled voids by using the average elevation change calculated over the entire study area (e.g., Rignot et al., 2003), over a given elevation band in the study area, or over a given radius around the void (Melkonian et al., 2013). The latter is most likely more accurate than the other two, as the mean elevation change around the void is more likely to be reflective of the changes in the void, at least when the void does not stretch over too many elevation bands

A further critical issue for post-processing are artefacts that might result from a failed matching during DEM generation instead of data voids. Typically, these can be found in regions of steep slopes, low contrast (shadow, snow) or self-similar structures. They also result when the spatial resolution is blown-up to a value not supported by the original data. In this case the surface might appear 'bumpy' over large regions, i.e. the amplitude of the artefact is smaller but its occurrence is more frequent. When two DEMs with artefacts are subtracted, the artefacts from both DEMs will be transferred to the difference grid. Depending on the region where they occur (e.g. accumulation or ablation area) and their frequency and amplitude, different measures to remove or reduce them can be applied (local smoothing, threshold cut-off). For example, strong negative (positive) elevation changes are unlikely in the accumulation (ablation region) and can be disregarded by using an elevation dependent threshold (Pieczonka and Bolch, 2015), either setting the outliers to zero or no data. For artefacts with the correct sign (e.g. mass gain in the accumulation area), correction is more difficult as changes up to a certain value might indeed have occurred (Le Bris and Paul, 2015). In this case it might be helpful to also analyse their spatial pattern to reveal a possibly natural or artificial cause. For example a speckled pattern over steep slopes in the accumulation region of a glacier is a typical DEM artefact and should be removed (data void) or replaced by one of the three methods (1) to (3) mentioned before.

4.3 Accuracy determination

There is a large number of possibilities to determine the accuracy of elevation change products from DEM differencing either related to the DEMs itself or the subtracted DEMs. However, several secondary effects (e.g. differences in spatial resolution, terrain slope, optical or microwave source data) interfere and could result in misleading results. Similarly, stable terrain that should not show any vertical or horizontal changes over time and be found near the glaciers has to be carefully selected (e.g., no trees, lakes, or buildings, low slopes, different aspect sectors) and might need to be manually delineated to avoid misleading conclusions; it is not just all terrain off glaciers. In Table 6 we provide an overview of some key measures for accuracy and precision (internal ones and those requiring additional data) that are discussed in detail afterwards.

Table 6: Overview of the measures to determine accuracy and precision of glacier elevation changes from DEM differencing (DEM). The level refers to section 5.3. All mean values and standard deviations (STD) are expressed in absolute units.

Nr.	Name	Level	Measure	Format	Section
DEM-1	Co-registration	L0	Fit accuracies (horizontal/vertical)	Mean, STD	4.3.1
DEM-2	Stable ground	L0	Elevation differences	Mean, STD	4.3.1
DEM-3	ICESat reference	L1	Difference to ICESat points (stable ground)	Mean, STD	4.3.2
DEM-4	Vector sum	L1	Sum of offset from 3 elevation sources	Residual value	4.3.2
DEM-5	High quality DEM	L2	Difference (gives accuracy and precision)	Mean, STD	4.3.3
DEM-6	Ground control	L2	Comparison to field-based validation points	Mean, STD	4.3.3

DEM-7	points Changes by LIDAR	L3	Difference to change rates from LIDAR	Mean, STD	4.3.4

4.3.1 Co-registration and stable ground off-sets

This is an internal measure that only requires the two DEMs. Before they are are subtracted, datums have to be aligned and a proper co-registration (horizontally and vertically) has to be performed. The co-registration vectors can be determined analytically using a short script described by Nuth and Kääb (2011). The elevation points selected for the co-registration should be located on stable terrain which might require manual selection (e.g. via a polygon). The accuracies of the fit are directly provided as standard errors of the fitted offsets. In addition, the mean, median, STD, and RMSE of the elevation differences (vertical component) is calculated and should be reported with the dataset. Whereas the horizontal offset should be applied in any case, consideration of the vertical offset should be carefully checked before it is applied to the difference DEM. In particular when DEMs of different source (microwave and optical), spatial resolution or geodetic projection are compared. It is also possible that elevation differences have a non-constant shift that is not easily corrected with a mean value but can be estimated with a trend surface (e.g. Racoviteanu et al., 2007).

4.3.2 ICESat reference data and vector sum

In the case ICESat data are available for the study site they can be used in two different ways. First, elevation differences of the source DEMs can be calculated along the ICESat track considering the side impacts described above (time of the year, radar penetration, cell size, stable terrain). This will give accuracy (mean difference) and precision (STD) of the source DEMs that can be considered in the error budget. Secondly, the elevation values from ICESat can also be used in the co-registration process with each of the two DEMs. Ideally, the sum of the three horizontal shift vectors as well as of the vertical offsets is zero. Practically, this will not exactly be the case and a residual offset vector and vertical shift will remain. These values should be reported as well.

4.3.3 Comparison to reference data (high-quality DEM and GCPs)

In the case one of the two DEMs subtracted has a much higher quality than the other (e.g. it is derived from aerial photography or laser scanning) it can be used as a reference DEM to calculate accuracy and precision of the second DEM over stable terrain. To avoid a bias related to spatial resolution, it would be required to aggregate the higher quality DEM to the cell size of the second DEM (which likely has a lower resolution). A direct comparison is also possible with ground based GCPs, but these might only seldom be available and sample size is likely much smaller than for a reference DEM. The advantage of the latter could be that the high-quality reference DEM is only available for a small region whereas the GCPs might be available over the entire study region.

If the two DEMs are temporally consistent (e.g. SRTM C and X-band), comparison over glaciers provides glacier-specific biases (e.g., penetration of radar signals into snow/ice; e.g. Gardelle et al., 2012). This would be an important correction factor when one of the DEMs is subtracted in the same region from another dataset. It also provides a measure of uncertainty for the random differences. Before processing, the difference DEM should also be visually examined for any internal scene biases that may exist, for example due to errors in the sensor attitude determination (e.g., Surazakov and Aizen, 2006; Berthier et al., 2007). Removal of such signals is necessarily sensor- and scene-specific, as it depends on the source data used for DEM generation, and cannot be universally standardized.

4.3.4 LIDAR DEM differences

The above methods refer to the accuracy assessment of the source data rather than to the derived elevation changes. In rare cases it might be possible to directly compare them over a longer period of time as derived from high-resolution LIDAR data (acquired by air planes or drones) to the changes derived from DEM differencing (Jörg et al., 2012). Of course, the time periods analysed should be the same, but the pattern of changes or mean annual values per elevation band can also indicate accuracy. Over short time periods, however, one also has to carefully consider the timing (winter snow fall and summer ablation) and glacier dynamics (e.g. emergence and submergence velocities). They might have a considerable impact on the obtained differences and are difficult to correct.

4.3.5 Example for the region around Kronebreen (Svalbard)

We compared three DEMs over the region surrounding Kronebreen, Northwest Svalbard, to exemplify some of the methods applied for estimating accuracy and precision from DEM differencing. In Fig. 5, we show elevation differences (Fig. 2a and 2b) between an aerial photogrammetric DEM from 1990, a SPOT5 IPY-SPIRIT DEM from 2007 (Korona et al., 2009) and the recent TanDEM-X Intermediate DEM from December 2010 (https://tandemx-science.dlr.de/pdfs/TD-GS-PS-0021_DEM-Product-Specification_v3.1.pdf). Co-registration between the different DEMs was performed (measure DEM-2), using only the stable terrain, after resampling all DEMs to a spatial resolution of 40 m using a block averaging routine to minimize effects related to resolution (e.g., Paul, 2008; Gardelle et al., 2012). After co-registration, the mean and median bias are all less than a metre while the standard deviations are less than about 10 m for all three comparison (Table 7). Fig 2c shows the histograms of the elevation differences on stable terrain and on the glaciers (DEM-2), revealing the significance of the changes over the glaciers during the 17 and 3-year periods.

Table 7: Results of the co-registration and stable terrain statistics for the DEM differencing example shown in Fig. 2. All mean values and standard deviations (STD) are expressed in absolute units.

	Coregistration parameters (m)			**Stable terrain statistics**		
DEM difference	**dx**	**dy**	**dz**	**mean**	**median**	**STD**
2007 (slave) - 1990 (master)	-6.7	-4.95	4.17	-0.13	0.13	9.81
2007 (slave) - 2010 (master)	2.59	-9.52	2.9	-0.05	0.04	6.35
2010 (slave) - 1990 (master)	-10.38	3.41	1.98	0.71	0.22	10.01
2010 (slave) - 1990 (master)	-10.38	3.41	1.98	0.71	0.22	10.01
1990 (slave) - ICESat (master)	0.21	-2.24	-1.57	-1.65	-0.14	17.57
2007 (slave) - ICESat (master)	-6.99	-6.04	4.56	-0.18	0.07	8.27
2010 (slave) - ICESat (master)	-10.63	1.51	1.4	-0.03	-0.07	6.26
Vector SUM (1990/2007/2010)	*-1.09*	*-1.16*	*0.71*			
Vector SUM (1990/2007/ICESat)	*0.5*	*-1.15*	*-1.96*			
Vector SUM (1990/2010/ICESat)	*0.46*	*-0.34*	*-0.99*			
Vector SUM (2007/2010/ICESat)	*-1.05*	*-1.97*	*-0.26*			

Furthermore, we used ICESat as reference for co-registration (DEM-3) and calculated the vector sum (triangulation) between co-registration vectors (DEM-4). They are all less than 2 m for each combination of DEM and ICESat. These precisions are much higher than the original DEM resolutions of 40 m and that of the 90 m ICESat footprint. The largest standard deviation between the 1990 DEM and ICESat is a result of rather limited stable terrain on the DEM resulting in a sample size of less than 1000 points. Finally, an elevation change profile is shown along the first 25 km of Kronebreen in Fig 2d, revealing the larger thinning rates on this glacier in the most recent 3-year period as compared to the 17-year thinning averages since 1990.

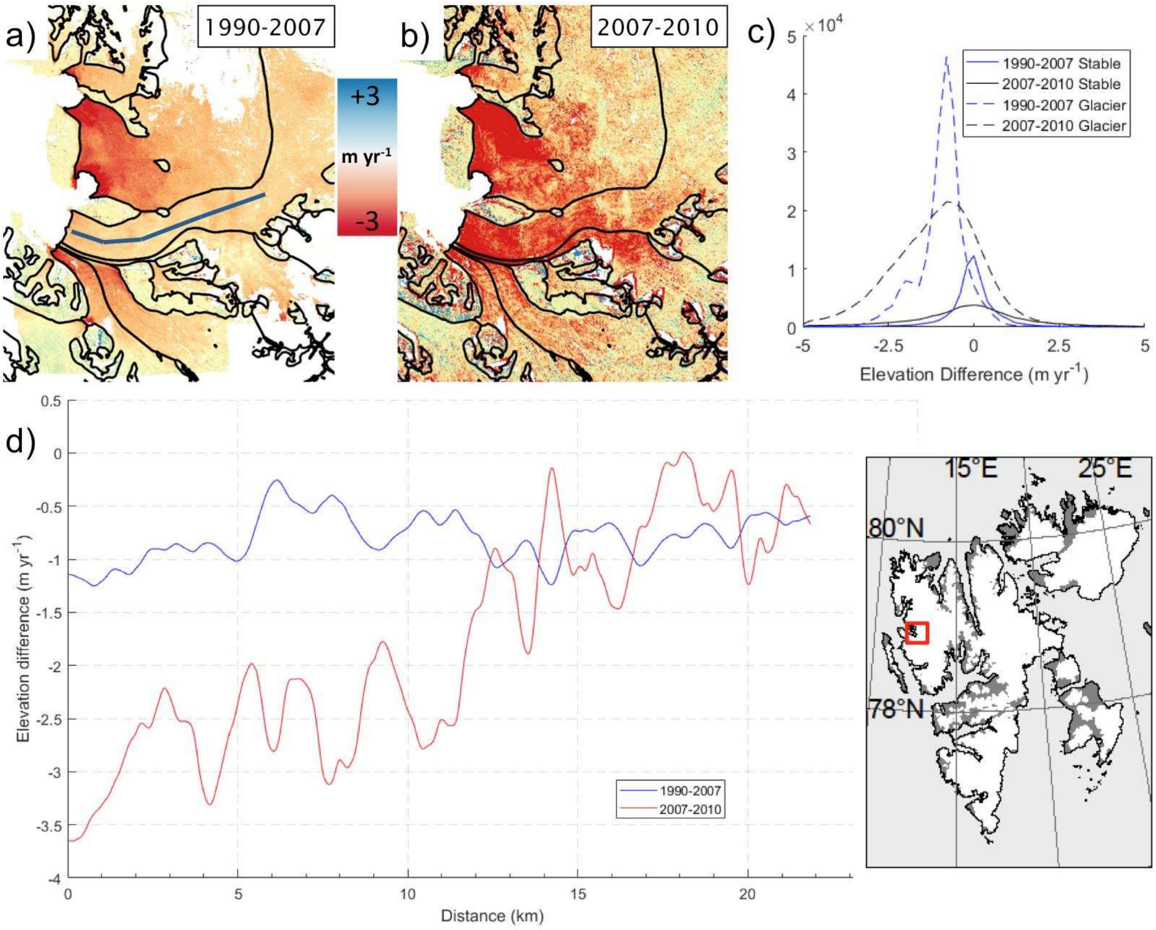

Fig. 5: Illustration of elevation differences on stable terrain and glaciers between a) 1990 and 2007 and b) 2007 and 2010 for Kronebreen in Svalbard (see red square on the inset for location). c) Elevation difference histograms for stable terrain and glacier ice. d) Elevation change centreline profiles along Kronebreen for both epochs, revealing higher loss rates near the terminus in the more recent period.

4.4 Recommended Strategy

Level 0

We recommend that co-registration of the two DEMs is always performed and the resulting horizontal and vertical shifts (mean and STD) over stable ground are always reported. This is an absolute minimum to determine whether the observed changes over glaciers are significant or not. It should also be reported if the mean vertical shift over stable ground was applied.

Level 1

In most glacierized regions at least some ICESat tracks also cover mountain ranges. It is thus recommended to use this information for accuracy assessment of the two DEMs used to obtain the elevation change over glaciers. Careful consideration of differences in spatial resolution needs to be considered. If the number of points from ICESat is sufficiently large, a small additional effort will reveal the co-registration offsets between all three elevation sources and the possible residual error. This would be one step closer to the truth as otherwise compensating systematic biases in both source DEMs can be revealed and reported. Overall, ICESat elevations can be (still) considered the best global elevation reference frame for glacier remote sensing (Nuth and Kääb, 2011) and is thus useful to check and potentially improve the accuracy of DEMs and derived elevation differences.

Level 2

This measure can only be applied if one of the two DEMs has a much higher quality than the other one or if an external DEM with superior quality (e.g. derived from airborne photogrammetry or LIDAR) is available. Differencing the two will provide accuracy and precision of the other (or both) DEMs over stable terrain. The same is true for GCPs but these might be even more rarely available.

Level 3

For some glaciers precise elevation changes from repeat aerial photogrammetry or laser scanning are available. In the case of a temporal coincidence with the satellite-based measurements, these can be used for validation of the latter.

5. Velocity

5.1 Processing line

Glacier surface velocities can be derived from both high-resolution optical (e.g., Scherler et al., 2008; Heid and Kääb, 2012; Dehecq et al., 2015) and SAR repeat satellite data (e.g., Strozzi et al., 2002; Quincey et al., 2009; Nagler et al., 2015; Schellenberger et al., 2016). Optical sensors are sensitive to surface features only, whereas microwave signals penetrate into dry snow and firn from depths of a few centimetres up to several tens of metres, depending on the signal frequency and properties of the snow and ice. However, radar penetration is in general neglected, as surface flow velocities do not change much with depth. Typically, block and offset matching techniques are employed to estimate surface motion from satellite images, with the kernel size adjusted to the resolution of the satellite data, the time period and the expected displacements (e.g. Debella-Gilo and Kääb, 2012). These techniques demand co-registered images with sub-pixel accuracy. For optical images, with an almost nadir view, accurate orthorectification is needed before matching. SAR images, with their peculiar side-looking geometry, are preferable matched in the SAR imaging geometry, e.g. slant range and along track coordinate system, to avoid distortions caused by geocoding in regions of layover and shortening both of which are amplified by low quality DEMs. Offset matching techniques provide two-dimensional displacements in ground-projected geometry for optical imagery and in slant-range geometry for SAR imagery. The latter are then geocoded into a map projection using a DEM and converted to horizontal or slope parallel velocity components (e.g. Paul et al., 2015). Post-processing includes filtering of outliers based on correlation strength, magnitude and angle of displacement, or neighbourhood similarity. Glacier outlines are used to obtain ice-free terrain for accuracy assessment.

5.2 Factors influencing product accuracy

5.2.1 External factors and source data

Glacier surface conditions, structure and terrain complexity all have a direct impact on the quality of image correlations. Generally, cross-correlation algorithms perform best when distinctive intensity features are present for tracking with regard to the size of the applied matching kernel and the spatial resolution of the satellite images. As with DEM generation, for optical imagery the presence of snow or clouds reduce precision. In addition, illumination conditions on the ground can complicate the matching process of optical images, in particular in areas where there is little to no visual contrast or sensor saturation (e.g., shadow, fresh snow, or the accumulation areas of many glaciers), features that are self-similar (e.g., seracs or ogives), or contrast that defines only one offset dimension (e.g., longitudinal moraines or flow strips with no variations in contrast). Many of these issues have been reduced with the transition to 12-bit radiometric resolution in the recent Landsat-8 OLI and Sentinel-2 MSI instruments (Kääb et al., 2016). SAR sensors are sensitive to snow and ice conditions on the glacier surface, in particular to the presence of liquid water, which can significantly reduce the quality of the results.

Vertical error components in the DEMs used for orthoprojection of optical images translate to horizontal displacement errors. This effect is typically negligible when utilizing data from the same track but if data from different orbits are used, horizontal displacements on stable ground will be visible (Kääb et al, 2016). Because DEM errors that propagated into the orthorectified images are not analytical in nature, they cannot be corrected or removed. However, displacements for stable ground provide an estimate for the overall effect of these errors, at least when disregarding surface elevation changes and the often existing temporal mismatch between DEM and image acquisition. Systematic errors in the provided or modelled sensor attitude angles (i.e., jitter) lead to corresponding patterns in displacements calculated from optical data. Depending on their nature, and provided that many well-distributed off-glacier offsets are available, they could be statistically modelled, and on-glacier displacements could be corrected (e.g., Scherler et al., 2008; Nuth and Kääb, 2011). SAR sensors, on the other hand, are sensitive to ionospheric scintillations, causing shifts in azimuthal position ("azimuthal streaking", Wegmüller et al., 2006; Strozzi et al., 2008; Nagler et al., 2015). They are especially visible in SAR images of high latitudes and depend on solar activity. The streaks are visible in azimuthal offset maps and can be reduced by high-pass filters along the range direction (Wegmüller et al., 2006). The wavelength employed by the radar sensor has a large impact on ionospheric artefacts, which are typically larger at lower frequencies.

It should also be noted that cross-correlation algorithms provide displacement estimations for the time period between image acquisitions. Thus, the derived velocities represent the mean value over the observation period and cannot account for short-term velocity variations between the image acquisition dates. This fact is particularly important when time series of glacier velocities are analysed.

5.2.2 Algorithm application

In the implementation of the normalized cross-correlation algorithm, the choice of the matching window size and the oversampling factor have a direct consequence on the precision of the estimates, the noise level, as well as the computational time required. The choice of the matching window size will also depend on the target being observed and on the spatial resolution of the source data (Debella-Gilo and Kääb, 2012). For SAR sensors, estimates using very large window sizes (e.g., 512 x 512 pixels) are generally more precise for large structures, but are not applicable to small (e.g., < 500 m width) glaciers, nor do they provide information in shear zones (Strozzi et al., 2002; Paul et al., 2015). This drawback can be overcome by using iterative algorithms with a variable matching window size (Debella-Gilo and Kääb, 2012; Nagler et al., 2015; Euillades et al., 2016). For optical sensors, these window sizes are typically 10-30 pixels wide, and in general, larger window sizes produce better accuracy for large structures, though the same drawback applies. Thus, a necessary trade-off exists and must be considered in the implementation of the algorithm (Debella-Gilo and Kääb, 2012). The implementation of the cross-correlation algorithm (that is, the choice of window sizes used) has a direct impact on the noise levels, and therefore the accuracy, in the resulting displacement estimates.

When working with SAR images, apparent offsets between two images are a result of the different orbit configurations of the two images, stereo offsets, ionospheric effects, noise, and the actual surface displacement between the image acquisition times. To accurately determine the displacement of the surface, then, all of the other contributions to the offsets must be carefully characterised and removed. Orbital offsets are determined by fitting a bilinear polynomial function to offset fields computed globally from the SAR images, assuming no displacement in most of the image. Stereo offsets are relevant for the range-offset field, and depend on the height of the target, the baseline between the two satellite orbits, the height of the satellites above the Earth’s surface, and the incidence angle of the satellite. Stereo offsets can be avoided by co-registering the two SAR images with topography considered, which necessarily requires an accurate DEM. Ionospheric contributions are discussed in section 6.1.1, noise removal will be handled in section 6.1.3. Residual errors on stable ground are used to inspect the results against systematic residual offsets.

5.2.3 Post processing and editing

Filtering the results of the matching outcomes is a critical processing step. A trade-off is necessary at this stage, as well, in terms of the number of estimates versus confidence level, or the number of mismatches kept and correct matches discarded as a result of the filtering process. This filtering step can be implemented by using a simple threshold of the signal-to-noise ratio or correlation coefficient, by iteratively discarding matches based on the angle and size of displacement vectors in the surrounding area (e.g., Burgess et al., 2012), by using high- or low-pass filters on the resulting displacement fields, or through some combination of these approaches (Paul et al., 2015). In image series of higher temporal resolution, triplet matches can be performed over all three pair combinations in three images and the results be triangulated to indicate inconsistent measurements and thus outliers (Kääb et al., 2016).

5.3 Determination of precision and accuracy

Validation of glacier displacements measured from spaceborne sensors compared to ground-based data is inherently difficult. This difficulty arises from the following main sources:

- Coincident observation: As a consequence of highly-variable sub-glacial hydrology, glacier surface velocities are variable temporally, with diurnal, seasonal, and interannual cycles (e.g. Vieli et al., 2004; Allstadt et al., 2015). Therefore, comparisons should be done between coincidently acquired data sets.
- Spatial scale: Measuring glacier displacements from satellite images requires the comparison of image windows. As such, the motion estimated results from motion of large areas of features,

and is not necessarily representative of the motion of individual features or points. This representativeness is furthermore 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. Thus, direct comparison to point measurements such as GPS displacements are suitable for areas with homogeneous velocity fields, but are not necessarily straightforward in shearing zones or regions with significant spatial velocity variations such as calving fronts.

- Different velocity components: In-situ surface velocity is measured by GPS at stakes, representing the 3D displacement of the surface due to several processes (horizontal, displacement, ablation, movement along slope, etc.). From space, cross-correlation techniques using optical images determine the horizontal displacement at the surface while SAR images measure Line-Of-Sight (LOS) and along-track displacement. To validate or compare products from these different methods requires first transforming measurements to the same velocity component.

If suitable reference data exist, accuracy or bias of ice surface velocity data can be estimated with field measurements and independent images, respectively. In the absence of suitable ground-based data for comparison, uncertainties in velocity-based products should be characterized based on internal measures. For practical purposes, we suggest the tiered system of levels as summarized in Table 8 and section 5.4.

Table 8: Overview of the possibilities to determine the accuracy and precision of glacier velocity products.

Nr.	Name	Level	Application	Measures	Section
IV-1	Overlay of outlines, spatial consistency of flow field	L0	Visualization, outlier detection	Descriptive	5.3.1
IV-2	CC/SNR	L1	Quality map of correlation coefficients and/or signal-to-noise ratio values	Coefficient	5.3.2
IV-3	Stable ground velocities	L1	Statistical measures	Mean, STD	5.3.3
IV-4	Consistency of time series	L2	Analysis of time series of ice velocity at profiles and points	Mean, STD Trends	5.3.4
IV-5	Comparison to higher resolution data (different sensors)	L2	Cross-validation with very-high resolution reference images	Mean, STD	5.3.5

IV-6	In-situ data (dGPS)	L3	Validation with temporally and spatially coincident ground-truth	Mean, STD	5.3.6

5.3.1 Overlay of outlines and outlier detection

The computed surface velocity maps can be visually inspected with overlaid glacier outlines by (i) evaluating the spatial consistency of ice flow patterns regarding both direction and magnitude, (ii) checking for outliers remaining after filtering, (iii) checking for unnatural patterns in the displacement field considering that ice flow is in a (roughly) downslope direction. Though subjective, these qualitative checks rely on basic physical principles, such as the incompressibility of ice or glacier flow under gravity, and should be done as a final step before validation.

The physical properties of glacier ice, such as incompressibility and transfer of stresses, combined with the low spatial variation in gravity that drives glacier flow means that glacier velocities tend to be relatively smooth and coherent. As a result, different frequencies of the velocity field can be compared, and results that differ too much from expected low-frequency values can be discarded. The qualitative (visual) check of the spatial coherence of the flow field allows application of a quantitative measure (a filter) to remove related outliers (e.g. Skvarca et al., 2003). This typically gives good results, but it fails entirely where entire zones of measurement are inaccurate, or where a glacier has high local velocity gradients.

5.3.2 Matching quality measures

Most algorithms will either provide directly, or with some additional processing, quantities to describe the degree of similarity between the matching image windows; typically these are either the correlation coefficient (CC) or signal-to-noise ratio (SNR). These parameters provide an indication for the reliability of an individual match, though this measure is not strict: bad matches may still reflect the true displacement, and matches with a high score may not. Thus, this measure should not be used on its own for validation.

5.3.3 Stable ground

Stable and ice-free ground in the images can be matched to give a good indication for the overall co-registration of the two images, and some general idea of the matching accuracy under the specific image conditions. The representativeness depends on the image content similarity between the stable ground and the glacier areas. Additionally, as a side quality indicator, the percentage of successful matches over ice can be provided. The above-described triplet matching and subsequent triangulation of displacement vectors includes the idea of independent matches into the post-processing step.

5.3.4 Consistency of velocity time series

This test is suitable for glaciers with systematic acquisition of time series of satellite images. Especially, since the launch of Sentinel-1 and Sentinel-2 in 2014 and 2015, respectively and the systematic acquisition planning and short repeat observation intervals over many mountain regions the test becomes increasingly useful. For example, Sentinel-1 A/B provides a 6-day repeat interval. The test assumes that over short time intervals the ice velocity of most glaciers is stable or shows trends over several observation cycles. The test can be applied at selected regions of the glaciers with homogeneous velocity providing the temporal mean and standard deviation, and temporal trend of the velocity, or the velocity along selected profiles (e.g. central flow line). Obviously, systematic inconsistencies in the employed tracking algorithm will not be revealed by this test.

5.3.5 Comparison to higher resolution data

Satellite-derived displacements can be compared to products derived from independent image data when available. That is, they can be compared to measurements derived from data of equal or better resolution, accuracy, and precision. The discrepancy between the products is then a function of the accuracy of both matches, the co-registration between the two sets of images (that is, their relative geocoding), the representativeness of the displacement compared to the "true" displacement, and the temporal variations between the acquisition dates of the two sets of images.

5.3.6 Comparison to field measurements

Satellite-derived displacements can be compared to field measurements, provided that the above-described considerations about temporal and spatial consistency and different velocity components are taken into account. Though these field-based measurements tend to be very precise, the temporal and spatial representativeness of these measurements as compared to the satellite-derived measurements will vary and is not strictly known.

5.3.7 Examples for Kronebreen (Svalbard)

In Fig. 6 we show various examples of uncertainty assessments for Kronebreen in Svalbard (Luckman et al., 2015; Schellenberger et al., 2015). Figures 6a and 6c show flow velocities from Sentinel 2 and 1 along with correlation coefficients of the matching (IV-2) in Figs. 6b and 6d, respectively. Both images (Figs. 6a and c) depict the high velocities near the terminus and agree in the derived value of about 3 m day^{-1}. However, due to the large estimation window used for Sentinel 1 values at the calving front are underestimated. The correlation coefficients over the glacier are very high for Sentinel 2 apart from a region with a small cloud and topographic shadow (Fig. 6b). The radar image is more consistent in this regard apart from regions in radar shadow, but the correlation coefficient is generally higher over steep terrain. The stable ground measure (IV-3) revealed flow velocities of 1.2 ±0.85 m day^{-1} for Sentinel 2 and 0.05 ±0.11 m day^{-1} for Sentinel 1.

Results of a survey using two ground based radar interferometers (measures IV-5 and 6) acquired over a period of three hours on August 27, 2016 are depicted in Fig. 6e. They are thus obtained within the period used for satellite data retrieval and reveal a good match with the velocity pattern seen in Fig. 6a, even close to the calving front. Maximum values of 3 m d^{-1} are found at the same location. In Figure 6f a dense time series of flow velocities determined with Sentinel-1 along the central flow line of Kronebreen is shown starting at the top of the glacier. The very limited variability along large parts of the flow line reveal that measurements are consistent and vary only slightly (IV-4). Towards the terminus the variability increases, showing an increasing trend towards summer.

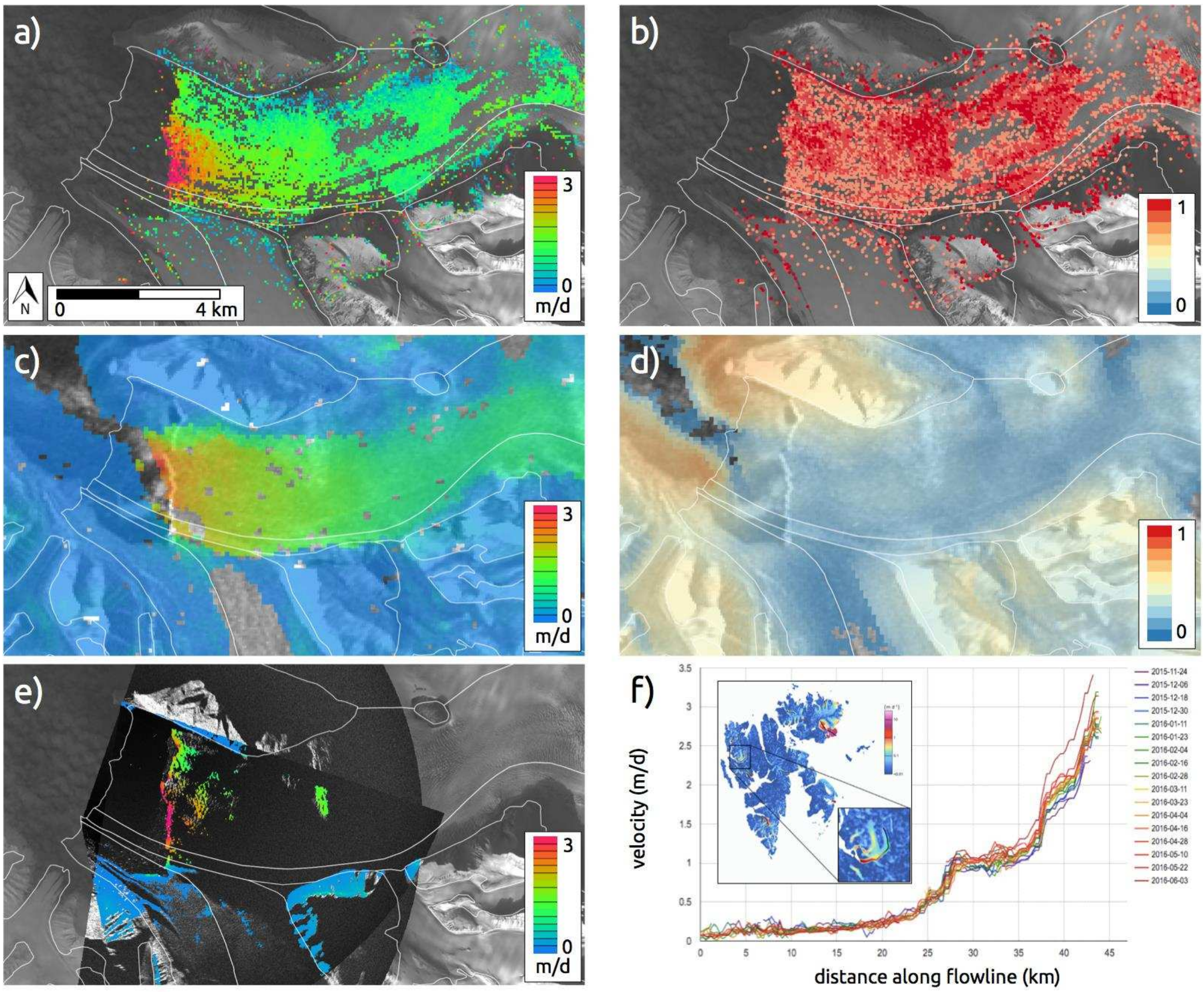

Fig. 6: Illustration of four methods used to determine accuracy for glacier velocity on the example of Kronebreen (see inset in Fig. 3f for location). a) Colour-coded flow velocities derived from a Sentinel 2 image pair acquired on 22.8. and 1.9. 2016. b) Related correlation coefficients for the image pair in a). c) As a) but with Sentinel 1 images acquired on 20.8. and 1.9. 2016. f) As in b but for the Sentinel 1 image pair used for c). e) Ground based determination of flow velocities obtained on 27.8. 2016 over three hours using the Gamma Portable Radar Interferometer (GPRI) using the same colour-coding as in a) and c). f) Multi-temporal analysis of flow velocities along the central flow line of Kronebreen. The inset shows the location of Kronebreen in Svalbard and the location of the profile line. The Svalbard map is colour-coded with flow velocities derived from Sentinel 1. The white glacier outlines are from the RGI 5.0 (source: glims.org/RGI) illustrating considerable frontal retreat until 2016.

5.4 Recommended Strategy

Level 0

Overlay of outlines: A map of the results and a comment from an experienced operator based on visual inspection of the resulting displacement field (i.e., whether the derived flow field is consistent, whether sensor effects are apparent, whether artefacts (e.g. jitter or ionosphere) are present, etc.) is important for a first order quality assessment.

Level 1

Matching CC or SNR: A map of correlation coefficients and/or signal to noise ratio values should be provided, to have an estimate of the strength of the matches behind each displacement. As noted previously, however, this is not suitable on its own to determine accuracy, as strong matches can still give erroneous displacements (and vice-versa).

Retrieval over stable ground

Statistical measurements (i.e., mean or median and standard deviation or RMSE) of the matches over stable ground should be included in the accuracy assessment. As a further quality indicator the percentage of successful matches over ice can be also provided.

Level 2

Analysis of ice velocity times series and consistency

This test is suitable for regions with a systematic acquisition of satellite images (Sentinel-1/2, Landsat 8). The test assumes that over short time intervals the ice velocity of most glaciers is stable or shows trends over several observation cycles and can thus be applied to regions with homogenous velocity. The test provides the temporal mean and standard deviation of velocity, its the temporal trend, or along selected profiles (e.g. a centre line).

Comparison of different sensors

If temporally consistent, higher-resolution images are available, the internal accuracy measurements described above can be supplemented with the deviation between the two displacement maps for the vector magnitude and direction or the vector easting, northing and vertical components. A summary of these deviations can be expressed by the mean and standard deviation (or root-mean square error) for the total number of coincident measurements.

Level 3

Validation with in-situ velocity measurements

If temporally consistent ground-based measurements of displacement are available, the deviation between product-type displacements and validation displacements gives product accuracy. A summary of these deviations can be expressed by the mean and standard deviation (or root-mean square error) for the total number of in-situ data with corresponding EO observations.

6. Summary of recommendations

We have presented methods to determine accuracy and precision of glacier area (Section 2), elevation change (Sections 3 & 4) and velocity (section 5) products based on the experiences gained in Glaciers_cci and earlier studies. We have not provided an explicit review of the literature or equations and theory on error propagation, but rather focus here on key practical issues that are relevant. For all products we identified possibilities to estimate precision using internal methods (e.g. elevation changes or flow velocities over stable ground), more laborious ones requiring extra effort (e.g. multiple manual digitization of glacier outlines), and those using reference data to also determine accuracy. Based on the various levels of complexity and workload, we have suggested for all products a tiered list of measures to guide analysts through the possibilities. We think that applying and providing the Level 0 assessments is mandatory and results from the measures at Level 1 should be provided whenever possible. The Level 2 measures already require a substantial additional workload but they are still based on internal calculations, i.e. they do not require external validation data. They often provide a more realistic measure of product precision than the measures at Level 0 and 1 and can thus be well used to determine the significance of a change. Real validation, however, can only be obtained with the measures at Level 3 that consider a comparison with appropriate reference data. For a correct result it is important to carefully remove potential biases between the two datasets that might, for example, be introduced by different spatial resolution. So far, this has rarely been done.

We are aware that there are several further factors influencing product accuracy that are not discussed here. In general, their impact on accuracy is rather small and/or requires investigations that are beyond the scope of this overview. Examples are the correction of spatial trends in elevation change, consideration of instrument jitter when calculating glacier volume changes from DEM differencing (Girod et al., 2016), or dealing with pixel shifts when processing descending and ascending orbits to estimate flow velocities. Uncertainty in the acquisition date of the DEM (e.g. national DEMs or the ASTER GDEM2) is also a factor directly impacting on the accuracy of the derived elevation change rate. Another one is the deformation of glacier outlines when an inappropriate DEM is used for orthorectification of the related satellite data. This is related to coarse resolution (e.g. using a 90 m DEM for 10 m satellite data) and the date of the DEM in relation to the image. In particular glaciers might show strong changes in elevation over a decadal period giving rise to uncertainty when an out-dated DEM is used (Kääb et al., 2016). There is thus an urgent need not only to use more appropriate DEMs for orthorectification of satellite data, but also for providing these DEMs to the community so that sub-sequent calculations (e.g. glacier drainage divides) have a good spatial match.

The results for our product examples show a general trend of reduced uncertainty (higher precision) when the more laborious, higher level measures are applied. As they might also be more realistic in regard to the dataset under consideration, they are worth the extra effort. We have not investigated here more subtle impacts on product accuracy (e.g. area in UTM projection) as well as very gross ones (e.g. removing attached snow fields) as they are highly variable and difficult to quantify. However, in general we suggest that products requiring strong interactions / editing by an analyst (such as glacier outlines) should be carefully investigated before being used for change assessment. The differences in interpretation might result in much larger changes than the real changes and be much higher than other uncertainties. Apart from the possibilities to provide quantitative numbers on product precision (and maybe accuracy), it is recommended to not forget the simplest measures (overlay of outlines or velocity vectors, visual inspection) to detect gross errors and check if results are reasonable.

Acknowledgements

This work was funded by the ESA project Glaciers_cci (4000109873/14/I-NB). CN further acknowledge support from the European Research Council under the European Union’s Seventh Framework Programme (FP/2007-2013)/ERC grant agreement no. 320816. Ground based radar velocity validation was provided through support from the Norwegian Research Council (244196/E10) and the Svalbard Science Forum. TanDEM-X Intermediate DEM was provided by DLR through proposal IDEM_GLAC0435.

References

Allstadt, K.E., Shean, D. E., Campbell, A., Fahnestock, M., & Malone, S. D. (2015). Observations of seasonal and diurnal glacier velocities at Mount Rainier, Washington, using terrestrial radar interferometry. The Cryosphere, 9, 2219-2235.

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

Ayoub, F., Leprince, S., Binety, R., Lewis, K.W., Aharonson, O. & Avouac, J.P (2008). Influence of camera distortions on satellite image registration and change detection applications. IEEE International Geoscience and Remote Sensing Symposium. Piscataway, NJ, 1072-1075, doi: 10.1109/IGARSS.2008.4779184.

Berthier, E., Arnaud, Y., Vincent, C., & Rémy, F. (2006). Biases of SRTM in high-mountain areas: Implications for the monitoring of glacier volume changes. Geophysical Research Letters, 33(8), L08502, doi: 10.1029/2006GL025862.

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

Berthier, E., Vincent, C., Magnússon, E., Gunnlaugsson, Á.Þ., Pitte, P., Le Meur, E., Masiokas, M., Ruiz, L., Pálsson, F., Belart, J.M.C., & Wagnon, P. (2014). Glacier topography and elevation changes derived from Pléiades sub-meter stereo images. The Cryosphere, 8, 2275-2291.

Berthling, I. (2011). Beyond Confusion: Rock Glaciers as Cryo-Conditioned Landforms. Geomorphology, 131, 98-106.

Bhambri, R., & Bolch, T. (2009). Glacier mapping: a review with special reference to the Indian Himalayas. Progress in Physical Geography, 33, 672-704.

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

Bolch, T., Sandberg Sørensen, L., Simonssen, S.B., Mölg, N., Machguth, H., Rastner, P., & Paul, F. (2013). Mass loss of Greenland’s glaciers and ice caps 2003-2008 revealed from ICESat laser altimetry data. Geophysical Research Letters, 40, 875-881, doi: 10.1002/grl.50270.

Borsa, A. A., Moholdt, G., Fricker, H. A., & Brunt, K. M. (2014). A range correction for ICESat and its potential impact on ice-sheet mass balance studies. The Cryosphere, 8, 345-357.

Bown, F., Rivera, A., & Acuña, C. (2008). Recent glacier variations at the Aconcagua basin, central Chilean Andes. Annals of Glaciology, 48, 43-48.

Burgess, E. W., Forster, R. R., Larsen, C. F., & Braun, M. (2012). Surge dynamics on Bering Glacier, Alaska, in 2008-2011. The Cryosphere, 6, 1251-1262.

Copland, L., Pope, S., Bishop, M., Shroder, J., Clendon, P., Bush, A., Kamp, U., Seong, Y., & Owen, L. (2009). Glacier velocities across the central Karakoram. Annals of Glaciology, 50, 41-49.

Davis, C. H. (1997). A robust threshold retracking algorithm for measuring ice-sheet surface elevation change from satellite radar altimeters. IEEE Transactions on Geoscience and Remote Sensing, 35 (4), 974-979.

Davis, C. H., Li, Y., McConnell, J. R., Frey, M. M., Hanna, E. (2005). Snowfall-driven growth in East Antarctic Ice Sheet mitigates recent sea-level rise. Science, 308 (5730), 1898 -1901.

Debella-Gilo, M., & 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.

Dehecq, A., Gourmelen, N., & Trouve, E. (2015). Deriving large-scale glacier velocities from a complete satellite archive: Application to the Pamir-Karakoram-Himalaya. Remote Sensing of Environment, 162, 55-66.

Dehecq, A., Millan, R., Berthier, E., Gourmelen, N., & Trouvé, E. (2016). Elevation changes inferred fromTanDEM-X data over the Mont-Blanc area: Impact of the X-band interferometric bias. Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9 (8), 3870-3882.

Duda, D. P., Spinhirne, J. D., & Eloranta, E. W. (2001). Atmospheric multiple scattering effects on GLAS altimetry - part I: Calculations of single pulse bias. IEEE Transactions on Geoscience and Remote Sensing, 39, 92-101.

Euillades, L. D., Euillades, P. A., Riveros, N. C., Masiokas, M. H., Ruiz, L., Pitte, P., Elefante, S., Casu F., & Balbarani, S. (2016). Detection of glaciers displacement time-series using SAR. Remote Sensing of Environment, 184, 188-198.

Fischer, M., Huss, M., Barboux, C., & Hoelzle, M. (2014). The new Swiss Glacier Inventory SGI2010: relevance of using high-resolution source data in areas dominated by very small glaciers. Arctic, Antarctic and Alpine Research, 46, 933-945.

Flament, T., & Rémy, F. (2012). Dynamic thinning of Antarctic glaciers from along-track repeat radar altimetry. Journal of Glaciology, 58 (211), 830-840.

Frey, H., Paul, F., & Strozzi, T. (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., Borsa, A., Minster, B., Carabajal, C., Quinn, K., & Bill, B. (2005). Assessment of ICESat performance at the salar de Uyuni, Bolivia. Geophysical Research Letters, 32, L21S06, doi: 10.1029/2005GL023423.

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

Gardelle, J., Berthier, E., Arnaud, Y., & Kääb, A. (2013). Region-wide glacier mass balances over the Pamir-Karakoram-Himalaya during 1999-2011. The Cryosphere, 7(4), 1263-1286.

Gardner, A. S., Moholdt, G., Cogley, J. G., Wouters, B., Arendt, A. A., Wahr, J., Berthier, E., Hock, R., Pfeffer, W. T., Kaser, G., Ligtenberg, S. R. M., Bolch, T., Sharp, M. J., Hagen, J. O., van den Broeke, M. R., & Paul, F. (2013). A reconciled estimate of glacier contributions to sea level rise: 2003 to 2009. Science, 340, 852-857.

GCOS (2006): Systematic observation requirements for satellite-based products for climate. GCOS Report 107, WMO/TD No. 1338, 103 pp.

Girod, L., Nuth, C., & Kääb, A. (2016). Glacier volume change estimation using time series of improved ASTER DEMs. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLI-B8, 489-494; doi: 10.5194/isprs-archives-XLI-B8-489-2016.

Gonzalez, J.H., Bachmann, M., Scheiber, R., & Krieger, G. K. (2010). Definition of ICESat selection criteria for their use as height references for TanDEM-X. IEEE Transactions on Geoscience and Remote Sensing, 48 (6,) 2750–2757

Granshaw, F. D. & Fountain, A. G. (2006). Glacier change (1958-1998) in the North Cascades National Park Complex, Washington, USA. Journal of Glaciology, 52, 251-256.

Gray, L., Burgess, D., Copland, L., Demuth, M. N., Dunse, T., Langley, K., & Schuler, T. V. (2015). CryoSat-2 delivers monthly and inter-annual surface elevation change for Arctic ice caps. The Cryosphere, 9, 1895-1913.

Gruber, A., Wessel, B., Huber, M., Roth, A. (2012). Operational TanDEM-X DEM calibration and first validation results. ISPRS J. Photogramm. Remote Sens., 73, 39-49.

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

Hall, D. K., Chang, A.T.C., & Siddalingaiah, H. (1988): Reflectances of glaciers as calculated using Landsat 5 Thematic Mapper data. Remote Sensing of Environment, 25, 311-321.

Hall, D.K., Bayr, K.J., Schöner, W., Bindschadler, R.A., & Chien, J.Y.L. (2003). Consideration of the errors inherent in mapping historical glacier positions in Austria from the ground and space (1893-2001). Remote Sensing of Environment, 86, 566-577.

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

Helm, V., Humbert, A., & Miller, H. (2014). Elevation and elevation change of Greenland and Antarctica derived from CryoSat-2. The Cryosphere, 8, 1539-1559.

Hilbert, C., & Schmullius, C. (2012). Influence of surface topography on ICESat/GLAS forest height estimation and waveform shape. Remote Sensing, 4(8), 2210-2235.

Holzer, N., Vijay, S., Yao, T., Xu, B., Buchroithner, M., & Bolch, T. (2015). Four decades of glacier variations at Muztagh Ata (eastern Pamir): a multi-sensor study including Hexagon KH-9 and Pléiades data. The Cryosphere, 9, 2071-2088.

Janke, J. R., Bellisario, A. C., & Ferrando, F. A. (2015). Classification of debris-covered glaciers and rock glaciers in the Andes of Central Chile. Geomorphology, 241, 98-121.

Jörg, P. C., Morsdorf, F., & Zemp, M. (2012). Uncertainty assessment of multi-temporal airborne laser scanning data: A case study at an Alpine glacier. Remote Sensing of Environment, 127, 118-129.

Jörg, P. C. & Zemp, M. (2014). Evaluating volumetric glacier change methods using airborne laser scanning data. Geografiska Annaler: Series A - Physical Geography, 96, 135-145.

Joughin, I. R., Winebrenner, D. P., Fahnestock, M. A., Kwok, R., & Krabill, W. B. (1996). Measurement of ice-sheet topgraphy using satellite-radar interferometry. Journal of Glaciology, 42(140), 10-22.

Kääb, A. (2008). Glacier volume changes using ASTER satellite stereo and ICESat GLAS laser altimetry. A test study on Edgeøya, Eastern Svalbard. IEEE Transactions on Geoscience and Remote Sensing, 46(10), 2823-2830.

Kääb, A., Berthier, E., Nuth, C., Gardelle, J., & Arnaud, Y. (2012). Contrasting patterns of early twenty-first-century glacier mass change in the Himalayas. Nature, 488, 495-498.

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

Kienholz, C., Hock, R., & Arendt, A. A. (2013). A new semi-automatic approach for dividing glacier complexes into individual glaciers. Journal of Glaciology, 59(217), 925–937.

Kohler, J., James, T. D., Murray, T., Nuth, C., Brandt, O., Barrand, N. E., et al. (2007). Acceleration in thinning rate on western Svalbard glaciers. Geophysical Research Letters, 34(18), L18502, doi: 10.1029/2007GL030681.

Korona, J., Berthier, E., Bernard, M., Rémy, F., & Thouvenot, E. (2009). SPIRIT. SPOT 5 stereoscopic survey of Polar Ice: Reference Images and Topographies during the fourth International Polar Year (2007-2009). ISPRS Journal of Photogrammetry and Remote Sensing, 64, 204-212.

Kronenberg, M., Barandun, M., Hoelzle, M., Huss, M., Farinotti, D., Azisov, E., et al. (2016). Mass-balance reconstruction for Glacier No. 354, Tien Shan, from 2003 to 2014. Annals of Glaciology, 57(71), 92-102.

Kropáček, J., Neckel, N., & Bauder, A. (2014): Estimation of mass balance of the Grosser Aletschgletscher, Swiss Alps, from ICESat laser altimetry data and digital elevation models. Remote Sensing, 6(6), 5614-5632; doi:10.3390/rs6065614.

Lambrecht, A., & Kuhn, M. (2007). Glacier changes in the Austrian Alps during the last three decades, derived from the new Austrian glacier inventory. Annals of Glaciology, 46, 177-184.

Le Bris, R. & Paul, F. (2015). Glacier-specific elevation changes in western Alaska. Annals of Glaciology, 56 (70), 184-192.

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

Luckman, A., Benn, D. I., Cottier, F., Bevan, S., Nilsen, F., & Inall, M. (2015). Calving rates at tidewater glaciers vary strongly with ocean temperature. Nature Communication, 6, 8566, doi: 10.1038/ncomms9566.

Malenovsky, Z., Rott, H., Cihlar, J., Schaepman, M. E., García-Santos, G., Fernandes, R. and Berger, M. (2012). Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. Remote Sensing of Environment, 120, 91-101.

McMillan, M., Shepherd, A., Sundal, A., Briggs, K., Muir, A., Ridout, A., Hogg, A., & Wingham, D. (2014). Increased ice losses from Antarctica detected by CryoSat-2. Geophysical Research Letters, 41 (11), 3899-3905, doi:10.1002/2014GL060111.

McMillan, M., Leeson, A., Shepherd, A., Briggs, K., Armitage, T. W. K., Hogg, A., Kuipers Munneke, P., van den Broeke, M., Noël, B., van de Berg, W., Ligtenberg, S., Horwath, M., Groh, A. , Muir, A., & Gilbert, L. (2016). A high resolution record of Greenland Mass Balance. Geophysical Research Letters. 43 (13), 7002-7010; doi:10.1002/2016GL069666.

Melkonian, A. K., Willis, M. J., Pritchard, M. E., Rivera, A., Bown, F., & Bernstein, S. A. (2013). Satellite-derived volume loss rates and glacier speeds for the Cordillera Darwin Icefield, Chile. The Cryosphere., 7(3), 823-839.

Melkonian, A. K., Willis, M. J., & Pritchard, M. E. (2014). Satellite-derived volume loss rates and glacier speeds for the Juneau Icefield, Alaska. Journal of Glaciology, 60(222), 743-760.

Menditto, A., Patriarca, M., & Magnusson, B. (2007). Understanding the meaning of accuracy, trueness and precision. Accreditation and Quality Assurance, 12, 45-47.

Merchant, C. J., Paul, F., Popp, T., Ablain, M., Bontemps, S., Defourny, P., Hollmann, R., Lavergne, T., Laeng, A., de Leeuw, G., Mittaz, J., Poulsen, C., Povey, A. C., Reuter, M., Sathyendranath, S., Sandven, S., Sofieva, V. F., & Wagner, W. (2017). Uncertainty information in climate data records from Earth observation. Earth System Science Data, 9, 511-527.

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

Müller, J., Vieli, A., & Gärtner-Roer, I. (2016). Rock glaciers on the run – understanding rock glacier landform evolution and recent changes from numerical flow modeling, The Cryosphere, 10, 2865-2886.

Nagai, H., Fujita, K., Sakai, A., Nuimura, T., & Tadono, T. (2016). Comparison of multiple glacier inventories with a new inventory derived from high-resolution ALOS imagery in the Bhutan Himalaya. The Cryosphere, 10, 65-85.

Nagler, T., Rott, H., Hetzenecker, M., Wuite, J., & Potin, P. (2015). The Sentinel-1 Mission: New Opportunities for Ice Sheet Observations. Remote Sensing, 2015, 7, 9371-9389, doi:10.3390/ rs70709371.

Neckel, N., Braun, A., Kropáček, J., & Hochschild, V. (2013). Recent mass balance of the Purogangri Ice Cap, central Tibetan Plateau, by means of differential X-band SAR interferometry. The Cryosphere, 7,

Nilsson, J., Sandberg-Sorensen, L., Barletta, V. R., & Forsberg, R. (2015). Mass changes in Arctic ice caps and glaciers: implications of regionalizing elevation changes. The Cryosphere. 9, 139-150.

Nilsson, J., Gardner, A., Sandberg Sørensen, L., & Forsberg, R. (2016). Improved retrieval of land ice topography from CryoSat-2 data and its impact for volume-change estimation of the Greenland Ice Sheet, The Cryosphere, 10(6), 2953-2969.

Nuimura, T., Sakai, A., Taniguchi, K., Nagai, H., Lamsal, D., Tsutaki, S., Kozawa, A., Hoshina, Y., Takenaka, S., Omiya, S., Tsunematsu, K., Tshering, P., & Fujita, K. (2015). The GAMDAM glacier inventory: a quality-controlled inventory of Asian glaciers. The Cryosphere, 9, 849-864.

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

Østrem, G. (1971). Rock glaciers and ice-cored moraines, a reply to D. Barsch. Geografiska Annaler Series A - Physical Geography, 53, 207–213.

Paul, F. (2008). Calculation of glacier elevation changes with SRTM: Is there an elevation-dependent bias? Journal of Glaciology, 54(188), 945-946.

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

Paul, F., Huggel, C., Kääb, A. , & 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., Andreassen, L. M., & 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.

Paul, F., Barrand, N., Berthier, E., Bolch, T., Casey, K., Frey, H., Joshi, S. P., Konovalov, V., Le Bris, R., Mölg, N., Nosenko, G., Nuth, C., Pope, A., Racoviteanu, A., Rastner, P., Raup, B.H., Scharrer, K., Steffen, S., & Winsvold, S.H. (2013). On the accuracy of glacier outlines derived from remote sensing data. Annals of Glaciology, 54 (63), 171-182.

Paul, F., Bolch, T., Kääb, A., Nagler, T., Nuth, C., Scharrer, K., et al. (2015). The glaciers climate change initiative: Methods for creating glacier area, elevation change and velocity products. Remote Sensing of Environment, 162, 408-426.

Paul, F., Winsvold, S.H., Kääb, A., Nagler, T., & Schwaizer, G. (2016). Glacier Remote Sensing Using Sentinel-2. Part II: Mapping Glacier Extents and Surface Facies, and Comparison to Landsat 8. Remote Sensing, 8(7), 575; doi:10.3390/rs8070575.

Peipe, J., Reiss, P., & Rentsch, H. (1978). Zur Anwendung des digitalen Geländemodells in der Gletscherforschung. Zeitschrift fur Gletscherkunde und Glazialgeologie, 14(2), 161-172.

Pope, A., Rees, W. G., Fox, A. J., & Fleming, A. (2014). Open Access Data in Polar and Cryospheric Remote Sensing. Remote Sensing, 6, 6183-6220.

Pfeffer, W. T., Arendt, A. A., Bliss, A., Bolch, T., Cogley, J. G., Gardner, A. S., Hagen, J.-O., Hock, R., Kaser, G., Kienholz, . C., Miles, E. S., Moholdt, G., Mölg, N., Paul, . F., Radic, V., Rastner, P., Raup, B. H., Rich, J., Sharp, M.J. & the Randolph Consortium (2014). The Randolph Glacier Inventory: a globally complete inventory of glaciers. Journal of Glaciology, 60 (221), 537-552.

Pieczonka, T., & Bolch, T. (2015). Region-wide glacier mass budgets and area changes for the Central Tien Shan between ~1975 and 1999 using Hexagon KH-9 imagery. Global and Planetary Change 128, 1-13.

Quincey, D., Luckman, A., & Benn, D. (2009). Quantification of Everest region glacier velocities between 1992 and 2002, using satellite radar interferometry and feature tracking. Journal of Glaciology, 55(192), 596-606.

Racoviteanu, A. E., Manley, W.F., Arnaud, Y., & Williams M. W. (2007). Evaluating digital elevation models for glaciologic applications: An example from Nevado Coropuna, Peruvian Andes. Global and Planetary Change, 59, 110-125.

Racoviteanu, A. E, Paul, F., Raup, B., Khalsa, S. J. S., & Armstrong, R. (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.

Rankl, M. & Braun, M. (2016). Glacier elevation and mass changes over the central Karakoram region estimated from TanDEM-X and SRTM/X-SAR digital elevation models. Annals of Glaciology 51(71), 273-281.

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

Rastner, P., Bolch, T., Notarnicola, C., & Paul, F. (2014). A comparison of pixel- and object-based glacier classification with optical satellite images. Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7 (3), 853-862.

Raup, B. H., & Khalsa, S. J. S. (2007). GLIMS Analysis Tutorial, vers. 22/05/2007. Boulder. http://glims.org/MapsAndDocs/assets/GLIMS_Analysis_Tutorial_a4.pdf.

Raup, B. H., Kääb, A., Kargel, J. S., Bishop, M. P., Hamilton, G., Lee, E., Paul, F., Rau, F., Soltesz, D., Khalsa, S. J. S., Beedle, M. and Helm, C. (2007): Remote sensing and GIS technology in the Global Land Ice Measurements from Space (GLIMS) Project. Computers and Geosciences, 33, 104-125.

Raup, B., Khalsa, S. J. S., Armstrong, R., Sneed, W., Hamilton, Paul, F., Cawkwell, F., Beedle, M., Menounos, B., Wheate, R., Rott, H., Liu, S., Li, X., Shangguan, D., Cheng, Kargel, J., Larsen, Molnia, B., Kincaid, J., Klein, A. and Konovalov, V. (2014): Quality in the GLIMS Glacier Database. In: Kargel, J.S., Bishop, M.P., Kääb, A. and Raup, B.H. (Eds.): Global Land Ice Measurements from Space - Satellite Multispectral Imaging of Glaciers. Praxis-Springer, Chapter 7, 163-180.

Reinhardt, W., & Rentsch, H. (1986): Determination of changes in volume and elevation of glaciers using digital elevation models for the Vernagtferner, Oetztal Alps, Austria. Annals of Glaciology, 8, 151-155.

Rignot, E., Echelmeyer, K., & Krabill, W. (2001). Penetration depth of interferometric synthetic-aperture radar signals in snow and ice. Geophysical Research Letters, 28(18), 3501-3504.

Rignot, E. J., Rivera, A., & Casassa, G. (2003). Contribution of the Patagonia Icefields of South America to Sea Level Rise. Science, 302(5644), 434-437.

Rinne, E., Shepherd, A., Muir, A., & Wingham, D. (2011a). A Comparison of Recent Elevation Change Estimates of the Devon Ice Cap as Measured by the ICESat and EnviSAT Satellite Altimeters. IEEE Transactions on Geoscience and Remote Sensing, 49, 1902-1910.

Rinne, E., Shepherd, A., Palmer, S., van den Broeke, M., Muir, A., Ettema, J., & Wingham, D. (2011b). On the recent elevation changes at the Flade Isblink Ice Cap, northern Greenland. Journal of Geophysical Research, 116, F03024, doi: 10.1029/2011JF001972.

Schellenberger, T., Dunse, T., Kääb, A., Kohler, J., & Reijmer, C. H. (2015). Surface speed and frontal ablation of Kronebreen and Kongsbreen, NW Svalbard, from SAR offset tracking. The Cryosphere, 9, 2339-2355.

Schellenberger, T., Van Wychen, W., Copland, L., Kääb, A., & Gray, L. (2016). An inter-comparison of techniques for determining velocities of maritime Arctic glaciers, Svalbard, using Radarsat-2 Wide Fine Mode data. Remote Sensing, 8(9), 785.

Scherler, D., Leprince, S., & Strecker, M. R. (2008). Glacier surface velocities in alpine terrain from optical satellite imagery - Accuracy improvement and quality assessment. Remote Sensing of Environment, 112 (10), 3806-3819.

Schiefer, E., Menounos, B., & Wheate, R. (2007). Recent volume loss of British Columbian glaciers, Canada. Geophysical Research Letters, 34(16), L16503 doi: 10.1029/2007GL030780.

Scott, J. B. T., Nienow, P., Mair, D., Parry, V., Morris, E., & Wingham, D.J. (2006). Importance of seasonal and annual layers in controlling backscatter to radar altimeters across the percolation zone of an ice sheet, Geophysical Research Letters, 33, L24502, doi: 10.1029/2006GL027974.

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

Shugar, D. H., Rabus, B. T., & Clague, J. J. (2010). Elevation changes (1949 - 1995) of Black Rapids Glacier, Alaska, derived from a multi-baseline InSAR DEM and historical maps. Journal of Glaciology, 56(198), 625-634.

Shuman, C. A., Zwally, H. J., Schutz, B. E., Brenner, A. C., DiMarzio, J. P., Suchdeo, V. P., & Fricker, H. A. (2006): ICESat Antarctic elevation data: Preliminary precision and accuracy assessment. Geophysical Research Letters, 33(7), L07501, doi: 10.1029/2005GL025227.

Skvarca, P., Raup, B. H., & De Angelis, H. (2003): Recent behaviour of Glaciar Upsala, a fast-flowing calving glacier in Lago Argentino, southern Patagonia. Annals of Glaciology, 36, 184-188.

Sørensen, L. S., Simonsen, S. B., Nielsen, K., Lucas-Picher, P., Spada, G., Adalgeirsdottir, G., Forsberg, R., & Hvidberg, C. S. (2011). Mass balance of the Greenland ice sheet (2003-2008) from ICESat data - the impact of interpolation, sampling and firn density. The Cryosphere, 5, 173-186.

Strozzi, T., Luckman, A., Murray, T., Wegmüller, U., & Werner, C. (2002). Glacier motion estimation using SAR Offset-Tracking procedures. IEEE Transactions On Geoscience and Remote Sensing., 40(11), 2384-2391.

Strozzi, T., Kouraev, A., Wiesmann, A., Wegmüller, U., Sharov, A. & Werner, C. (2008). Estimation of Arctic glacier motion with satellite L-band SAR data. Remote Sensing of Environment, 112, 636-645.

Surazakov, A. B., & Aizen, V. B. (2006). Estimating volume change of mountain glaciers using SRTM and map-based topographic data. IEEE Transactions on Geoscience and Remote Sensing, 44(10), 2991-2995.

Trantow, T., & Herzfeld, U. C. (2016). Spatiotemporal mapping of a large mountain glacier from CryoSat-2 altimeter data: surface elevation and elevation change of Bering Glacier during surge (2011-2014). International Journal of Remote Sensing, 37, 2962-2989.

Treichler, D., & Kääb, A. (2016). ICESat laser altimetry over small mountain glaciers. The Cryosphere, 10, 2129-2146.

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, 26, 167-181.

Vaughan, D.G., Comiso, J. C., Allison, I., Carrasco, J., Kaser, G., Kwok, R., Mote, P., Murray, T., Paul, F., Ren, J., Rignot, E., Solomina, O., Steffen, K., & Zhang, T. (2014): Observations: Cryosphere. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the IPCC. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 317-382.

Vieli, A., Jania, J., Blatter, H., & Funk, M. (2004). Short-term velocity variations on Hansbreen, a tidewater glacier in Spitsbergen. Journal of Glaciology, 50(170), 389-398.

Wang, D. & Kääb, A. (2015). Modeling glacier elevation change from DEM time series. Remote Sensing, 7, 10117-10142.

Wang, X., Cheng, X., Gong, P., Huang,H., Li, Z., & Li, X. (2011). Earth science applications of ICESat/GLAS: a review. International Journal of Remote Sensing, 32 (23), 8837-8864.

Wegmüller, U., Werner, C., Strozzi, T., & Wiesmann, A. (2006). Ionospheric electron concentration effects on SAR and INSAR. IEEE International Symposium on Geoscience and Remote Sensing IGARSS, 3731-3734.

Willis, M. J., Melkonian, A. K., Pritchard, M. E., & Rivera, A. (2012). Ice loss from the Southern Patagonian Ice Field, South America, between 2000 and 2012. Geophysical Research Letters, 39(17), L17501, doi: 10.1029/2012GL053136.

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

Winsvold, S. H., Kääb A., & Nuth C. (2016). Regional glacier mapping using optical satellite data time series. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(8), 3698-3711.

Zemp, M., Thibert, E., Huss, M., Stumm, D., Rolstad Denby, C., Nuth, C., Nussbaumer, S. U., Moholdt, G., Mercer, A., Mayer, C., Joerg, P. C., Jansson, P., Hynek, B., Fischer, A., Escher-

Vetter, H., Elvehøy, H., and Andreassen, L. M. (2013): Reanalysing glacier mass balance measurement series. The Cryosphere, 7, 1227-1245.

Zemp, M., et al. (2015). Historically unprecedented global glacier decline in the early 21st century. Journal of Glaciology, 61 (228), 745-762.

Zwally H.J., Schutz, B., Abdalati, W., Abshire, J., Bentley, C., Brenner, A., Bufton, J., Dezio, J., Hancock, D., Harding, D., Herring, T., Minster, B., Quinn, K., Palm, S., Spinhirne, J., & Thomas R. (2002). ICESat's laser measurements of polar ice, atmosphere, ocean, and land. Journal of Geodynamics, 34, 405-445.

Figure Captions

Fig. 1: The two false colour Landsat images (path-row: 147-031) in the top row cover the region around North and South Inylcheck Glacier in the central Tien Shan (see blue square in inset map for location) and show clouds (white) at different locations (ice and snow in shades of blue-green). They were acquired on a) 21.08.2006 and b) 24.08.2007. c) The digital combination of the classified glacier maps (2006: grey/blue, 2007: grey/red) allows creating a near complete glacier coverage. Inset map: screen shot from Google Earth, Landsat images: USGS/NASA.

Fig. 2: The region around Baspa Glacier at the headwater of the Baspa river basin (see blue square in inset map for location) as seen on two false colour Landsat images (path-row: 146-038) acquired on a) 20. Aug. 2014 and b) 10. Sep. 2016. Although a) looks usable for glacier mapping at first sight, it suffers from abundant seasonal snow (circle) and avalanche cones hiding glacier parameters. In b) snow outside of glaciers has largely disappeared and glacier mapping is much more easy. However, some clouds are now hiding some of the glaciers and need to be mapped by other scenes (see Fig. 1). Inset map: screen shot from Google Earth, Landsat images: USGS/NASA.

Fig. 3: a) Glaciers, debris-covered ice, rock glaciers, ice-cored moraines and other periglacial features in a small catchment of the Baspa basin (see inset for location). In this region the glacier terminus is clearly defined, but the other marked periglacial landforms containing ice are based on

subjective interpretation. b) A small cirque glacier (upper right) that continuously evolves into a debris-covered glacier and a rock glacier with its steep front in the lower left (there is a further rock glacier to the right). In this case several possibilities exist to assign a glacier terminus (indicated by the transition zone). Images and inset map: Screen shots from Google Earth, (C) 2017 CNES / Airbus.

Fig. 4: Illustration of three methods used to determine uncertainty for glacier outlines. a) Location of the study glaciers in Austria (the main image is a screenshot from Google Earth), b) buffer method GO-3 (±1/2 pixel) illustrated for the smaller glacier, c) multiple digitizing (GO-6) for the glacier in b), and d) comparison to a reference area (GO-7) for the glacier in b). Panels b) and c) are based on 30 m Landsat images whereas d) is from Quickbird (screenshot from Google Earth). The white bar measures 100 m, North is up.

Fig. 5: Illustration of elevation differences on stable terrain and glaciers between a) 1990 and 2007 and b) 2007 and 2010 for Kronebreen in Svalbard (see red square on the inset for location). c) Elevation difference histograms for stable terrain and glacier ice. d) Elevation change centreline profiles along Kronebreen for both epochs, revealing higher loss rates near the terminus in the more recent period.

Fig. 6: Illustration of four methods used to determine accuracy for glacier velocity on the example of Kronebreen (see inset in Fig. 3f for location). a) Colour-coded flow velocities derived from a Sentinel 2 image pair acquired on 22.8. and 1.9. 2016. b) Related correlation coefficients for the image pair in a). c) As a) but with Sentinel 1 images acquired on 20.8. and 1.9. 2016. f) As in b but for the Sentinel 1 image pair used for c). e) Ground based determination of flow velocities obtained on 27.8. 2016 over three hours using the Gamma Portable Radar Interferometer (GPRI) using the same colour-coding as in a) and c). f) Multi-temporal analysis of flow velocities along the central flow line of Kronebreen. The inset shows the location of Kronebreen in Svalbard and the location of the profile line. The Svalbard map is colour-coded with flow velocities derived from Sentinel 1. The white glacier

outlines are from the RGI 5.0 (source: glims.org/RGI) illustrating considerable frontal retreat until 2016.