ARTICLE IN PRESS

Remote Sensing of Environment xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

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[a,*], Tobias Bolch[a], Kate Briggs[b], Andreas Kääb[c], Malcolm McMillan[b], Robert McNabb[c], Thomas Nagler[d], Christopher Nuth[c], Philipp Rastner[a], Tazio Strozzi[e], Jan Wuite[d]

[a] Department of Geography, University of Zurich, Winterthurerstr. 190, 8057 Zurich, Switzerland
[b] Centre for Polar Observation and Modelling, University of Leeds, Leeds LS2 9JT, UK
[c] Institute of Geosciences, University of Oslo, P.O. Box 1047, 0316 Oslo, Norway
[d] ENVEO IT GmbH, Technikerstrasse 21a, 6020 Innsbruck, Austria
[e] GAMMA Remote Sensing, Worbstr. 225, 3073 Gümligen, Switzerland

ARTICLE INFO

Keywords:
Glaciers
Outlines
Elevation change
Velocity
Quality measures
ECV monitoring
Satellite data
CCI

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

* Corresponding author.
E-mail address: frank.paul@geo.uzh.ch (F. Paul).

http://dx.doi.org/10.1016/j.rse.2017.08.038
Received 7 November 2016; Received in revised form 8 August 2017; Accepted 31 August 2017

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)

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.

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.

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 here provide a systematic overview on the determination of product accuracy and precision for each of the four products glacier area (outlines) in Section 2, elevation changes from altimetry (Section 3) and DEM differencing (Section 4), and velocity from space borne optical sensors and Synthetic Aperture Radar (SAR) using offset tracking in Section 5 (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 sometimes also 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