Snow depth mapping from stereo satellite imagery in mountainous terrain: evaluation using airborne laser-scanning data

César Deschamps-Berger, Simon Gascoin, Etienne Berthier, Jeffrey Deems, Ethan Gutmann, Amaury Dehecq, David Shean, Marie Dumont

HAL Id: hal-02965637
https://hal.inrae.fr/hal-02965637

Submitted on 13 Oct 2020

The Cryosphere

Snow depth mapping from stereo satellite imagery in mountainous terrain: evaluation using airborne laser-scanning data

César Deschamps-Berger[1,2], **Simon Gascoin**[1], **Etienne Berthier**[3], **Jeffrey Deems**[4], **Ethan Gutmann**[5], **Amaury Dehecq**[6,7], **David Shean**[8], **and Marie Dumont**[2]

[1]Centre d'Etudes Spatiales de la Biosphère, CESBIO, Univ. Toulouse, CNES/CNRS/INRA/IRD/UPS, 31401 Toulouse, France
[2]Université Grenoble Alpes, Université de Toulouse, Météo-France, CNRS, CNRM, Centre d'Etudes de la Neige, 38000 Grenoble, France
[3]Centre National de la Recherche Scientifique (CNRS-LEGOS), 31400 Toulouse, France
[4]National Snow and Ice Data Center, Boulder, CO, USA
[5]Research Applications Lab, National Center for Atmospheric Research (NCAR), Boulder, CO, USA
[6]Laboratory of Hydraulics, Hydrology and Glaciology (VAW), ETH Zurich, Zurich, Switzerland
[7]Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland
[8]Dept. of Civil and Environmental Engineering, University of Washington, Seattle, WA, USA

Correspondence: César Deschamps-Berger (cesar.deschamps-berger@cesbio.cnes.fr)

Abstract. Accurate knowledge of snow depth distributions in mountain catchments is critical for applications in hydrology and ecology. Recently, a method was proposed to map snow depth at meter-scale resolution from very-high-resolution stereo satellite imagery (e.g., Pléiades) with an accuracy close to 0.5 m. However, the validation was limited to probe measurements and unmanned aircraft vehicle (UAV) photogrammetry, which sampled a limited fraction of the topographic and snow depth variability. We improve upon this evaluation using accurate maps of the snow depth derived from Airborne Snow Observatory laser-scanning measurements in the Tuolumne river basin, USA. We find a good agreement between both datasets over a snow-covered area of 138 km^2 on a 3 m grid, with a positive bias for a Pléiades snow depth of 0.08 m, a root mean square error of 0.80 m and a normalized median absolute deviation (NMAD) of 0.69 m. Satellite data capture the relationship between snow depth and elevation at the catchment scale and also small-scale features like snow drifts and avalanche deposits at a typical scale of tens of meters. The random error at the pixel level is lower in snow-free areas than in snow-covered areas, but it is reduced by a factor of 2 (NMAD of approximately 0.40 m for snow depth) when averaged to a 36 m grid. We conclude that satellite photogrammetry stands out as a convenient method to estimate the spatial distribution of snow depth in high mountain catchments.

1 Introduction

The snow depth or height of the snowpack (symbol: HS; Fierz et al., 2009) is a key variable for both water resource management and avalanche forecasting in mountain regions. However, determination of the spatial distribution of HS in complex terrain remains challenging due to its high spatial variability at horizontal scales below 100 m (Deems et al., 2006; Fassnacht and Deems, 2006). Current approaches to map HS are based on either sparse in situ measurements (Lopez-Moreno et al., 2011; Sturm et al., 2018), area-limited unmanned aircraft vehicle (UAV) campaigns (Bühler et al., 2016; De Michele et al., 2016; Harder et al., 2016; Redpath et al., 2018), terrestrial laser-scanning (Prokop et al., 2008; Fey et al., 2019) or costly airborne campaigns (Bühler et al., 2015; Dozier et al., 2016; Painter et al., 2016).

Recently a method was introduced to retrieve HS maps from satellite data at meter-scale resolution, typically 1 to

4 m (Marti et al., 2016; McGrath et al., 2019; Shaw et al., 2019). The method is based on the differencing of snow-on (winter) and snow-off (in general end of summer) digital elevation models (DEMs) that are generated from very-high-resolution satellite stereo imagery (e.g., Pléiades; Digital-Globe/Maxar WorldView-1, WorldView-2 and WorldView-3; and GeoEye-1). The method was first tested using two Pléiades stereo triplets over the Bassiès catchment in the Pyrenees (Marti et al., 2016). The snow-on and snow-off DEMs were generated using the Ames Stereo Pipeline (ASP; Shean et al., 2016; Beyer et al., 2018) and coregistered before differencing (Berthier et al., 2007). The accuracy of the method was evaluated using 451 probe measurements of snow depth. The HS satellite-derived map was also compared to one obtained from a UAV photogrammetric survey over a small portion of the catchment (3.1 km^2). The results showed that snow depth could be retrieved from Pléiades images with an accuracy of roughly ~ 0.5 m (standard deviation of residuals of 0.58 m for a pixel size of 2 m), suggesting that the method had the potential to become a viable alternative to airborne campaigns in mountain catchments with the benefits of a space-based platform: access to any point on the globe and lower cost for the end user. HS maps from stereo satellite images were evaluated in two recent studies besides Marti et al. (2016) against terrestrial laser-scanning data over a small area (< 1 km^2; Shaw et al., 2019) and against ground-penetrating radar measurements, which were limited to roughly 50 km^2 of relatively flat terrain (McGrath et al., 2019).

However, these works provided only a partial validation of the method since the reference data did not homogeneously sample the topographic and HS variability of the study area. For example, in Marti et al. (2016), accumulation due to snow drifts on the lee side of high-elevation ridges was not surveyed for safety reasons. The sampling depth was also limited to 3.2 m, which was the length of the snow probes. Furthermore, the areas with steep slopes were undersampled. Half of the points sampled in the field were on slopes lower than 10°, while the median terrain slope in this catchment is $\sim 30°$. This lack of validation data in steep slope areas was an important limitation of this study since DEMs from stereoscopic images are known to be less accurate on steep slopes due to a higher sensitivity to horizontal error and to local image distortion (Lacroix, 2016; Shean et al., 2016). In addition, snow probe measurements may fail to represent the mean HS at a scale of a 2 m pixel, especially in mountain terrain (Fassnacht et al., 2018). Furthermore, the impact of the photogrammetric software configuration on the accuracy of the HS map has never been evaluated. The semiglobal matching algorithm (Hirschmüller, 2005) was, for instance, added to the catalog of algorithms that can be used to derive the disparity map from stereo images in the ASP and has never been used with satellite images to derive HS. This algorithm is expected to perform better in low-texture terrain (Bühler et al., 2015; Shean et al., 2016; Harder et al., 2016; Beyer et al.,

2018) and therefore has the potential to reduce the number of missing values in the snow depth map.

Given the aforementioned limitations, we present a more comprehensive validation study by taking advantage of the NASA Airborne Snow Observatory (ASO) campaigns in the Sierra Nevada, USA. In this area ASO routinely acquires HS measurements by airborne laser scanning (ALS). We used two Pléiades stereo triplets over the Tuolumne river basin (snow-on and snow-off). The snow-on triplet was acquired on 1 May 2017, the day before the ASO flight, close to the accumulation peak. The ASO product was used as a reference as it should exhibit no bias and was found to have an accuracy roughly an order of magnitude better than Pléiades HS maps (Painter et al., 2016). We use it to test the impact of the DEM processing options, the stereo image acquisition geometry and the HS map resolution on the accuracy of HS. In addition we are able to evaluate an error model (Rolstad et al., 2009), which would enable us to calculate the error of Pléiades HS maps in other study areas where no reference datasets are available.

2 Study site

The study site is located in the Tuolumne river basin in the Sierra Nevada mountain range, California, USA (Fig. 1). The Tuolumne river supplies water to the agricultural plain of the Great Valley and the densely populated area of San Francisco. The region recently experienced a 5-year drought from 2012 to 2016 (Roche et al., 2019), increasing the interest for water resource monitoring. The ASO flights cover 1100 km^2 in the basin, while this study focuses on a 280 km^2 subzone, which was selected to cover a large elevation range. The elevation within this subzone ranges from 1800 to 3500 m a.s.l. Typical winter accumulation can reach several meters at high elevations (Painter et al., 2016). The 2016–2017 winter was characterized by near-record snow accumulation that has been referred to as the "snowpocalypse" (Painter et al., 2017).

3 Data

3.1 Pléiades images

The study area is too large to be imaged by Pléiades in tri-stereo mode with a single scene; therefore the area was imaged in two strips, which overlapped by 3 km in winter and 1.5 km in summer in the along-track direction. The snow-on triplets were both acquired on 1 May 2017, while the snow-off triplets were acquired on 8 and 13 August 2017 (Fig. 1, Table 1, image ID in Table S1). The imaged area covers 280 km^2 in total. Images were acquired in panchromatic and multispectral mode, with incidence angles along-track between -7 and 9°. The base to height (B / H) ratio of successive pairs is around 0.1 (Table 1). The panchromatic
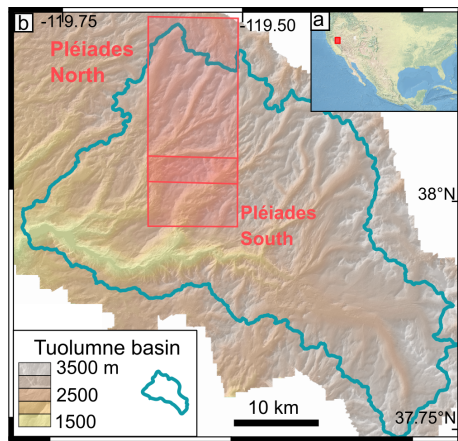
Figure 1. The Tuolumne basin is located in California, USA **(a)**. Pléiades image footprint (red polygon) in the Tuolumne basin (blue line) **(b)**. The terrain elevation in the background is the snow-off digital elevation terrain from ASO used in the coregistration step.

images have a resolution of 0.5 m at nadir and are used to calculate the DEMs. For the snow-on acquisition, we requested to reduce the number of time domain integration (TDI) lines used to image the scene. This is recommended to curb image saturation over sun-exposed snow surfaces (Berthier et al., 2014). As a result, there are no saturated pixels in the images of this study. Pléiades multispectral images have a resolution of 2 m. We only use the multispectral image that was acquired closest to the nadir view angle to compute the multispectral orthoimage. Pléiades images were obtained at no cost through the DINAMIS program (only for French scientists; see https://dinamis.teledetection.fr/, last access: 1 May 2020). It is also open to European scientists working at public research institutions. Otherwise Pléiades images can be ordered from Airbus Defense and Space.

3.2 ALS data from the Airborne Snow Observatory (ASO)

A snow-off DEM on 13 October 2015 and a snow depth map on 2 May 2017 from the ASO are used for comparison with the Pléiades products (Fig. 1, Table 1). The ASO program, operating since 2012, provides snow depth, snow water equivalent (SWE) and snow albedo maps over full mountain watersheds to support scientific campaigns and operational water management (Painter et al., 2016). The ASO laser-scanning system measures the distance between the target and aircraft and is combined with aircraft position and orientation measurements to generate a collection of reflection points: a "point cloud". Ground points are aggregated to a 3 m grid to derive a gridded DEM (Painter et al., 2015). Snow depth maps are obtained from the difference of a snow-on and snow-off DEM in unforested areas. The values in the snow-free areas are used to bias-correct the snow-on elevations and are set to zero. From comparison with 80 in situ
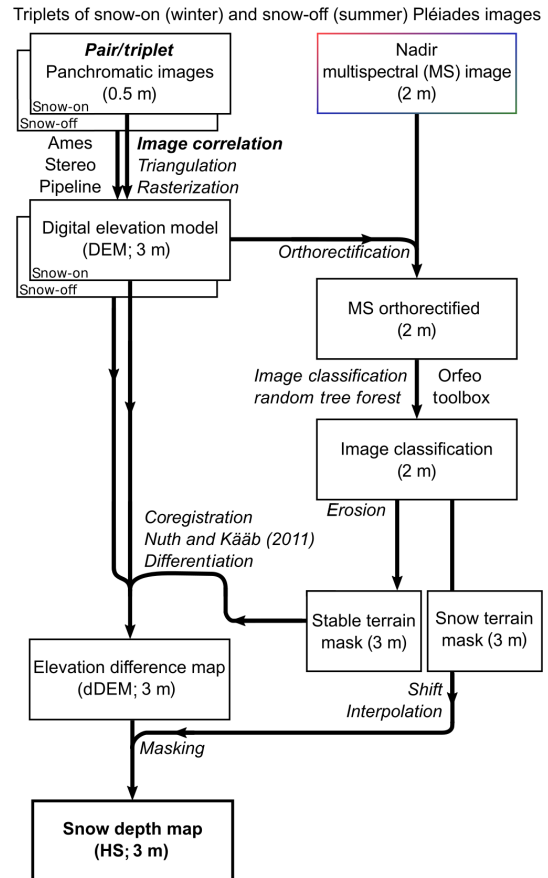

Figure 2. Workflow for the processing of the panchromatic and multispectral Pléiades images. Intermediate products are in the boxes, while the processing steps are in italics between the boxes. Text in bold italic characters indicates steps for which we tested different options.

manual measurements, no bias is observed on the HS maps, and the root mean square error (RMSE) per pixel at a 3 m resolution is 0.08 m (Painter et al., 2016). For the evaluation of the Pléiades HS maps, we excluded 25 km^2 near the catchment divide in the northeast part of the study area because we observed artifacts in the ASO HS map probably due to issues with the aircraft position and orientation data.

4 Methods

4.1 Workflow for calculation of Pléiades snow depth maps

Figure 2 presents the workflow we developed to produce HS maps from Pléiades images using the ASP version 2.6.2 (Shean et al., 2016; Beyer et al., 2018) and the Orfeo Tool-Box (Grizonnet et al., 2017). We detail below the calculation of the DEMs, the HS maps and the land cover classifications.

Table 1. Summary of the data used in this study. The base-to-height ratio (B / H) between the front–nadir (F–N), nadir–back (N–B) and front–back (F–B) pair of images is given for the stereo images.

Type	Source	Zone	Date	Horizontal resolution	B / H (F–N|N–B|F–B)	Snow-on/snow-off
Digital terrain model	Airborne laser scanning (ASO)	North + south	13 October 2015	3.0 m	–	Off
Snow depth map	Airborne laser scanning (ASO)	North+south (minus 25 km^2)	2 May 2017	3.0 m	–	On
Tri-stereo images	Satellite optical images (Pléiades)	South	1 May 2017	PAN: 0.5 m MS: 2.0 m	0.12|0.12|0.23	On
Tri-stereo images	Satellite optical images (Pléiades)	North	1 May 2017	PAN: 0.5 m MS: 2.0 m	0.12|0.12|0.23	On
Tri-stereo images	Satellite optical images (Pléiades)	South	8 August 2017	PAN: 0.5 m MS: 2.0 m	0.12|0.08|0.20	Off
Tri-stereo images	Satellite optical images (Pléiades)	North	13 August 2017	PAN: 0.5 m MS: 2.0 m	0.11|0.11|0.22	Off

4.1.1 DEM calculation

A DEM is computed with the Ames Stereo Pipeline (ASP) using two utilities: *stereo* and *point2dem*. All the options of *point2dem* were set to their default values. We use an iterative approach to obtain a refined point cloud with *stereo* and a DEM with *point2dem* from each triplet of stereo images. The first iteration uses L1B input images to produce a coarse DEM at 50 m resolution. During the second iteration, the L1B input images are orthorectified using this coarse DEM with the ASP utility *mapproject*. The orthorectified images are then processed to obtain a fine DEM at 3 m resolution. The options of the *stereo* command for this second run were empirically adjusted as explained in Sect. 4.1.2. This iterative processing was shown to improve computation time and reduce artifacts in the final DEM (Shean et al., 2016; Beyer et al., 2018). The output DEM resolution and coordinate system was defined to match those of the ASO product (UTM 11 north, WGS 84).

4.1.2 Photogrammetric processing of the images

First, *stereo* generates a dense disparity map (i.e., the pixel displacement between the two images of a stereo pair) using image correlation. The disparity map is used to calculate a point cloud with a triangulation algorithm. Then, *point2dem* interpolates the point cloud on a regular grid (Shean et al., 2016; Beyer et al., 2018). We compared three sets of options in *stereo*. These sets of options were empirically selected but do not cover all the options available in the ASP. The first set

of options is the one used by Marti et al. (2016). This set uses the local-search-window stereo algorithm and the normalized cross-correlation parametric cost function with windows of 25 px × 25 px (these options are hereafter called local search). The subpixel refinement algorithm uses an affine method. The two other sets of options use the semiglobal matching (SGM) stereo algorithm (Hirschmüller, 2005) combined with two different cost functions. The semiglobal matching is often used with nonparametric cost functions. Here we compare the two nonparametric cost functions implemented in the ASP: the binary census transform (options hereafter called SGM-binary) and the ternary census transform (hereafter called SGM-ternary). The subpixel refinement is operated during the SGM correlation with the option *Poly4* of the ASP. We evaluated the three sets of options based on the completeness of the maps and the agreement of the snow depth with the ASO using the mean bias, normalized median absolute deviation (NMAD) and RMSE of the residuals. The complete options are available in the Supplement (Table S2).

The SGM algorithm (Hirschmüller, 2005) differs from the local-search-window algorithm during the disparity map calculation. The local-search algorithm calculates the disparity for each pixel independently. The SGM algorithm optimizes the disparity over the whole image by assuming that disparity from neighboring pixels is likely to be close. This introduces more continuity in the disparity map and then in the DEM. The matching of subsets of the images of a stereo pair is measured with a cost function. The binary and ternary census transforms are two cost functions that convert a kernel centered on a pixel into a binary number. For the binary census

transform, each pixel of the kernel is compared to the central pixel of the kernel and gives 1 if it is superior to it, 0 otherwise. All the digits are concatenated in a binary number associated with the central pixel. For the ternary census transform, each comparison of a pixel with the central pixel can give three different values: 00, 01 or 11 depending on whether it is smaller, within or greater than a buffer centered on the central pixel value.

4.1.3 Comparison of bi- and tri-stereo images for DEM calculation

We calculated five DEMs from each stereo triplet by selecting a pair of images (front–nadir, nadir–back, front–back) or the complete triplet (front–nadir–back, nadir–front–back). This provided combinations of different B / H (called image geometry further in the article), ranging between 0.08 and 0.23 (Table 1). The three sets of options of *stereo* were tested on these different geometries. In the tri-stereo case, the ASP calculates two disparity maps and performs a joint triangulation when calculating the point cloud. In the first tri-stereo case (front–nadir–back), the ASP calculates a disparity map between the front and the nadir image and between the front and the back image. In the second case (nadir–front–back), the ASP calculates a disparity map between the nadir and the front image and between the nadir and the back image. The order of the images matters in the tri-stereo case since the B / H is different between front–nadir and front–back or nadir–back and front–back. We did not evaluate the third possible tri-stereo combination (back–nadir–front) as we expect results to be similar to the front–nadir–back case.

4.1.4 Snow depth (HS) maps

We coregistered the Pléiades DEMs to the ASO snow-off DEM to enable a pixel-wise comparison between both datasets. We first coregistered the Pléiades snow-off DEM to the ASO snow-off DEM. We then separately coregistered the Pléiades snow-on DEM to the Pléiades-registered snow-off DEM before computing the difference between the Pléiades snow-on and Pléiades snow-off DEMs (hereafter referred to as dDEMs). The north and south Pléiades dDEMs were mosaicked, and the north dDEM value was preserved in the overlapping area. The coregistration vectors were calculated using the algorithm by Nuth and Kääb (2011) on areas where no elevation change is expected (i.e., stable terrain). The stable-terrain areas, which are snow-free terrain without trees, were determined by a supervised classification of the Pléiades multispectral images into a land cover map (see Sect. 4.1.5). From the same land cover map, the Pléiades dDEM values were set to zero in snow-free areas to obtain the HS map. Pléiades HS values below −1 m and above 30 m were set to no data to exclude unrealistic outliers based on expert judgment and considering the minimal value that Pléiades HS could reach for actual HS close to zero.

4.1.5 Land cover classification

Snow-covered areas and stable terrain were analyzed separately, and their location was determined with a land cover supervised classification calculated from the multispectral images. The winter and summer scenes were classified into four categories: snow, forest, open water and stable terrain, the latter corresponding to snow-free areas with low vegetation or bare rock. First, we orthorectified the nadir multispectral images using *mapproject* on their corresponding DEM. For each image, we manually extracted training data covering 0.1–1.0 km^2 from a composite image of red, green, blue and near-infrared bands and the derived normalized difference vegetation index (NDVI). A maximum of 33 polygons were manually drawn for the snow class on the winter north image. These samples were used to train a random-forest classifier with *otbcli_TrainVectorClassifier* from the Orfeo ToolBox.

The stable-terrain and snow masks were shrunk (morphological erosion) to a radius of 2 pixels (4 m), and patches smaller than 30 pixels (270 m^2) were removed. The masks were shifted according to the DEM coregistration vector and then interpolated with the nearest-neighbor method onto the ASO grid. Lakes and snow patches remaining in the summer land cover map were removed from the winter snow mask. Lakes were manually delineated on snow-off images. This workflow was automated except for the training dataset, which was generated by human interpretation of the images.

4.2 Evaluation of the snow depth maps

We evaluated the Pléiades HS maps for the area where both the Pléiades (snow mask) and ASO (HS greater than zero) HS maps had snow. The term *HS residuals* in the rest of the article refers to the difference between the Pléiades and the ASO HS (Pléiades HS minus ASO HS). We also evaluated the Pléiades dDEM over stable terrain, where we expect no elevation difference over time. The stable-terrain residuals are the Pléiades dDEM values as ASO products are set to zero over snow-off terrain. The distribution of the residuals was characterized with the bias (both the mean and the median), the root mean square error (RMSE) and the normalized median absolute deviation (NMAD) of the residuals. The NMAD is a measure of the dispersion suited for populations with outliers (Höhle and Höhle, 2009).

The accuracy of HS maps is often discussed at (or close to) the highest resolution that is allowed by the sensor (e.g., Nolan et al., 2015; Marti et al., 2016). In practice however, HS maps may be subject to spatial averaging to assimilate in a snowpack model, to estimate catchment-scale HS, or to compare with coarser satellite products and model output (Painter et al., 2016; Margulis et al., 2019; Shaw et al., 2019). The accuracy of the mean HS of a set of contiguous pixels is expected to be higher than that of a single pixel but depends on the spatial correlation of the errors (Rolstad et

Table 2. Comparison of the snow depth residual (HS Pléiades minus HS ASO) and stable-terrain elevation difference (Pléiades) for different image acquisition geometries for SGM-binary options only. HS Pléiades are calculated from a combination of images taken looking to the front (F), the nadir (N) or the back (B). All metrics are in meters except the mean B / H for bi-stereo geometries, which is dimensionless. The bold line is common to this table and Table 3.

	Mean	Area (km^2)		Mean		Median		NMAD		RMSE		SD	
	B / H	snow	stable	snow	stable	snow	stable	snow	stable	snow	stable	snow	stable
FB	0.22	138.11	5.2	0.16	−0.03	0.18	0.01	0.68	0.39	0.80	1.35	0.79	1.35
FN	0.12	138.13	5.28	0.01	−0.01	0.03	0.02	1.13	0.70	1.21	1.15	1.21	1.15
NB	0.10	137.25	5.25	0.08	−0.02	0.10	0.02	1.07	0.71	1.18	1.17	1.18	1.17
FNB	**–**	**138.02**	**5.30**	**0.08**	**−0.01**	**0.10**	**0.02**	**0.69**	**0.40**	**0.80**	**1.16**	**0.79**	**1.16**
NFB	–	137.51	5.29	0.13	−0.06	0.15	−0.00	0.78	0.44	0.90	1.06	0.89	1.06

al., 2009; Anderson, 2019). Hence, we performed an empirical assessment of the evolution of the accuracy of Pléiades HS as a function of resolution by aggregating the HS residual map to resolutions ranging between 3 and 180 m (Berthier et al., 2016; Brun et al., 2017; Miles et al., 2018). An average resampling scheme was used, which calculates the average value of all valid contributing pixels. For each resolution, we compared the distribution of the HS residual or measured error to the standard error that is obtained from the error model of Rolstad et al. (2009). Using a spherical semivariogram model to measure the spatial correlation, Rolstad et al. (2009) estimate the random error of the spatially averaged residual, σ_A, as

$$\sigma_A^2 = \sigma^2 \left(1 - \frac{L}{l_{cor}} + \frac{1}{5} \frac{L^3}{l_{cor}} \right) \text{ if } L < l_{cor} \tag{1}$$

$$\sigma_A^2 = \frac{\sigma^2 l_{cor}^2}{L^2} \text{ if } L > l_{cor}, \tag{2}$$

where σ is the standard deviation of the elevation difference residual, l_{cor} is the semivariogram range or length of autocorrelation, and L is the length of aggregation (half of the pixel spacing). A is the area of aggregation (pixel area) and is related to L ($A = 4 \times L^2$). This formula assumes no spatial trend in the HS residual map. We estimated l_{cor} from a semivariogram analysis of the HS residuals at the highest resolution (3 m). The value of σ was taken as the NMAD of the HS residuals at the highest resolution (3 m). We also tested this equation using the stable-terrain residuals to set the value of l_{cor} and σ. The measured error was taken as the NMAD of the aggregated HS residual. By comparing the measured error and the modeled error, we aim to verify whether (i) the error model from Eqs. (1) and (2) is valid and (ii) its parameters (σl_{cor}) can be estimated from stable-terrain residuals only. It is important to evaluate whether the stable-terrain residuals can be used to parameterize the error model because that is the only available information in regions without HS reference data (Deschamps-Berger et al., 2019).
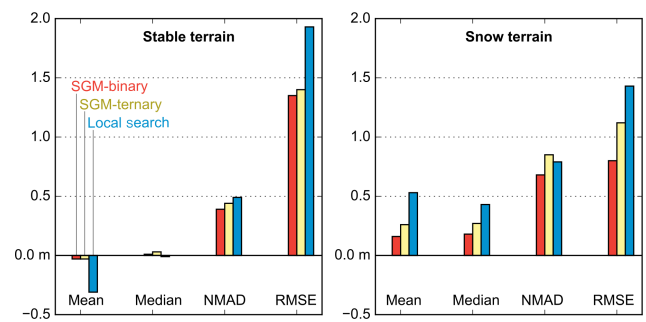

Figure 3. Mean, median, NMAD and RMSE of the residual of HS maps depending on the ASP *stereo* correlation option. The options compared are the SGM algorithm with the binary census transform cost function (SGM-binary in red), the ternary census transform cost function (SGM-ternary in yellow) and the local-search algorithm (local search in blue).

5 Results

5.1 Evaluating the impact of bi- or tri-stereo images as input

We first investigate the impact of different image geometries on the HS maps while keeping the stereo SGM-binary option fixed. The NMAD of the HS residuals with respect to ASO data (Table 2) is larger for maps from pairs of images with B–H around 0.12 (1.13 m for front–nadir, 1.07 m for nadir–back) than from pairs of images with B / H around 0.20 (0.68 m for front–back) or triplets of images. The NMAD of the snow depth residuals from the front–nadir–back triplets (0.69 m) is slightly better than from the nadir–front–back triplets (0.78 m) and very similar to the NMAD from the front–back pair. The NMAD over stable terrain is lower, but relative values between two geometries are similar (Table 2). For the different image geometries, the RMSE evolves similarly to the NMAD over snow-covered areas but very differently over stable terrain. The largest RMSE over stable terrain is 1.35 m for front–back, and the smallest is 1.06 m for nadir–front–back. The mean differences over snow-covered areas range from +0.01 m (front–nadir) to +0.16 m (front–

Table 3. Comparison of the snow depth residual (HS Pléiades minus HS ASO) and stable-terrain elevation difference (Pléiades) depending on the ASP *stereo* options for front–nadir–back geometry only. All metrics are in meters. The bold line is common to this table and Table 2.

	Area (km^2)		Mean		Median		NMAD		RMSE		SD	
	snow	stable	snow	stable	snow	stable	snow	stable	snow	stable	snow	stable
SGM-binary	138.02	5.30	0.08	−0.01	0.10	0.02	0.69	0.40	0.80	1.16	0.79	1.16
SGM-ternary	138.14	5.21	0.24	−0.03	0.25	0.03	0.85	0.44	1.11	1.30	1.09	1.30
Local search	135.96	5.32	0.49	−0.32	0.39	−0.00	0.80	0.51	1.41	1.94	1.32	1.92

back). The absolute means and medians over stable terrain are all less than 0.06 m. The relative results for the different geometries are similar to the SGM-ternary and local-search options except for the mean error (not shown here). In the following sections, the HS map from the front–nadir–back geometry is used as it yields the lowest bias, RMSE and NMAD of all the geometries, although it is similar to the front–back geometry.

5.2 Sensitivity to the photogrammetric processing

We compare the *stereo* options on the HS maps from the front–nadir–back geometry (Table 3 and Fig. 3). The SGM sets of options provide DEMs without data gaps. The local-search option produces snow-on DEMs with gaps, which results in ∼ 2 km^2 missing in the HS maps compared to the SGM options (Table 3). Visual examination of the winter DEMs shows large differences in snow fields and forest. Linear artifacts are observed over snow in the DEM produced with the SGM-ternary option (Fig. S1). The same regions are noisy in SGM-binary. Patches of typically 20 m × 20 m with abnormally large HS (> 10 m) compared to ASO (∼ 3 m) are also observed with the local-search options around isolated trees. These artifacts are not visible with the SGM-binary or ternary options (Fig. S1).

The mean differences from the ASO snow depth data range from +0.08 m (SGM-binary) to +0.49 m (local-search option). The mean differences are larger for SGM-ternary (+0.24 m) than SGM-binary. The NMAD of the residuals is smaller for SGM-binary (0.68 m) than the local-search (0.80 m) and SGM-ternary options (0.85 m). Over stable terrain, the absolute mean and median of the elevation differences are less than 0.03 m except for the mean of the local-search option, which is −0.32 m. The mean of the elevation differences for local search decreases to −0.03 m when the elevation differences are excluded if they exceed 3 times the NMAD value. This is expected as the same filtering is used during the coregistration process to remove outliers. In the following, the SGM-binary was selected since it gives the lowest bias and NMAD with respect to ASO data and the lowest NMAD over stable areas (Table 3).

5.3 Spatial distribution of the residuals

The Pléiades HS map calculated with the selected image geometry and ASP configuration (front–nadir–back images, SGM-binary) compares well with the ASO HS map (Fig. 4). Typical mountain snowpack features (e.g., avalanche deposits and snow drift accumulation) can be identified on the Pléiades HS map (Figs. 4d, e, 5). Pléiades HS data are available over 215 km^2 of open terrain but not for the 23 km^2 of forest. No HS was higher than 30 m, but 0.25 km^2 of HS were excluded because HS was less than −1 m. This occurred in areas covered with low-density deciduous vegetation, which was classified as snow. The intersection area of Pléiades and ASO snow-covered areas is 138 km^2 after erosion of the Pléiades snow mask. The Pléiades mean (median) HS is 4.05 m (4.13 m) against 3.96 m (4.02 m) for ASO over the common snow-covered area. ASO and Pléiades HS exhibit a similar relationship between HS and elevation (Fig. 6) except between 1900 and 2100 m and between 3500 and 3700 m, where the mean residual over snow-covered areas is greater than 0.25 m (Fig. 7). This corresponds, however, to elevation ranges that cover less than 0.05 km^2 each.

The NMAD of the Pléiades dDEMs over the 4.07 km^2 of stable terrain is 0.40 m against 0.69 m for the HS residual. The distribution of residuals on stable terrain is similar for most aspect classes with the exception of the north-facing slopes (0.26 km^2, aspect classes 315–360° and 0–45°; Fig. 7). Based on a visual analysis of the residual map, we attribute these errors to shaded slopes of steep summits. The distribution of HS shows a similar spread for all aspects but a larger positive bias (∼ 0.20 m) for south-facing slopes (90–270°, Fig. 7). The distribution of HS residuals against the terrain slope is similar between 0 and 50° but has a greater spread in steeper terrain, which covers 2.13 km^2. The same trend is observed over stable terrain but only above 70°.

The semivariogram of the residual increases from 0.2 to 0.8 linearly for lag distances between 3 and 20 m (Fig. 9a). Low-amplitude undulation for lag distances between 2000 and 8000 m (Fig. 9b) is related to a low-frequency undulation in the HS residual map, which has an amplitude of approximately 0.30 m and a wavelength of about 4 km (Fig. 8). The crests of the undulation are oriented in the east–west direction (Fig. 8). Such an undulation pattern was observed in other Pléiades products, ASTER images (Girod et al., 2017)
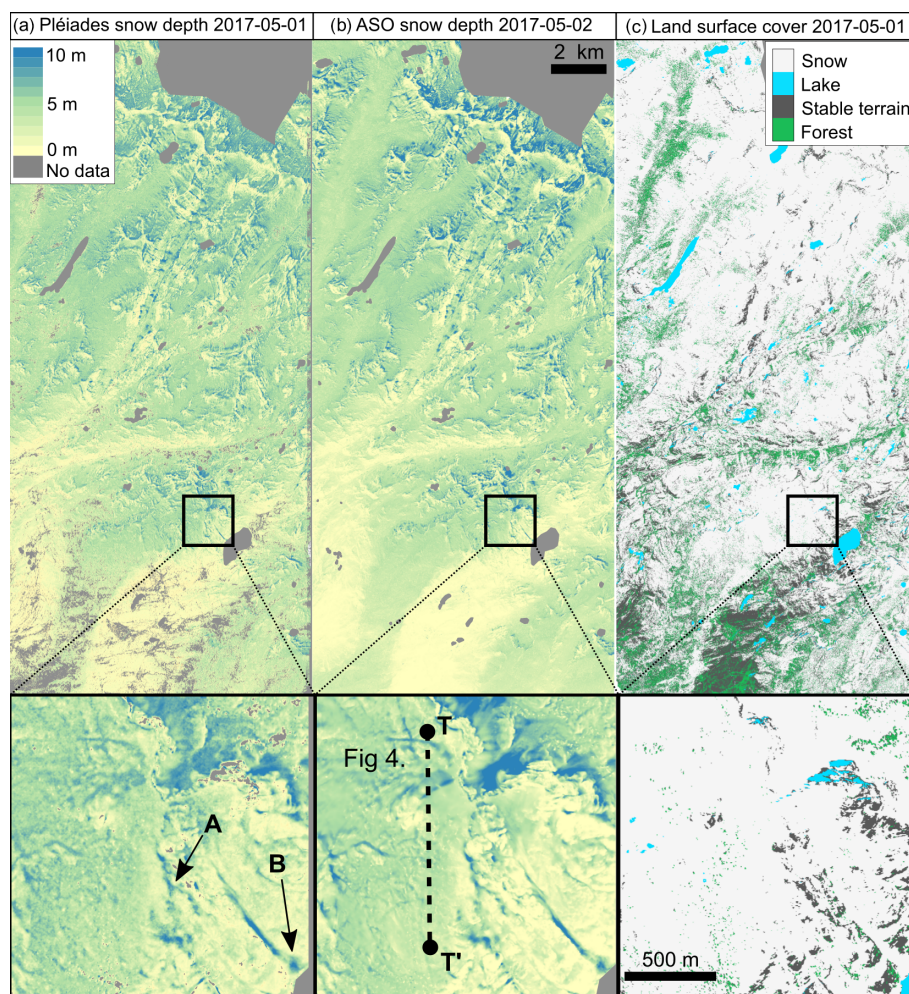
Figure 4. Snow depth maps from Pléiades data on 1 May 2017 **(a, d)** and from ASO on 2 May 2017 **(b, e)**. Cornice (spot A) and avalanche deposits (spot B) are visible on Pléiades HS maps **(d)**. The land surface cover is shown in **(c)** and **(f)** over the same area. Black squares in **(a–c)** are the area shown in **(d–f)**. The transect $T–T'$ is shown in Fig. 5. All datasets have the same spatial resolution (3 m). The difference of the maps **(a)** and **(b)** (Pléiades minus ASO) is in Fig. 8.

and WorldView DEMs (Fig. 10 in Shean et al., 2016; Fig. 6 in Bessette-Kirton et al., 2018). It is attributed to nonmodeled satellite attitude oscillations along-track (jitter). A similar semivariogram shape is obtained over stable terrain. From this semivariogram analysis we estimate that the correlation length of the residuals (see Sect. 4.2) is about 20 m for both snow and stable areas.

5.4 Evaluation of the Rolstad error model

The measured error of the HS map decreases with increasing resampling resolution (Fig. 10). The NMAD of the HS residuals is reduced by a factor of almost 2 by resampling from the original resolution of 3 m (NMAD = 0.69 m) to 36 m (NMAD = 0.38 m). As explained in Sect. 4.2, we computed two error models using either the HS residuals ($l_{cor} = 20$ m, $\sigma = 0.69$ m) or the stable-terrain residuals ($l_{cor} = 20$ m, $\sigma =$
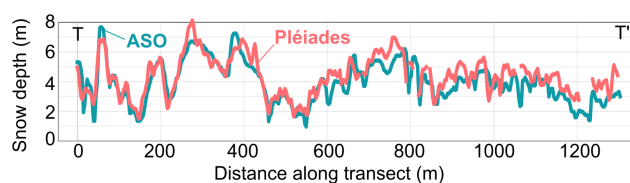

Figure 5. Transect $T–T'$ of snow depth visible in Fig. 4e from Pléiades data (pink) and ASO (blue). More transects are available in the Supplement (Figs. S3 and S4).

0.40 m) to parameterize Eqs. (1) and (2). We find that the NMAD of the HS residual well matches the error modeled for averaging areas smaller than 10^3 m^2 when l_{cor} and σ are calculated with the HS residuals (Fig. 10). However it does not match with the modeled error for averaging areas larger than 10^3 m^2 (Fig. 10). This is due to the lower decrease
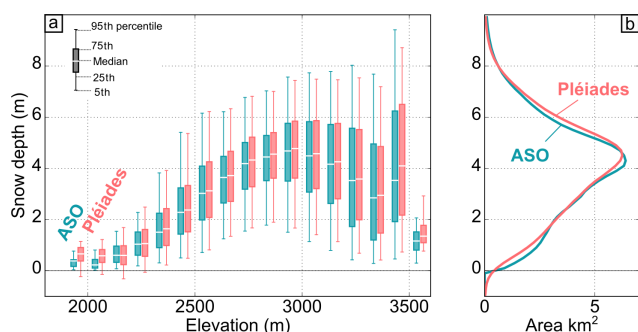
Figure 6. Snow depth against elevation **(a)** and total distribution **(b)** from Pléiades data (pink) and ASO (blue). The box plots show the median value (white line), the 25th and 75th percentile (box), and the 5th and 95th percentile (whiskers).
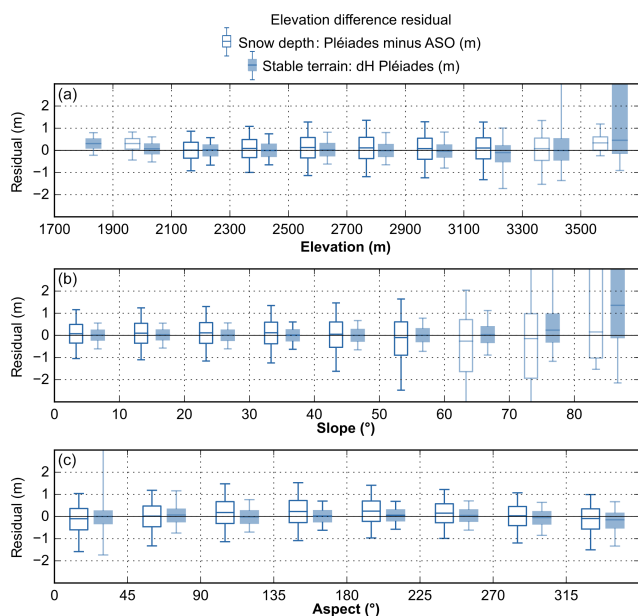

Figure 7. Distribution of the residuals between the Pléiades and ASO snow depth maps over the snow-covered area (empty box) and stable terrain (filled box) against elevations **(a)**, slopes **(b)** and aspect **(c)**. Over stable terrain, the ASO product is set uniformly to zero. Boxes where data were covering less than 1 km^2 are slightly transparent.

in the residual dispersion with spatial resolution. The measured NMAD decreased by 0.07 m between 36 m resolution and 180 m resolution, while the modeled error decreased by 0.22 m between the same resolutions. We attribute this mismatch to the undulation pattern identified in Sect. 5.3 (see Sect. 6.5 in Discussion).
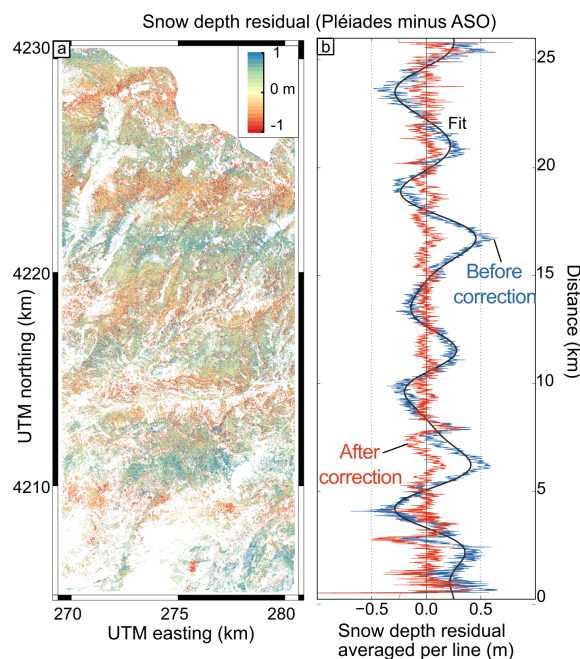

Figure 8. Residual snow depth (Pléiades minus ASO) over the complete study area **(a)** and average per line **(b)**. In **(b)**, the HS residual before correction (blue) is corrected for the low-frequency undulation (black) to obtain a corrected signal (red).

6 Discussion

6.1 Comparison to other studies using satellite photogrammetry

By comparing the Pléiades HS with the ASO data, we find an NMAD of 0.69 m in the best case (i.e., best acquisition geometry and ASP options), which is close to or higher than most previous evaluations (Table 4). Only Marti et al. (2016) measured a larger NMAD (0.78 m), with a reference HS map of 3.15 km^2 that was obtained by UAV photogrammetry. The spread in accuracy between studies in Table 4 could be due to differences in (i) the satellite data (i.e., acquisition geometry, image resolution), (ii) the characteristics of the study site and (iii) the representativeness of the validation data. The comparison with snow probe measurements showed NMAD about a third lower than this study, at 0.45 m ($n = 442$; Marti et al., 2016) and 0.47 m ($n = 36$; Eberhard et al., 2020), but covered limited portions of the studied sites. The B / H for the images of Marti et al. (2016) range between 0.21 and 0.25 for all consecutive stereo pairs, while our B / H ranges between 0.08 and 0.12. This is consistent with photogrammetry theory, which states that the accuracy of the DEM increases with the B / H up to a certain limit (Delvit and Michel, 2016). We find a similar NMAD to Eberhard et al. (2020), who calculated an HS map from a Pléiades snow-on DEM and an airplane structure from motion (SfM) snow-off DEM and compared it to HS from the
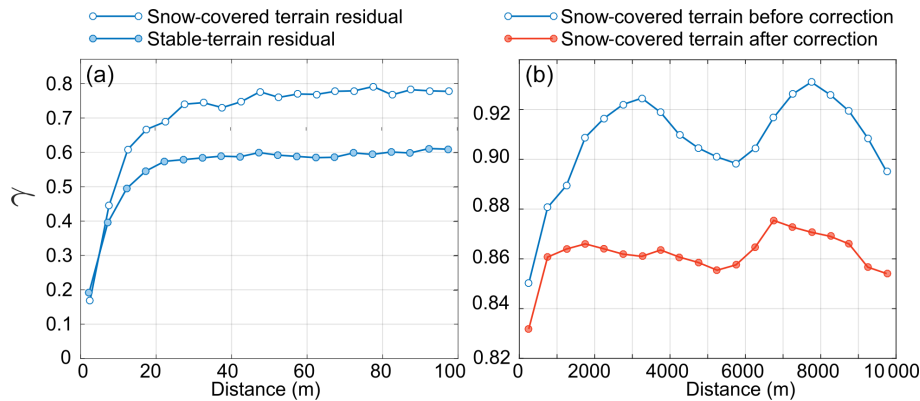
Figure 9. (a) Semivariogram or spatial autocorrelation (γ) against lag distance of the HS residuals (empty circles) and Pléiades elevation difference over stable terrain (filled circle). **(b)** Semivariogram of the HS residuals for large distances before (blue line) and after correcting the undulation pattern (red line), illustrating the reduction in spatial variance at greater lag distances due to this correction (Sect. 6.5).
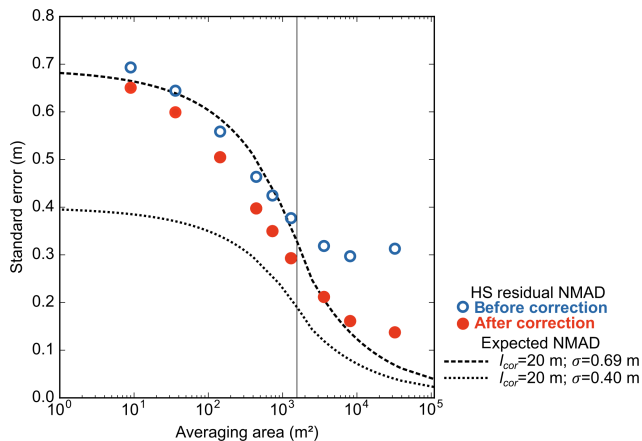

Figure 10. Measured error and modeled error (Rolstad et al., 2009) of the HS averaged over different areas. Modeled error (dashed line) is predicted based on different random error per pixel (σ) and autocorrelation length (l_{cor}; Eq. 3). The lines are the modeled error based on l_{cor} from the semivariogram and σ taken as the NMAD derived from stable-terrain residuals (dotted line) and HS residuals (dashed line). Empty blue circles are the NMAD of the residual HS maps averaged at different resolutions before the undulation correction. Filled red circles are the NMAD of the residual HS maps averaged at different resolutions after the undulation correction.

airplane SfM over $75\,\mathrm{km}^2$ (NMAD $= 0.65\,\mathrm{m}$). Finally, Mc-Grath et al. (2019) found an NMAD of 0.24 m for HS from WorldView-3 stereo DEMs using 2107 point observations from ground-penetrating radar surveys over a flat area of roughly $50\,\mathrm{km}^2$. This lower NMAD value might result from the higher resolution of the WorldView-3 images (0.3 m) together with the flatter terrain used for evaluation. As the ASO provides a much larger reference dataset over a complex terrain, we argue that our study provides a more robust evaluation of the HS accuracy that can be expected from Pléiades in high mountain regions. While the ASO data itself may add

some error, the published accuracy of the ASO HS data is significantly better than Pléiades. In all these studies, the absolute mean biases range between 0.01 m (McGrath et al., 2019) and 0.35 m (Eberhard et al., 2020).

6.2 Sensitivity to image geometry

We find that the HS map accuracy is sensitive to the B / H ratio of the input images and to the configuration details of the photogrammetric processing. We do not find a large added value of the tri-stereo images for the map accuracy compared to an optimal bi-stereo configuration. Tri-stereo might provide greater benefits in case of image occlusion in steep slopes, which is more prone to occur with higher B / H.

The NMAD of the Pléiades HS is improved by 36 % when using images with a B / H of 0.22 instead of 0.11 (Table 3). Marti et al. (2016) used pairs of front–nadir and nadir–back images (B / H = 0.2) as they observed that the front–back pair (B / H = 0.4) led to too many no-data pixels. From these two studies and for similar terrain, a triplet of images with a B / H for consecutive images around 0.2 seems to be a good compromise. It should ensure high coverage and good DEM precision. Further work is needed to confirm this statement by testing varying B / H values.

Using tri-stereo instead of bi-stereo images did not significantly improve the Pléiades HS map accuracy. It seems like the processing of a triplet of stereo images (front, nadir, back) with the ASP *stereo* function is equivalent to the processing of the best stereo pair of the triplet, the front–back pair in our case. There were no data gaps due to view obstruction by steep relief in this study area. Should this be the case, the tri-stereo may offer better coverage. Several studies have evaluated the benefits of tri-stereo imagery against bi-stereo (Berthier et al., 2014; Zhou et al., 2015; Bagnardi et al., 2016; Marti et al., 2016). However, these studies used different photogrammetric software that does not handle the combination of three images in the same way. For example,

Table 4. Comparison of HS accuracy with studies using satellite photogrammetry.

	Satellite (resolution)	HS map resolution	Validation data	Area	Number of measurements	Mean	Median	NMAD	RMSE
This study	Pléiades (0.5 m)	3 m	Airplane lidar	138 km^2		0.08	0.10	0.69	0.80
Marti et al. (2016)	Pléiades (0.5 m)	2 m	Snow probing		442		−0.16	0.45	
			UAV SfM	3.15 km^2		−0.06	−0.14	0.78	
McGrath et al. (2019)	WorldView-3 (0.3 m)	8 m	Ground-penetrating radar		2107	+0.01	+0.03	0.24	
Shaw et al. (2019)	Pléiades (0.5 m)	4 m	Terrestrial lidar	0.74 km^2		−0.10	−0.22	0.36	0.52
Eberhard et al. (2020)*	Pléiades (0.5 m)	2 m	Snow probing		36	−0.35	−0.36	0.47	0.52
			UAV SfM	4 km^2		−0.18	−0.18	0.38	0.44
			Airplane SfM	75 km^2		−0.02	−0.18	0.65	0.92

* Eberhard et al. (2020) used a Pléiades DEM for snow-on and UAV or airplane SfM DEM for snow-off.

either multiple disparity maps, point clouds or DEMs can be calculated and merged to produce a final single DEM. The use of tri-stereo results in increasing the density of the point cloud (Zhou et al., 2015; Bagnardi et al., 2016) and decreasing the area with missing data in the final DEM (Berthier et al., 2014; Zhou et al., 2015). The accuracy of elevation products from tri-stereo was slightly modified in Berthier et al. (2014) and Marti et al. (2016) compared to bi-stereo, with an increase or decrease in the NMAD by a few percent.

To our knowledge, surface elevation changes were never computed from a large number of very-high-resolution satellite stereo images (> 10), but a study suggests that the combination of multiview images can improve the DEM quality. The fusion of 16 WorldView-3 images improved the NMAD of the residual by 20 % compared to a set of 6 images over an industrial zone (Rupnik et al., 2018). Therefore, the most important use of tri-stereo may not be to improve the accuracy of HS maps but rather to obtain complete coverage of complex terrain and have a less distorted nadir orthoimage for land surface classification. We did not evaluate the extent to which the front and back images would provide a different land surface classification from the one obtained with the nadir image.

6.3 Sensitivity to photogrammetric processing

The choice of the photogrammetric options has an impact on the elevation difference accuracy over stable terrain and snow-covered areas. The NMAD over snow-covered areas is improved by 0.16 m by modifying the cost function (binary census transform instead of ternary census transform).

However, such a decrease in the dispersion will hardly impact the HS averaged over a region of interest since the random error decreases rapidly with increases in averaging area (see Sect. 6.5). More important is the larger mean bias over snow-covered areas introduced with the SGM-ternary option (0.24 m) and local-search option (0.49 m) compared to the SGM-binary option (0.08 m). This bias is particularly important for south-facing slopes. It seems to result from difficulties in image matching in bright areas for the three options and from the impact of isolated trees for local search. The impact of the trees is likely due to the larger kernel size (25 pixels) used in the local-search option. The exact origin of the bias on south-facing slopes remains unknown.

6.4 Attribution of the HS error

We found a mean difference of +0.08 m between Pléiades (SGM-binary, front–nadir–back) and ASO HS despite the correction of the vertical offset between the snow-on and snow-off DEM using stable terrain after coregistration. This bias is low given the differences in the characteristics of the ASO and the Pléiades products. This can be due to many factors, including the effect of vegetation. First, the ASO snow-off DEM is a digital terrain model, while the Pléiades snow-off DEM is a digital surface model. Tall vegetation (i.e., trees) is identified during the classification of the MS images and does not impact the HS evaluation. But short vegetation completely covered with snow in winter is not identified in the classification. For ASO products, filtering based on the multiple lidar returns produced by vegetation should provide the ground elevation, but short vegetation often does

not produce multiple returns (Painter et al., 2015). Furthermore, there is a large known error in vegetation height measured with Pléiades DEMs (Piermattei et al., 2018). Thus, it is still unclear which surface is sensed by each method between the top of the vegetation and the underlying ground.

We found that the random error is larger on snow-covered terrain (NMAD = 0.69 m) than on stable terrain (NMAD = 0.40 m). This is true for all slopes and most aspect classes (Fig. 7). Although mountainous snow surfaces tend to have smoother topography, thereby increasing the accuracy of the photogrammetric processing, bright snow surfaces also tend to have less texture than snow-free surfaces, which decreases the accuracy of the photogrammetric processing. The lower accuracy of snow areas is not due to saturation since no pixel was saturated in the panchromatic images. In addition, the residuals over stable terrain were computed from Pléiades data only, while residuals over snow-covered areas were computed from Pléiades and ASO data. Finally, the coregistration of the snow-on DEM was optimized using the stable terrain (Sect. 4.1.4); therefore a lower NMAD on stable terrain may be due to the coregistration step and not the photogrammetric processing in itself. Based on the above, we cannot conclude whether the larger dispersion over snow-covered areas results from the properties of the surface.

We further compared Pléiades snow-off DEM with the ASO snow-off DEM and Pléiades snow-on DEM with the ASO snow-on DEM. The latter was calculated by adding the ASO snow-off DEM and the ASO HS. Both Pléiades DEMs are coregistered as described in Sect. 4.1.4. We find a mean bias over snow-covered terrain of +0.13 m for snow-off conditions and +0.21 m for snow-on conditions (Table S3). These biases are of the same order of magnitude and suggest that a bias in the Pléiades snow-on DEM is partially compensated by the difference of the surface observed in the snow-off DEM (see above). In addition, the ASO snow-off DEM was acquired in October 2015 and the Pléiades snow-off DEM in August 2017. Growth or decay of the vegetation can occur over almost 2 years, leading to elevation differences between the snow-off DEMs. The NMAD is larger for snow-off DEMs (0.80 m) and snow-on DEMs (0.93 m) compared to the HS residual (0.69 m). This shows that some errors are consistently present in the snow-off and snow-on DEMs of each type (airplane lidar or satellite photogrammetry). Pléiades DEMs indeed overestimate the surface elevation as the terrain slope increases (Fig. S2). This suggests that combining satellite photogrammetry and airplane lidar DEMs may lead to larger errors than comparing DEMs from the same platform.

6.5 Evaluation of an error model at different resolutions

The error predicted with Eqs. (1) and (2) does not agree with the NMAD of measured HS error for averaging areas larger than 10^3 m^2 (Fig. 10). This is likely because Eqs. (1) assumes

a randomly distributed error beyond the short-distance correlation length (here 20 m), while the undulation pattern identified in Fig. 8 introduces an additional spatial correlation at larger scales in the HS residual map. To verify this explanation, we applied an empirical correction to remove the undulation pattern from the residual map. We averaged the HS residuals by pixel rows in the across-track direction and used a Fourier transform to identify the undulation frequencies (adapted from Girod et al., 2017). Then, we modeled this error by selecting the frequencies lower than 4×10^{-4} m^{-1} (i.e., wavelength longer than 2.5 km) and removed it from the HS map. As expected, this correction makes the semivariogram of the HS residual flatter for lag distances between 2000 and 8000 m (Fig. 9b). As a result, there is a better agreement between the HS residuals' NMAD and the modeled error, with σ and l_{cor} estimated from the HS residuals ($l_{cor} = 20$ m, $\sigma = 0.69$ m; Fig. 10). The improvement is more marked at lower resampling resolution. For instance, the HS NMAD is reduced after correction by 50 % at a resolution of 180 m. The improvement is under 10 % at 20 m resolution as expected since the correction only dampers a low-frequency signal. When the stable-terrain residuals are used to compute Eqs. (1) and (2) ($l_{cor} = 20$ m, $\sigma = 0.40$ m), the modeled error is lower than the measured error. This is expected since the NMAD of the stable-terrain residuals is lower than the NMAD of the HS residual. However, the discrepancy between both models decreased at coarser resolution.

This analysis shows that the model proposed by Rolstad et al. (2009) provides a good first-order estimation of the random error after spatial aggregation under the assumption that there is no spatial drift in the error at scales beyond the correlation length. In most cases, the statistics of the HS residuals are not available and might only be measured on stable terrain. Interestingly, in this study the correlation length of the error is similar over stable terrain and snow terrain. However, the dispersion (NMAD, standard deviation) is 2 times larger over snow-covered terrain than stable terrain, which leads to a proportional underestimation of the error. Finally, although the bias or systematic error is corrected on stable terrain, there remains a bias on HS of the order of ∼ 0.20 m (Table 4) that should be taken into account in the error calculation. According to the literature, this bias can be estimated by comparing the mean and median of elevation differences over stable terrain (Gardelle et al., 2013) or by calculating the residual of the coregistration vector when more than two elevation datasets are available (Nuth and Kääb, 2011).

6.6 Comparison of satellite photogrammetry with airborne methods

ALS provides HS maps with a better accuracy (RMSE < 0.10 m) than Pléiades and potentially a finer horizontal resolution too (Painter et al., 2016). One significant advantage of ALS is that it can measure HS under the tree canopy and in shaded areas. It is also able to acquire data

in overcast conditions provided that the clouds are above the aircraft. However, from this study and that of Marti et al. (2016), it appears that the accuracy of Pléiades HS maps is sufficient to provide valuable information in regions where there is no ALS monitoring capability (the vast majority of mountain regions with snow cover). A limitation of current very-high-resolution sensors such as Pléiades is their narrow swath (20 km for Pléiades), which impedes the acquisition of large areas with a frequent revisit. In particular, there are areas of high tasking competition at lower latitudes, where it can be challenging to obtain a stereo pair at the right time of the snow season. More frequent acquisitions should, however, become easier as new stereo satellite fleets are to be launched in the coming years (Pléiades Neo, WorldView Legion). The acquisition of visible images will always be limited by the presence of clouds, making some regions hard to study at least during some seasons.

HS maps from UAV SfM typically exhibit a centimeter-scale bias (0.05 to 0.11 m) and an RMSE between 0.05 and 0.30 m based on comparison with snow probe and Global Navigation Satellite System (GNSS) measurements. This is more accurate than what is currently achieved with satellite photogrammetry. However, UAV campaigns are currently limited to areas of a few square kilometers due to battery limitation and often rely on numerous ground control points. This greatly limits the possibility to cover large and remote areas. Airplane SfM exhibits accuracy close to UAV SfM, with an NMAD typically of 0.30 m (Bühler et al., 2015), and presents the same potential and logistic limitations as airplane laser-scanning campaigns. The reader is referred to the study of Eberhard et al. (2020) for a detailed discussion on the different approaches to map snow depth with photogrammetry.

6.7 Generalization to other regions

Several snow applications could benefit from HS maps from satellite photogrammetry. First, this study could be reproduced in any place of the globe provided that (i) there is a window to acquire snow-off images and (ii) there is a way to coregister the series of DEMs, for example with stable terrain. This method is particularly suited for snow volume evaluation at a basin scale in alpine areas (this study site; Marti et al., 2016; McGrath et al., 2019; Shaw et al., 2019). Observing shallow snowpack (HS roughly below 0.5 m, e.g., polar environments) might not be as straightforward as the typical spatial variability lies within our range of uncertainty (roughly 0.5 m). However, even landscapes with shallow snowpack often feature local accumulation of snow, which would be measurable with satellite photogrammetry. Therefore it is hard to qualify this method as unfit to any region, but future studies are required to confirm its usefulness in these challenging contexts. Study of shallow snowpack would clearly benefit from higher-accuracy DEMs through correction of the satellite jitter or increases in image resolution.

A lack of well-distributed stable terrain in snow-on and snow-off DEMs can complicate the coregistration process in some regions. The horizontal component of the coregistration vector can be measured without differencing stable terrain and snow-covered terrain (Marti et al., 2016), but the vertical component requires some stable terrain or an elevation reference. Ground control points (GCPs) could be used but would limit the applicability of the method in remote mountains. Besides, it remains to be tested how many GCPs would be required and how precisely their position should be measured.

There are already a number of efficient free and open-access photogrammetric software tools that are under continuous development. These tools enable a high level of automation and are compatible with high-performance computing environments (Howat et al., 2019). In our workflow, the last step to automate is the collection of training samples for image classification. This could be done by using an unsupervised classification algorithm or by using an external land cover classification. Preliminary work with a time series of Pléiades images in the Pyrénées (not shown here) suggests that it is not possible to simply use the classification model from a previous year to generate the classification of the current year. A possibility may be to use a Sentinel-2 snow cover map to extract training samples in the Pléiades multispectral images since Sentinel-2 images have a shortwave infrared band, which enables a robust and unsupervised detection of snow cover (Gascoin et al., 2019). Differentiating terrain covered with vegetation from stable terrain would remain challenging.

We find that the selection of the image configuration and the processing options can lead to changes in the NMAD up to ∼ 0.3 m. Figure 10 suggests that this variation is likely to become insignificant if the HS map is aggregated at a larger spatial scale (grid spacing larger than $100\,m \times 100\,m$). Such optimization is therefore more important for the study of small-scale features (wind drift, avalanches, typically about a few tens of meters) or to decrease bias on specific terrain (south slopes, fields with isolated trees). The optimization of the photogrammetric processing can also be important when little stable terrain is available for the coregistration step.

7 Conclusion

We found a good agreement between snow depth (HS) maps from high-resolution stereo satellite images with airplane laser-scanning HS maps over 138 km² of mountainous terrain in California. The mean residual is +0.08 m, the NMAD is 0.69 m, and the RMSE is 0.80 m. Comparison of individual DEMs shows a growing positive bias with slope in Pléiades DEMs. This bias is of similar magnitude in both snow-on and snow-off Pléiades DEMs and thus cancels out in the HS map, leading to agreement between Pléiades and airplane laser-scanning HS for all slopes up to 60°. South-

facing slopes seem prone to a positive bias in the Pléiades HS (~ 0.2 m). These areas were found to have less texture in the panchromatic images. The main drawbacks of the satellite stereo HS method are the lack of data under dense tree cover, the reduced accuracy in shaded areas and the current challenge to image large regions in a short time. We found that the accuracy of the maps was sensitive to the B / H and the photogrammetric processing options. Using the current ASP multiview triangulation routines, we could not find a clear benefit from the use of a triplet of images compared to a pair with optimal B / H (about 0.2). The accuracy of the HS maps can be improved by decreasing their resolution. This improvement cannot be described with a well-accepted statistical model partly due to an undulation pattern commonly observed in DEMs derived from satellite photogrammetry. We observe that the accuracy is improved by 50 % when decreasing the HS map resolution from 3 to 36 m. We conclude that satellite photogrammetric measurements of HS are relevant for snow studies as they offer accuracy of ~ 0.70 m at 3 m resolution, a high level of automation and the potential to cover remote regions around the world.

Author contributions. CDB, SG, EB, and EG designed the study. CDB performed the data curation and formal analysis. CDB, SG, and EB wrote the first draft of the manuscript, which was later edited by all authors. SG, EB, MD, and CDB participated in the funding acquisition.

References

Anderson, S. W.: Uncertainty in quantitative analyses of topographic change: error propagation and the role of thresholding, Earth Surf. Proc. Land., 1033, 1015–1033, 2019.

Bagnardi, M., González, P. J., and Hooper, A.: High-resolution digital elevation model from tri-stereo Pleiades-1 satellite imagery for lava flow volume estimates at Fogo Volcano, Geophys. Res. Lett., 43, 6267–6275, https://doi.org/10.1002/2016GL069457, 2016.

Berthier, E., Arnaud, Y., Kumar, R., Ahmad, S., Wagnon, P., and Chevallier P.: Remote sensing estimates of glacier mass balances in the Himachal Pradesh (Western Himalaya, India), Remote Sens. Environ., 108, 327–338, https://doi.org/10.1016/j.rse.2006.11.017, 2007.

Berthier, E., Vincent, C., Magnússon, E., Gunnlaugsson, Á. Þ., Pitte, P., Le Meur, E., Masiokas, M., Ruiz, L., Pálsson, F., Belart, J. M. C., and Wagnon, P.: Glacier topography and elevation changes derived from Pléiades sub-meter stereo images, The Cryosphere, 8, 2275–2291, https://doi.org/10.5194/tc-8-2275-2014, 2014.

Berthier, E., Cabot, V., Vincent, C., and Six, D.: Decadal Region-Wide and Glacier-Wide Mass Balances Derived from Multi-Temporal ASTER Satellite Digital Elevation Models. Validation over the Mont-Blanc Area, Front. Earth Sci., 4, 1–16, https://doi.org/10.3389/feart.2016.00063, 2016.

Bessette-Kirton, E. K., Coe, J. A., and Zhou, W.: Using Stereo Satellite Imagery to Account for Ablation, Entrainment, and Compaction in Volume Calculations for Rock Avalanches on Glaciers: Application to the 2016 Lamplugh Rock Avalanche in Glacier Bay National Park, Alaska, J. Geophys. Res.-Earth Surf., 123, 622–641, https://doi.org/10.1002/2017JF004512, 2018.

Beyer, R. A., Alexandrov, O., and McMichael, S.: The Ames Stereo Pipeline?: NASA's Open Source Software for Deriving and Processing Terrain Data Special Section, Earth Space Science, 5, 537–548 https://doi.org/10.1029/2018EA000409, 2018.

Brun, F., Berthier, E., Wagnon, P., Kääb, A., and Treichler, D.: A spatially resolved estimate of High Mountain Asia glacier mass balances from 2000 to 2016, Nat. Geosci., 10, 668–673, https://doi.org/10.1038/NGEO2999, 2017.

Bühler, Y., Marty, M., Egli, L., Veitinger, J., Jonas, T., Thee, P., and Ginzler, C.: Snow depth mapping in high-alpine catchments using digital photogrammetry, The Cryosphere, 9, 229–243, https://doi.org/10.5194/tc-9-229-2015, 2015.

Bühler, Y., Adams, M. S., Bösch, R., and Stoffel, A.: Mapping snow depth in alpine terrain with unmanned aerial systems (UASs): potential and limitations, The Cryosphere, 10, 1075–1088, https://doi.org/10.5194/tc-10-1075-2016, 2016.

Deems, J. S., Fassnacht, S. R., and Elder K. J.: Fractal Distribution of Snow Depth from Lidar Data, J. Hydrometeorol., 7, 285–297, https://doi.org/10.1175/JHM487.1, 2006

Delvit, J. and Michel, J.: Modèles Numériques de Terrain à partir d'images optiques, Observation des Surfaces Continentales par Télédétection optique: Techniques et méthodes, 366 pp., 2016.

De Michele, C., Avanzi, F., Passoni, D., Barzaghi, R., Pinto, L., Dosso, P., Ghezzi, A., Gianatti, R., and Della Vedova, G.: Using a fixed-wing UAS to map snow depth distribution: an evaluation at peak accumulation, The Cryosphere, 10, 511–522, https://doi.org/10.5194/tc-10-511-2016, 2016.

Deschamps-Berger, C., Gascoin, S., Berthier, E., Lacroix, P., and Polidori, L.: La Terre en 4D: apport des séries temporelles de modèles numériques d'élévation par photogrammétrie spatiale pour l'étude de la surface terrestre, Revue Française de Photogrammétrie et de Télédétection, 221 pp., 2019.

Deschamps-Berger, C., Gascoin, S., Berthier, E., and Dehecq, A.: Snow depth and land surface cover in Tuolumne basin (California) from Pléiades images, Data set, Zenodo, https://doi.org/10.5281/zenodo.4013939, 2020.

Dozier, J., Bair, E. H., and Davis, R. E.: Estimating the spatial distribution of snow water equivalent in the world's mountains, WIREs Water, 3, 461–474, https://doi.org/10.1002/wat2.1140, 2016.

Eberhard, L. A., Sirguey, P., Miller, A., Marty, M., Schindler, K., Stoffel, A., and Bühler, Y.: Intercomparison of photogrammetric platforms for spatially continuous snow depth mapping, The Cryosphere Discuss., https://doi.org/10.5194/tc-2020-93, in review, 2020.

Fassnacht, S. R. and Deems, J. S.: Measurement sampling and scaling for deep montane snow depth data. Hydrol. Process., 838, 829–838, https://doi.org/10.1002/hyp.6119, 2006.

Fassnacht, S. R., Brown, K. S. J., Blumberg, E. J., López-Moreno, J. I., Covino, T. P., Kappas, M., Huang, Y., Leone, V., and Kashipazha, A. H.: Distribution of snow depth variability, Front. Earth Sci., 12, 683–692, https://doi.org/10.1007/s11707-018-0714-z, 2018.

Fey, C., Schattan, P., Helfricht, K., and Schöber, J.: A compilation of multitemporal TLS snow depth distribution maps at the Weisssee snow research site (Kaunertal, Austria), Water Resour. Res., 55, 5154–5164, https://doi.org/10.1029/2019WR024788, 2019.

Fierz, C., Armstrong, R. L., Durand, Y., Etchevers, P., Green, E., McClung, D. M., Nishimura, K., Satyawali, P. K., and Sokratov, S. A.: The International Classification for Seasonal Snow on the Ground, IHP-VII Tech. Doc. Hydrol. 83, 2009.

Gardelle, J., Berthier, E., Arnaud, Y., and Kääb, A.: Regionwide glacier mass balances over the Pamir-Karakoram-Himalaya during 1999–2011, The Cryosphere, 7, 1263–1286, https://doi.org/10.5194/tc-7-1263-2013, 2013.

Gascoin, S., Grizonnet, M., Bouchet, M., Salgues, G., and Hagolle, O.: Theia Snow collection: high-resolution operational snow cover maps from Sentinel-2 and Landsat-8 data, Earth Syst. Sci. Data, 11, 493–514, https://doi.org/10.5194/essd-11-493-2019, 2019.

Girod, L., Nuth, C., Kääb, A., Mcnabb, R., and Galland, O.: MMASTER?: Improved ASTER DEMs for Elevation Change Monitoring, Remote Sens., 9, 704, https://doi.org/10.3390/rs9070704, 2017.

Grizonnet, M., Michel, J., Poughon, V., Inglada, J., Savinaud, M., and Cresson, R.: Orfeo ToolBox: open source processing of remote sensing images, Open Geospatial Data, Softw. Stand., 2, 15, https://doi.org/10.1186/s40965-017-0031-6, 2017.

Harder, P., Schirmer, M., Pomeroy, J., and Helgason, W.: Accuracy of snow depth estimation in mountain and prairie environments by an unmanned aerial vehicle, The Cryosphere, 10, 2559–2571, https://doi.org/10.5194/tc-10-2559-2016, 2016.

Hirschmüller, H.: Accurate and efficient stereo processing by semi-global matching and mutual information, Proc. – 2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition, CVPR 2005 II, 807–814, https://doi.org/10.1109/CVPR.2005.56, 2005.

Höhle, J. and Höhle, M.: Accuracy assessment of digital elevation models by means of robust statistical methods, ISPRS J. Photogramm. Remote Sens., 64, 398–406, https://doi.org/10.1016/j.isprsjprs.2009.02.003, 2009.

Howat, I. M., Porter, C., Smith, B. E., Noh, M.-J., and Morin, P.: The Reference Elevation Model of Antarctica, The Cryosphere, 13, 665–674, https://doi.org/10.5194/tc-13-665-2019, 2019.

Lacroix, P.: Landslides triggered by the Gorkha earthquake in the Langtang valley, volumes and initiation processes, Earth Planet. Space, 68, 46, https://doi.org/10.1186/s40623-016-0423-3, 2016.

López-Moreno, J. I., Fassnacht, S. R., Beguería, S., and Latron, J. B. P.: Variability of snow depth at the plot scale: implications for mean depth estimation and sampling strategies, The Cryosphere, 5, 617–629, https://doi.org/10.5194/tc-5-617-2011, 2011.

Margulis, S. A., Fang, Y., Li, D., Lettenmaier, D. P., and Andreadis K.: The utility of infrequent snow depth images for deriving continuous space-time estimates of seasonal snow water equivalent, Geophys. Res. Lett., 46, 5331–5340, https://doi.org/10.1029/2019GL082507, 2019.

Marti, R., Gascoin, S., Berthier, E., de Pinel, M., Houet, T., and Laffly, D.: Mapping snow depth in open alpine terrain from stereo satellite imagery, The Cryosphere, 10, 1361–1380, https://doi.org/10.5194/tc-10-1361-2016, 2016.

McGrath, D., Webb, R., Shean, D., Bonnell, R., and Marshall, H. P.: Spatially Extensive Ground - Penetrating Radar Snow Depth Observations During NASA's 2017 SnowEx Campaign?: Comparison With In Situ, Airborne, and Satellite Observations, Water Resour. Res., 55, 10026–10036, https://doi.org/10.1029/2019WR024907, 2019.

Miles, E. S., Watson, C. S., Brun, F., Berthier, E., Esteves, M., Quincey, D. J., Miles, K. E., Hubbard, B., and Wagnon, P.: Glacial and geomorphic effects of a supraglacial lake drainage and outburst event, Everest region, Nepal Himalaya, The Cryosphere, 12, 3891–3905, https://doi.org/10.5194/tc-12-3891-2018, 2018.

Nolan, M., Larsen, C., and Sturm, M.: Mapping snow depth from manned aircraft on landscape scales at centimeter resolution using structure-from-motion photogrammetry, The Cryosphere, 9, 1445–1463, https://doi.org/10.5194/tc-9-1445-2015, 2015.

Nuth, C. and Kääb, A.: Co-registration and bias corrections of satellite elevation data sets for quantifying glacier thickness change, The Cryosphere, 5, 271–290, https://doi.org/10.5194/tc-5-271-2011, 2011.

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., Hedrick, A., Joyce, M., Laidlaw, R., Marks, D., Mattmann, C., McGurk, B., Ramirez, P., Richardson, M., Skiles, S. M. K., Seidel, F. C., and Winstral, A.: The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo, Remote Sens. Environ., 184, 139–152, https://doi.org/10.1016/j.rse.2016.06.018, 2016.

Painter, T. H., Bormann K., Deems, J. S., Hedrick, A. R., Marks, D. G., Skiles, M., and Stock, G. M.: Through the Looking Glass:

Droughtorama to Snowpocalypse in the Sierra Nevada as studied with the NASA Airborne Snow Observatory, AGU Fall Meeting Abstracts, C12C-08, 2017.

Piermattei, L., Marty, M., Karel, W., Ressl, C., Hollaus, M., Ginzler, C., and Pfeifer, N.: Impact of the Acquisition Geometry of Very High-Resolution Pléiades Imagery on the Accuracy of Canopy Height Models over Forested Alpine Regions, Remote Sens., 10, 1542, https://doi.org/10.3390/rs10101542, 2018.

Prokop, A., Schirmer, M., Rub, M., Lehning, M., and Stocker, M.: A comparison of measurement methods?: terrestrial laser scanning , tachymetry and snow probing for the determination of the spatial snow-depth distribution on slopes, Ann. Glaciol., 49, 210–216, 2008.

Redpath, T. A. N., Sirguey, P., and Cullen, N. J.: Repeat mapping of snow depth across an alpine catchment with RPAS photogrammetry, The Cryosphere, 12, 3477–3497, https://doi.org/10.5194/tc-12-3477-2018, 2018.

Roche, J. W., Rice, R., Meng, X., Cayan, D. R., Dettinger, M. D., Alden, D., Patel, S. C., Mason, M. A., Conklin, M. H., and Bales, R. C.: Climate, snow, and soil moisture data set for the Tuolumne and Merced river watersheds, California, USA, Earth Syst. Sci. Data, 11, 101–110, https://doi.org/10.5194/essd-11-101-2019, 2019.

Rolstad, C., Haug, T., and Denby, B.: Spatially integrated geodetic glacier mass balance and its uncertainty based on geostatistical analysis: Application to the western Svartisen ice cap, Norway, J. Glaciol, 55, 666–680, https://doi.org/10.3189/002214309789470950, 2009.

Rupnik, E., Pierrot-Deseilligny, M., and Delorme, A.: 3D reconstruction from multi-view VHR-satellite images in Mic-Mac, ISPRS J. Photogramm. Remote Sens., 139, 201–211, https://doi.org/10.1016/j.isprsjprs.2018.03.016, 2018.

Shaw, T. E., Gascoin, S., Mendoza, P. A., Pellicciotti, F., and McPhee, J.: Snow depth patterns in a high mountain Andean catchment from satellite optical tri-stereoscopic remote sensing, Water Resour. Res., 56, e2019WR024880, https://doi.org/10.1029/2019WR024880, 2019.

Shean, D. E., Alexandrov, O., Moratto, Z. M., Smith, B. E., Joughin, I. R., Porter, C., and Morin, P.: An automated, open-source pipeline for mass production of digital elevation models (DEMs) from very-high-resolution commercial stereo satellite imagery, ISPRS J. Photogramm. Remote Sens., 116, 101–117, https://doi.org/10.1016/j.isprsjprs.2016.03.012, 2016.

Sturm, M. and Holmgren J.: An Automatic Snow Depth Probe for Field Validation Campaigns, Water Resour. Res., 54, 9695–9701, https://doi.org/10.1029/2018WR023559, 2018.

Zhou, Y., Parsons, B., Elliott, J. R., Barisin, I., and Walker, R. T.: Assessing the ability of Pleiades stereo imagery to determine height changes in earthquakes: A case study for the El Mayor-Cucapah epicentral area, J. Geophys. Res.-Sol. Ea., 120, 8793–8808, https://doi.org/10.1002/2015JB012358, 2015.