Answer to Y. Bühler

The authors would like to thank Yves Bühler for his review. We provide below a point-by-point response to his comments and we explain how we intend to modify the manuscript in order to take them into account.

Y.B . 1. Error combination of snow-on and snow-off DSMs: In their study they compare HS maps form Pléiades snow-on minus Pléiades snow-off with HS maps generated from LiDAR snow-on minus LiDAR snow-off. Certainly, this comparison makes sense and provides interesting insights but only as a second step. First, we need to see how much error is coming from individual surface models. If you take the snow-off and the snow-on DSM from the same platform and processing the errors of the individual DSMs could a) be added to each other or b) erase each other. So, if the snow-on DSM has a high error on one side (e.g. being too high) and the snow-off DSM has the error to the same side, the resulting snow depth is accurate even though both DSMs are bad. In the comparison as it is done now in the paper, there is now chance to see the quality of the individual DSMs. Therefore, I request the authors to do a HS comparison applying also the LiDARsnow-off DSM for the Pléiadessnow-on DSM. This will reveal how the snow-on DSM performs excluding random error addition from the snow-off DSM. Also, a DSM comparison between the Pléiades snow-off and the LiDAR snow-off DSM should be performed and discussed."

We agree that an individual evaluation of each DSM can help understand the sources of error in the HS maps. We planned to compare Pléiades DEMs with ASO DSMs in order to identify how each dataset compare separately in snow-on and snow-off condition. The provider of the ASO data (JPL) could not provide us with the DSMs within the time of this review. Therefore we calculated a snow-on ASO DSM by adding the snow-off DTM and the ASO HS map. We used the Pléiades DEMs after the co-registration processing described in the present manuscript. We evaluated the areas labeled as snow at the snow-on date.

We include the figure S2, the table S3 in the supplement and this discussion in the 6.4 of the manuscript (L448):

"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 co-registered as described in 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 two 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 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 are indeed over-estimating the surface elevation as the terrain slope increases (Figure S3). This suggests that combining satellite photogrammetry and airplane lidar DEMs may lead to larger errors than keeping homogeneous sourced DEMs."

Table S3. Comparison of the snow depth residual (HS Pléiades minus HS ASO) and stable terrain elevation difference (Pléiades). All metrics are in meters. The bold line is common to this table and Table 2.

	Area (km²)	Mean	Median	NMAD	RMSE	standard deviation
difference of elevation difference ($HS_{Pléiades}$ minus HS_{ASO})	**138.02**	**+0.08**	**+0.10**	**0.69**	**0.80**	**0.79**
difference of elevation SNOW OFF ($DEM_{Pléiades}$ minus DEM_{ASO})	138.02	+0.13	+0.01	0.80	0.96	0.95
difference of elevation SNOW ON ($DEM_{Pléiades}$ minus DEM_{ASO})	138.02	+0.21	+0.13	0.93	1.09	1.07

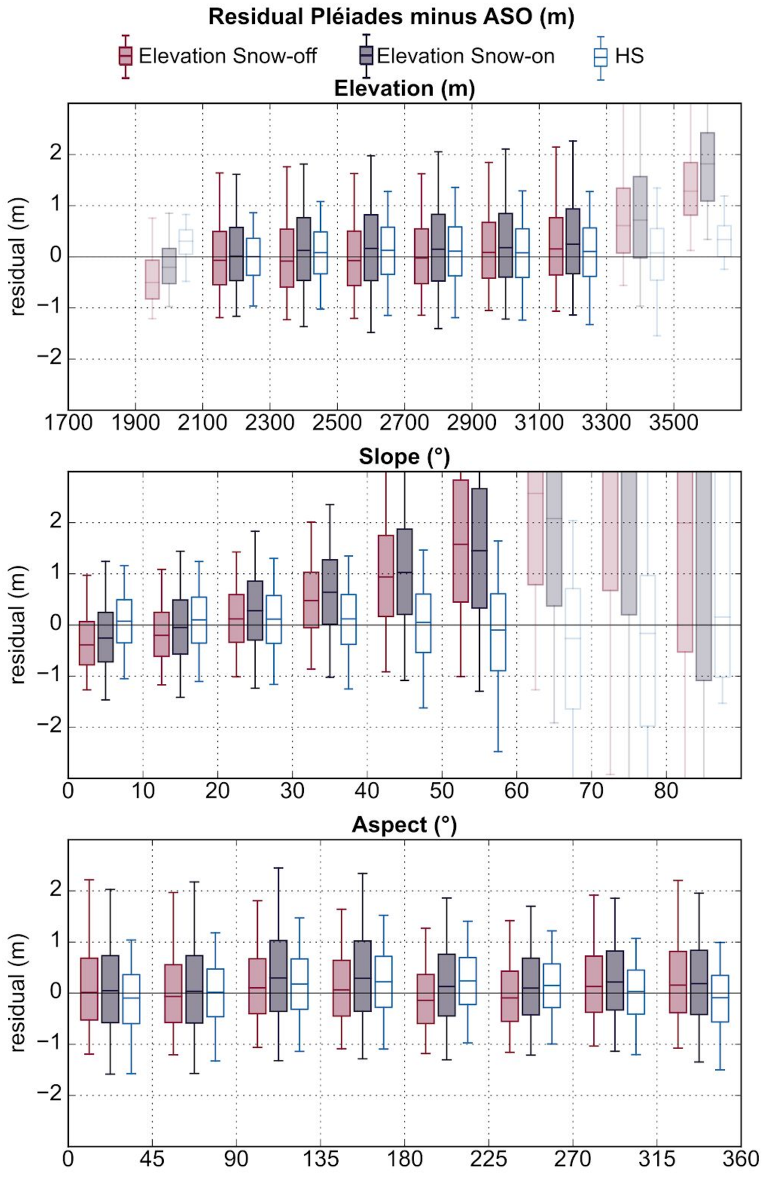


Fig S2. Snow-off DEMs difference,Snow-on DEMs difference, HS difference (Pléiades minus ASO) against elevation (top), slope (middle) and aspect (bottom).

2. Justification of selected photogrammetric processing method: The authors select three different processing methods in ASP (SGM-binary, SGM-ternary and Local search). But they do not justify why they selected these algorithms nor do they describe the algorithms and their differences. Are there other options? An overview on the available algorithms including their strength and weaknesses would be very helpful for the readers and would allow for the justification of the selection. Why does the SGM binary perform best? This section has to be expanded.

Many more stereo algorithms and options are available in ASP to process stereo images. In fact we tested around 1500 combinations of six options in a small sub-zone of the study area in order to determine an optimal set of options. This analysis is computer-demanding and complex because all options are not compatible. In addition, the results for a small sub-zone may not apply to the entire study area. Since we could not find a clear message from this analysis, we decided not to publish it at this point. From these experiments we still learnt that the choice of the correlation algorithm and the cost function (binary census transform or ternary census transform) had a noticeable impact on the quality of the DEMs and HS maps. The comparison of these three sets of options already enables to highlight the impact of processing options on the HS maps. Future work will hopefully address more thoroughly this issue.

We added this L160:

"*These set of options were empirically selected but do not cover all the options available in ASP.***"**

Explaining why one option performs better is beyond the scope of this study. However, we added some additional informations about the stereo algorithm and options L172:

"The SGM algorithm (Hirschmüller, 2005) differs from 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: 00,01,11 depending on whether it is smaller, within, or greater than a buffer centered on the central pixel value."

3. Comparison to results from other photogrammetric platforms (UAS and airplanes) and discussion of photogrammetry specific issues: The embedding of the work into the current literature is weak (section 6.1). It is true that they are among the first mapping snow depth with optical satellite data. But there has been a lot of additional work concerning photogrammetric snow depth mapping from UAS platforms (e.g. Van der Jagt et al. 2015, Bühler et al. 2016, De Michele et al. 2016, Harder et al. 2016, Adams et al. 2018). The authors should compare their results also to these publications and discuss more about the specific problems of photogrammetry on homogenous snow-covered areas (e.g. Bühler et al. 2017) and the effect of vegetation. I assume that there is distinct problem with noisy surfaces. This topic is not sufficiently discussed in the paper.

We added references to UAVs work in the introduction (Bühler et al., 2016, De Michele et al., 2016, Harder et al., 2016, Redpath et al., 2018) and we added the paragraph below in the Discussion about the comparison between satellite photogrammetry and other airborne methods (airplane, UAV). However we feel that it is difficult to go beyond these generalities in our

manuscript. We prefer to refer to Eberhard et al. (2020) who made a detailed comparison of photogrammetric approaches using datasets from different instruments. See L506:

"HS maps from UAV SfM typically exhibit a centimetric bias (0.05 m to 0.11 m) and a RMSE between 0.05 m and 0.30 m based on comparison with snow probe and GNSS measurements. This is more accurate than what is currently achieved with satellite photogrammetry. However, UAV campaigns are currently limited to areas of less than 1 km² due to battery limitation and often relied 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 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."

We also addressed the weakness of the embedding in the current literature by extending the section 6.1. We now include comparison of our results to McGrath et al. (2019), Shaw et al. (2019), Eberhard et al. (in TC discussion). These papers used satellite photogrammetry to measure at least the winter DEM. See L355 :

*"**6.1 Comparison to others studies using satellite photogrammetry***

By comparing the Pléiades HS with the ASO data, we find a 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 km² that was obtained by UAV photogrammetry. The spread in accuracy between studies in Tab. 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 probes 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 range 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) which calculated a HS map from a Pléiades snow-on DEM and an airplane SfM snow-off DEM and compared it to HS from airplane SfM only over 75 km² (NMAD=0.65 m). Finally, McGrath et al., (2019) found a 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 km². This lower NMAD value might result from the higher resolution of the WorldView-3 images (0.3 m). 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 and 0.35 m."

Table 4. *Comparison of HS accuracy with studies using satellite photogrammetry. *Eberhard et al. used a Pléiades DEM for snow-on and UAV or airplane SfM DEM for snow-off.*

	Satellite (resolution)	HS map resolution	Validation data	Area	#	Mean	Median	NMAD	RMSE
This study	Pléiades (0.5 m)	3 m	Airplane lidar	138 km²		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 km²		-0.06	-0.14	0.78	
McGrath et al. 2019	WorldView (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 km²		-0.10	-0.22	0.36	0.52
Eberhard et al.*	Pléiades (0.5 m)	2 m	Snow probing		36	-0.35	-0.36	0.47	0.52
			UAV SfM	4 km²		-0.18	-0.18	0.38	0.44
			Airplane SfM	75 km²		-0.02	-0.18	0.65	0.92

YB. 4. Discussion on which applications can benefit from this approach and which not: With the elaborated accuracies in mind, there should be a discussion on which applications could benefit from this new method and which not. For example, regions with shallow snow cover (0 – 50 cm) over the vast regions on the higher latitudes of the northern hemisphere would be within the error range. So, I assume the method would be not accurate enough. But for what applications could it be of big value? I think this would be a very interesting part for the readers.

As stated in the title of the article we think that this approach is mostly beneficial for mountainous terrain. However, we agree that we could further discuss its applicability in other contexts. We propose to include this paragraph (L515):

"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 co-register 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 (roughly HS below 50 cm, e.g. polar environments) might not be as straightforward as the typical spatial variability lays within our range of uncertainty (roughly 0.50 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 co-registration process in some regions. The horizontal component of the co-registration 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. 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 multi-spectral 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. Fig. 10 suggests that this variation is likely to become insignificant if the HS map is aggregated at a larger spatial scale (grids spacing larger than 100 m x 100 m). Such optimisation 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 co-registration step."

Detailed comments:

P3L65: Why are DSMs less accurate in steep terrain? Is it mainly because a small x,y shift results in a large z-shift? Please explain a bit more

DSM are expected to be less accurate in steep terrain because a small error in horizontal positioning results in a large vertical error. Horizontal errors are inherent to the photogrammetric process (see below the answer to comment about P13L325). Steep slopes might also be more prone to error as the image distortion also depends on the local incidence angle. Two images will look more similar on areas taken with small incidence angles than large incidence angles. Image matching process should be more accurate when images are more similar.

We modified the text for (L66):

"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)."

P3L76: the term airborne lidar altimetry sound a bit strange to me, I would use airborne laser scanning ALS.

We agree with the reviewer and replaced references to airborne laser scanning with ALS.

P3L87: Please give some references for this statement

We deeply modified this paragraph as we now evaluate a similar but well-accepted error model (Rolstad et al., 2009) L243:

"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 a single pixel accuracy but depends on the spatial correlation of the errors (Rolstad et al., 2009)."

P4L103: Why did you select exactly this zone? Can you give some justification?

We selected this zone to ensure a large overlap between Pléiades and ASO coverage while including a large range of elevation. We modified the text (L92):

"The ASO flights cover 1100 km² in the basin while this study focuses on a 280 km² subzone which was selected to cover a large elevation range."

P4L118: How was the image acquisition performed? What prices did you have to pay? What would be the conditions for people who want to do the same as you did?

We would now include this paragraph L117:

"Pléiades images were obtained at no cost through the DINAMIS program (https://dinamis.teledetection.fr/) opened to Europeans scientists working in public research institutions. Otherwise Pléiades images can be ordered from Airbus Defense and Space."

We prefer not to indicate the commercial cost of Pléiades imagery because it may be subject to change. The interested user can easily contact ADS to request a quote.

P5L135: Why is the resolution 3 m is there a justification for that? What was the point density per m2 of the LiDAR dataset? Double period at the end of the sentence.

We used ASO products described in Painter et al. (2016) and could not chose the HS map resolution. The 3m resolution balances computational efficiency and the lidar point density,

which varies from just under 1 pt/m2 in the worst case to > 10pts/m2 at highest terrain elevations, due to the constant flight altitude of the ASO aircraft. We added the reference to Painter et al. (2015) to make clear that we did not process these data L127:
"Ground points are aggregated to a 3 m grid to derive a gridded DEM (Painter et al., 2015)."

P7L173: In my experience it can be very hard to find snow free, stable terrain in high alpine regions and this with a good distribution over the entire investigation area. Please discuss this issue and give some recommendations what to do if not enough stable terrain can be found. Please also discuss why an absolute orientation via GCPs is not suitable.
We agree that stable terrain availability can be a limitation to this method. We identified areas with bare rock or grass as stable terrain because it was available in this study area. However this type of terrain is not necessarily always available in both snow-on and snow-off images. As in Marti et al. (2016), the horizontal component of the co-registration vector can be measured without differencing the stable or snow covered terrain. The vertical component of the co-registration vector should ideally be measured on well-distributed stable terrain (this study), but if this is not available, can be performed using stable terrain in a specific location of the images (e.g. a football field in Marti et al., 2016, snow-free road). If no stable terrain is available one could think of using snow depth measured in the field (snow probe, weather station).
The use of accurate and well-distributed GCPs improves the accuracy of individual DEMs (Berthier et al., 2014). However, we did not explore how many GCPs and what precision would be necessary to measure snow depth. The clear advantage of a method without GCP is that it can be readily applied elsewhere on Earth.
We added this paragraph in the discussion to comment on that point (L527):
"A lack of well distributed stable terrain in snow-on and snow-off DEMs can complicate the co-registration process in some regions. The horizontal component of the co-registration 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. 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. "

P7L176: Why did you choose the thresholds -1 to 30 m? Please justify
These thresholds are used to remove obvious outliers in the DEM difference map. Minus one (-1 m) is set to take into account that Pléiades elevation difference might be negative despite presence of snow due to its accuracy. Plus thirty (+30 m) is a conservative estimate based on expert judgment of the snow depth upper bound. We added this L207:
"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."

P7L182: In my experience a big problem are alpine bushes (0.5 to 2 m high). How do you treat these? It is not mentioned
We did not apply specific treatment for bushes, their impact is included in our residual calculation. We expanded comments on the impact of vegetation on our evaluation L425:
"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 co-registration. This bias is low given the differences in the characteristics of the ASO and the Pléiades products. It 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 do 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.”

P7L190: Why do you use nearest neighbor and not cubic resampling? Please justify.

We used nearest neighbour interpolation for the masks as they are binary, each pixel being one or zero. Other interpolation schemes (cubic, linear) would require another step of thresholding to obtain again a binary mask.

P10L244: Artifacts in the equations, please check all equations

The equations were replaced and should now be readable.

P11L266: It would be very helpful for the readers to have a figure showing the matching success and the hillshade of the DSMs generated with the different pairs This would also clarify the artifacts you mention.

The figure below shows the hillshade obtained with the different processing options, highlighting the described artifacts. We will add it in as a supplement to a revised version of this manuscript.

We unfortunately did not keep intermediate products of ASP, including the matching success maps. However, from our experience hillshade maps are more informative than matching success maps to visualize the impact of ASP configuration like these artefacts. We are not sure if the added value of a figure showing the matching success justifies the computational cost of running ASP again but we are ready to do it if the referee thinks it is important. We also added a description of the artifact L289:

“Patches of typically 20 m x 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 (Figure S1).”

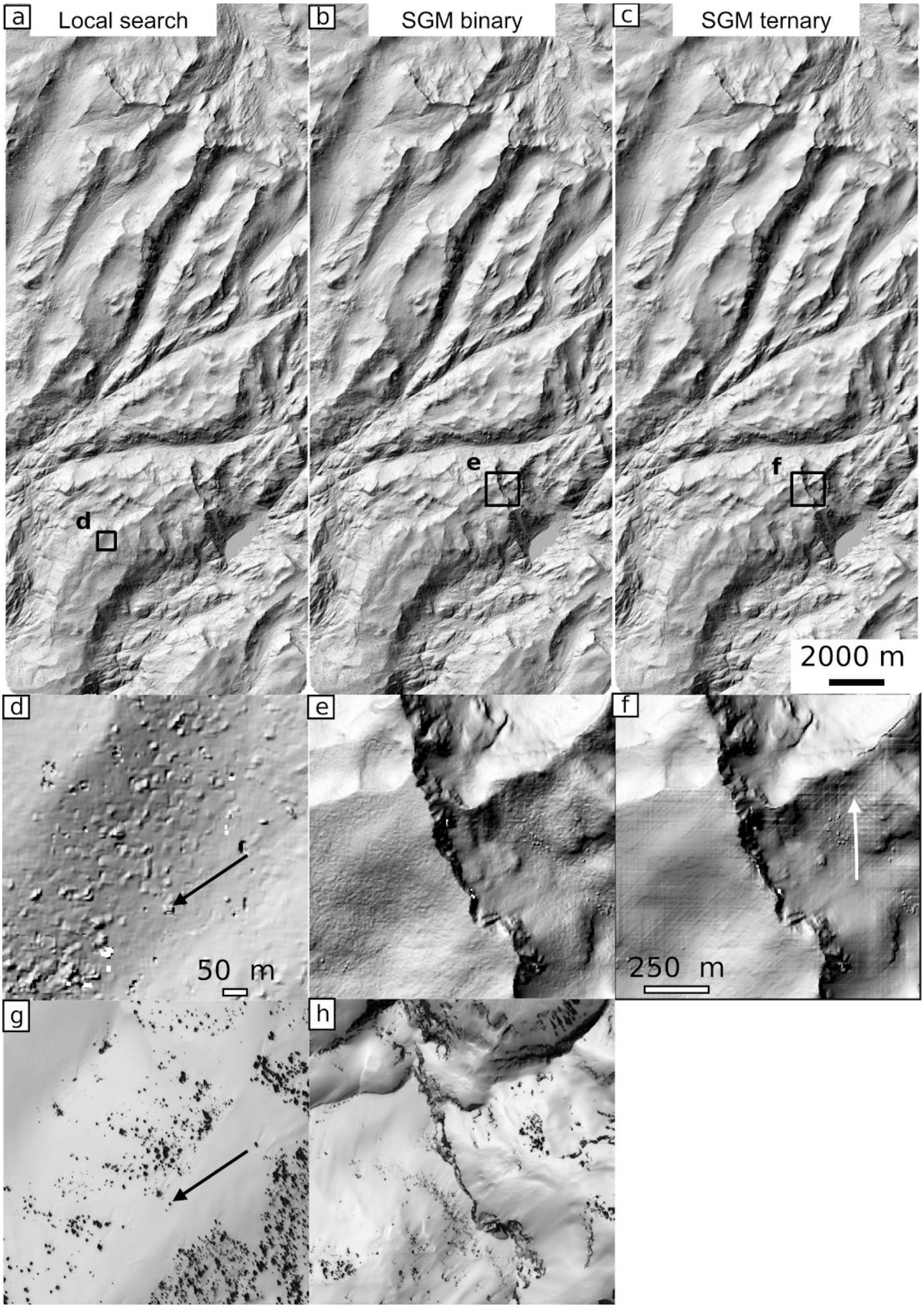


Figure S1. Hillshades of the snow-on Pléiades DEM calculated with the stereo options Local-Search (a), SGM-binary (b) and SGM-ternary (c). Artifacts around isolated tree were observed with Local-Search (arrow in hillshade (d) and pan-chromatic ortho-image (g)). Geometrical artifacts were observed in south facing slopes with SGM-ternary (arrow in hillshade (f)).

P12L296: Why do you use the threshold 3 x NMAD? Please justify.

This filtering is used to remove wrongly labeled pixels only in the co-registration process. It can for example happen that forest were included in stable terrain. If the residuals were following a Gaussian distribution, points further away than 3 x NMAD should occur less than 1% of the time. Hence such points are considered as very unlikely and excluded as outliers. This is an adaptation of the 3 x standard deviation filtering (Nuth and Kääb, 2011) with a metric adapted to elevation difference residual (NMAD, Höhle and Höhle, 2009). This type of filter is also sometimes used before averaging elevation aggregated by elevation but must be carefully used as it can eliminate valid extreme value and thus introduce bias (Dussaillant et al., 2019).

This is stated L298:

"This is expected as the same filtering is used during the co-registration process to remove outliers."

P13L325: Do you have a hypothes is where this error is coming from?

Random error could result from several phenomena. For example the fact that the images are a discrete representation of a continuous surface. This means that the images, the DEMs or intermediate products are necessarily interpolated during some operations. It is the case during the sub-pixel refinement in the images correlation. Another source of error is that the images are not corrected for the atmospheric or BRDF effect on the radiometry.

P13Table2: There is a 10 cm difference in RMSE and NMAD between front-nadir-back and nadir-front-back. Why is this showing, please explain as you take basically the same input data just in reversed order.

The processing in ASP depends on the order of the images. For a triplet of images (A-B-C), ASP calculates image matching for two pairs of images (A-B and A-C) defined by the order of the images. As the B/H ratio is different for each pair, the DEMs will be different depending on the order.

We expanded these explanation L188:

"In the tri-stereo case, ASP calculates two disparity maps and performs a joint triangulation when calculating the point-cloud. In the first tri-stereo case (front-nadir-back), 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), 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."

P14L346: Please extend this section substantially by discussing results from previous UAS and airborne investigations including the reached accuracies and performance on snow covered surfaces (see my main point 3)

We expanded this section as described in our answer to point 3.

P14L362: But please mention the benefit of additional coverage in particular within steep slopes.

We agree and added this L383:

"Tri-stereo might provide greater benefits in case of image occlusion in steep slopes, which is more prone to occur with higher B/H."

P14L366: Can you make the statement "B/H around 0.2 seems to be beneficial" stronger by providing more justification

We modified the text in L387:

"From these two studies and for similar terrain, a triplet of images with a B/H for consecutive images around 0.2 seems 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."

P15L400: what was calculated ASO minus Pléiades or the other way round? ASO should be higher as it uses a DTM as snow-free DSM

This is calculated as Pléiades minus ASO. As suggested ASO HS should be higher as it uses a DTM. This implies that there is a larger positive bias on Pléiades HS than what we measure with the HS comparison. This is confirmed by the comparison of snow-on DEMs which shows that Pléiades snow-on is on average 0.25 m above ASO snow-on over snow terrain (see above answer to main comment 1.).

P17L449: Please expand this section taking into account published UAS and airborne photogrammetric investigations

We expanded this section as described in our answer to point 3.

P18L473: The conclusions are a bit thin. Please also provide the RMES values. Could you provide more important information? The term "good accuracy" is vague.

We expanded the conclusion as follow L551:

"7 Conclusion

We found a good agreement between snow depth (HS) maps from high resolution stereo satellite images with ALS 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 show 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 cancel out in the HS map, leading to agreement between Pléiades and ALS 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 multi-view 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 the 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 maps resolution from 3 m 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."

P20Fig4: The scale bar is too large hindering a sound interpretation of the values. I would propose a range of 0 to >5 m. Here a difference image (HS_ASO minus HS_PLéiades would be essential. This would allow for a detailed interpretation of the results, also for the subsets. What you name an avalanche deposit in d) does not look like that to me at all. I saw many avalanche deposits in HS maps and I am pretty sure that this one is something else. Please check that. Why did you locate the transect there? Can you give a justification?

We provide below the Pléiades snow height map with a color bar ranging from 0 to 5 m. More areas appear saturated. We are not sure whether this makes the map more readable but would follow the editor's opinion.

The difference image is already provided in Fig. 8. It seems to us that figure 4 would get harder to read if a panel was added. We now refer to it in the caption of Fig. 4 so that reader can see the link.

We provide a zoom on what we called an avalanche in Fig. 3. We made this supposition based on the interpretation of the topography (steep couloir and flat area at the bottom), the pan-chromatic image (texture indicating dynamic deposition) and the snow height (accumulation at the bottom of the couloir and upward). We are still of the opinion that it is an avalanche. We agree that the arrow we positioned on the figure was misleading as it points at the couloir. We now moved it so that it points at the deposition area. Similar observation that we also interpreted as avalanche deposition are visible below in profile B.

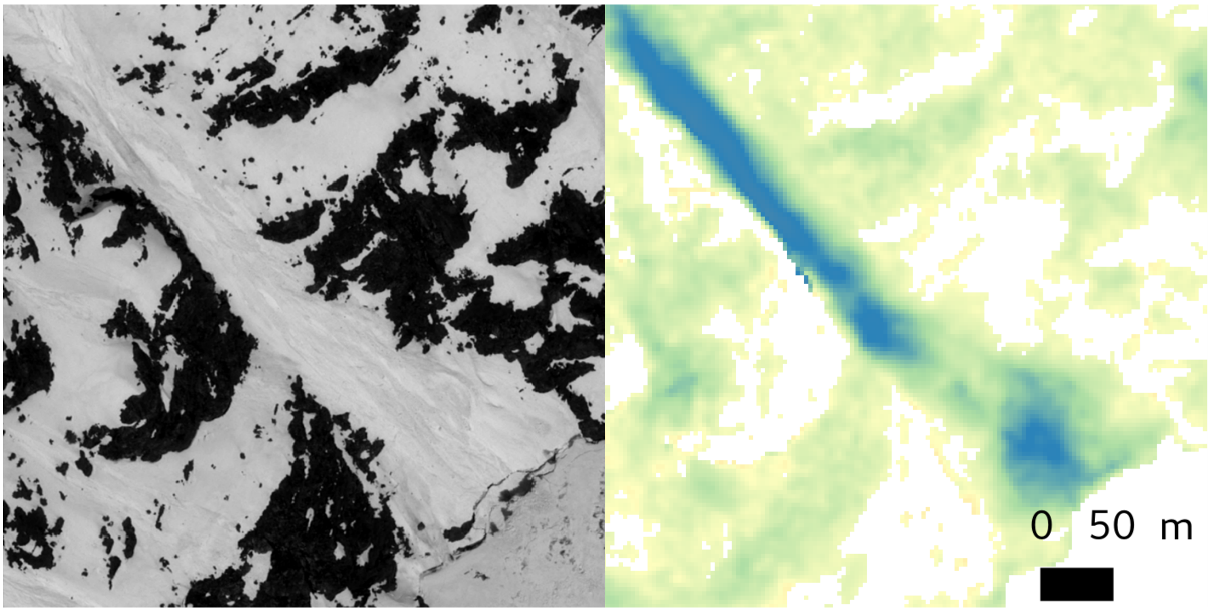


Fig 3. Zoom on the avalanche deposit in the pan-chromatic image (left) and snow depth map (right).

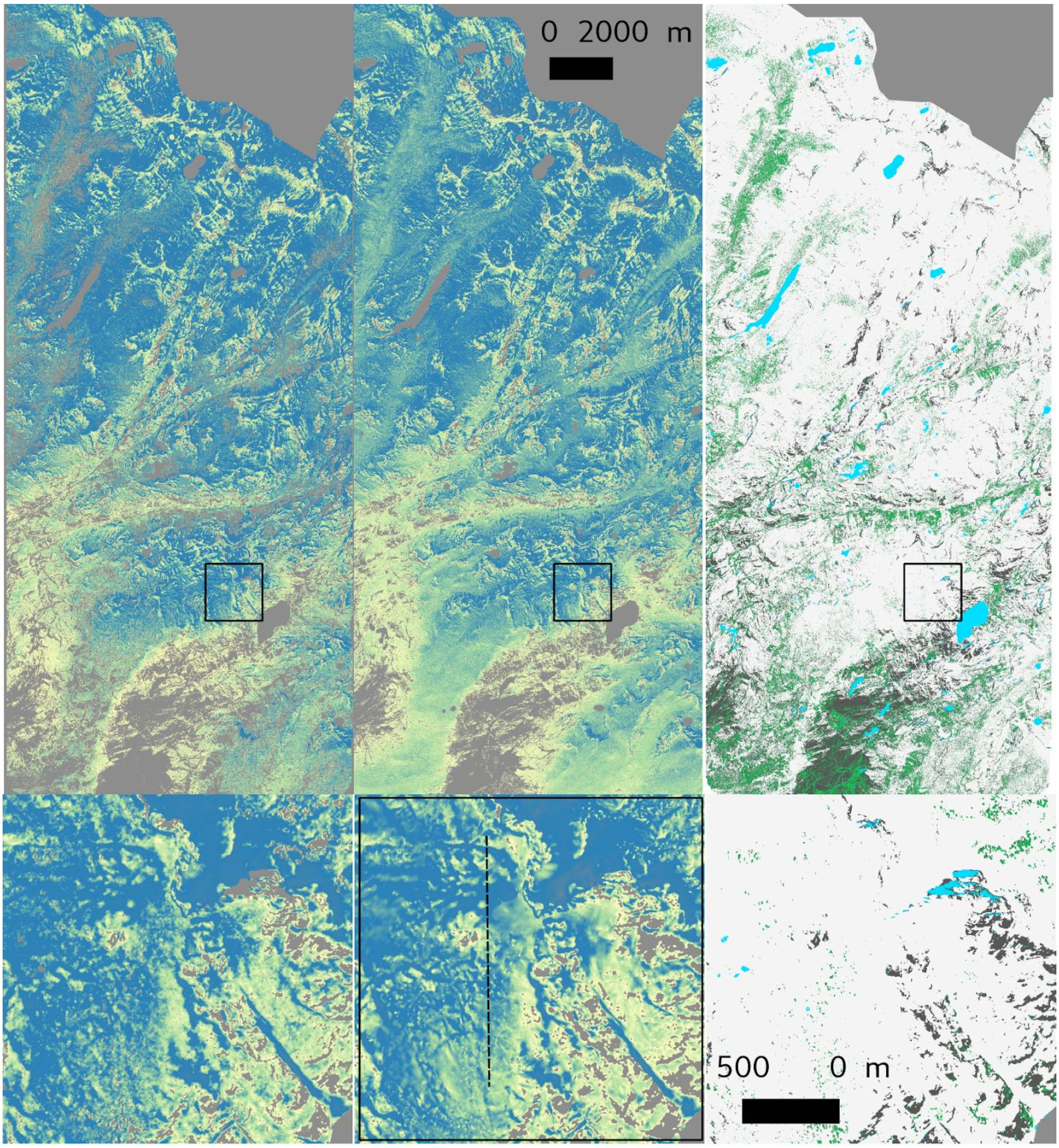


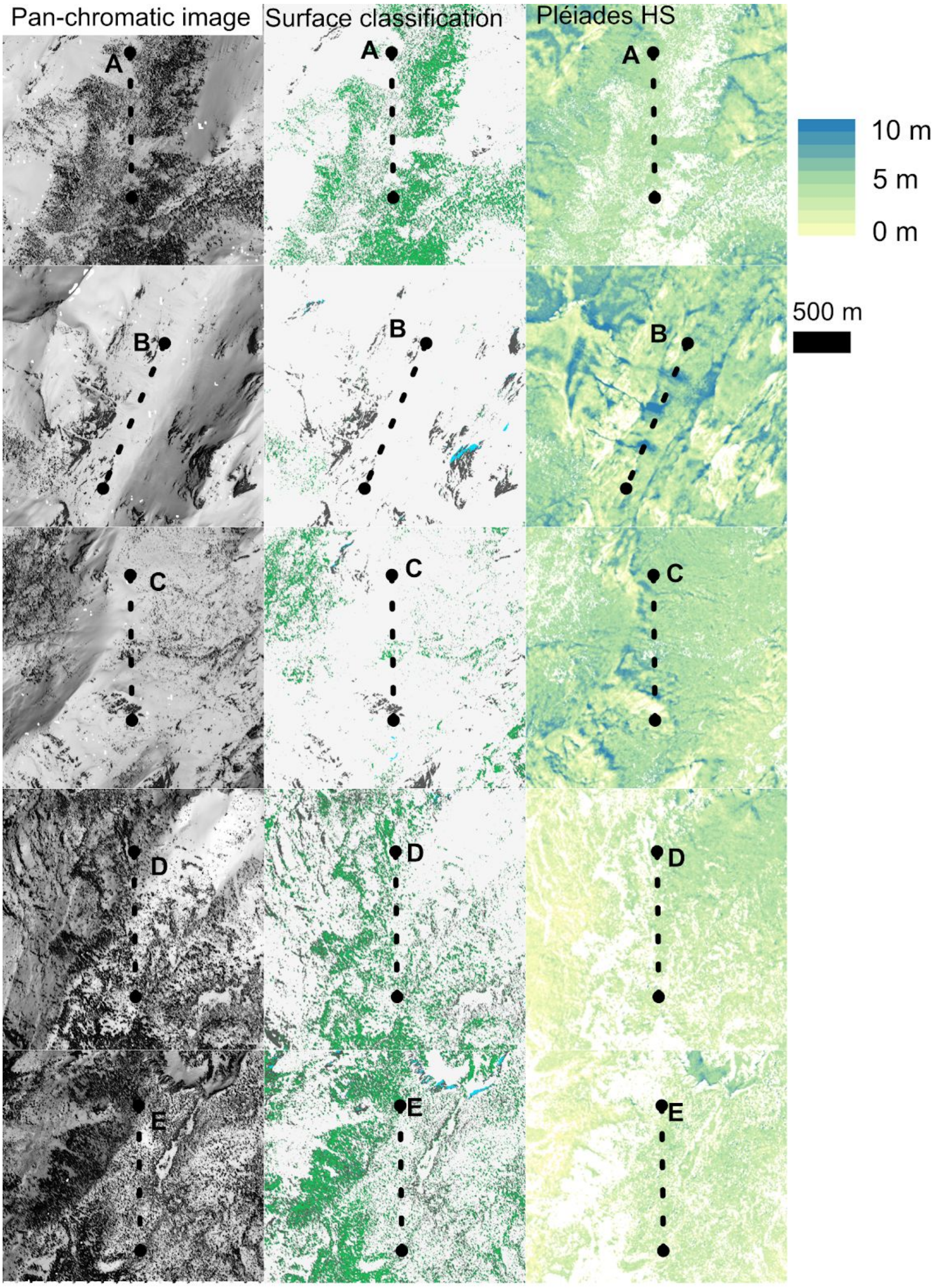
Fig. 4 with the colorbar updated to the range [0,5 m].

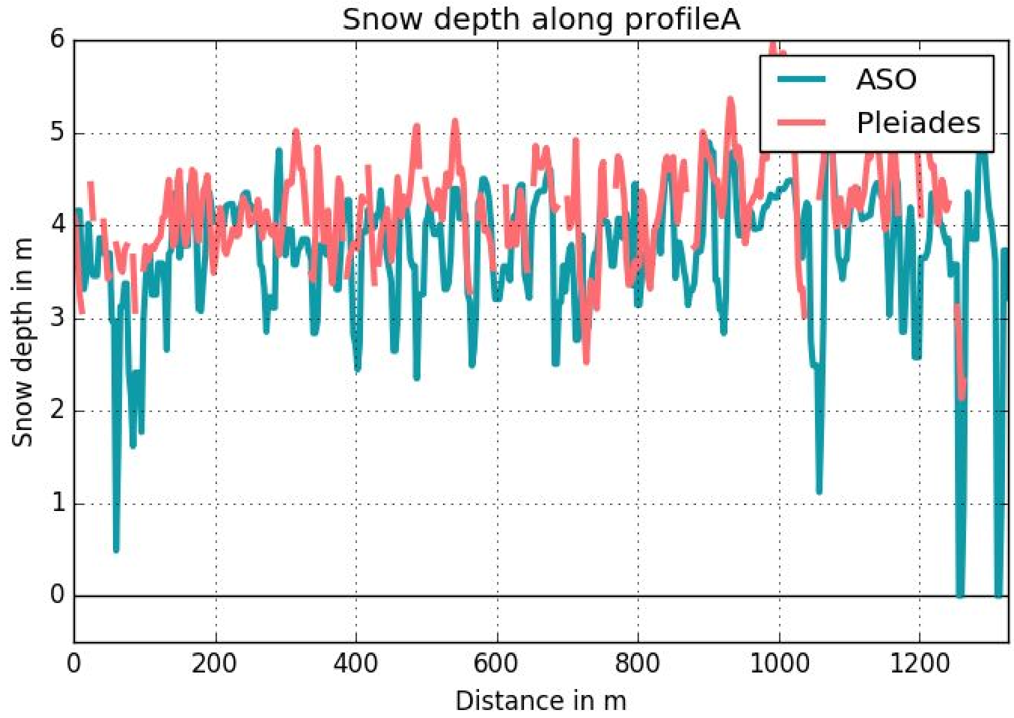
P21Fig5: Here it would help if you could draw some more transects at further locations to see the differences

We drew more profiles trying to cover different ranges of snow depth and vegetation condition. HS seems more noisy and less accurate in forest environment (profile A, D, E). Another case of what seems to be avalanches deposit is visible in profile B.

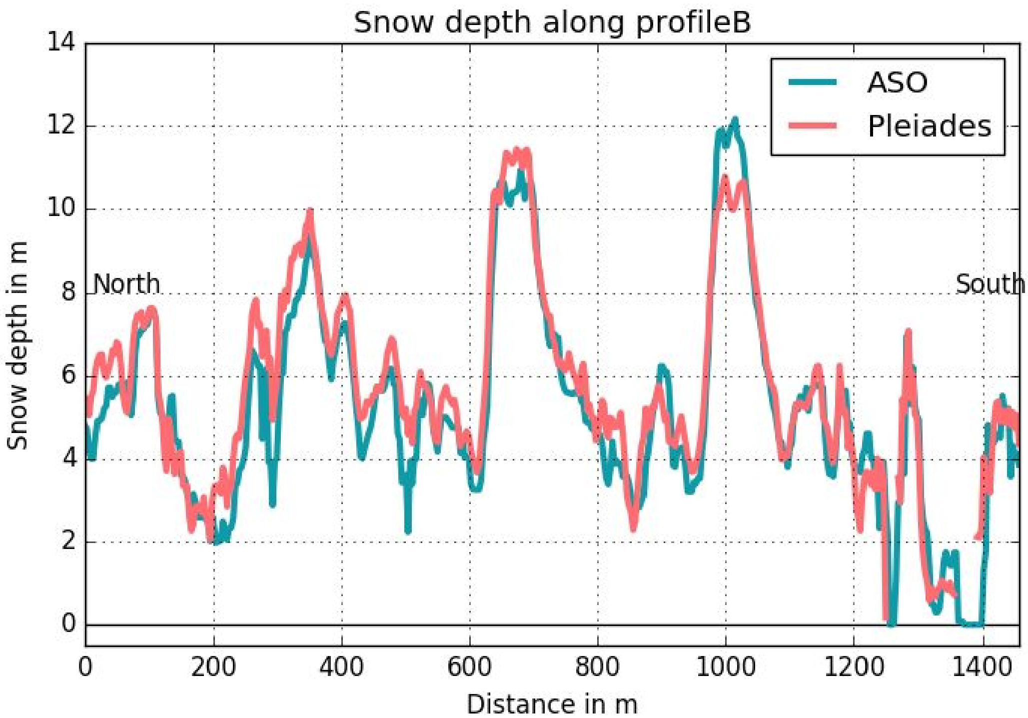
We propose to add these profiles in the supplement. We mention it in Fig. 5 caption L601:

*"**Figure 5.** Transect T-T' of snow depth visible on Figure 4. e. from Pléiades data (pink) and ASO (blue). More transects are available in supplement."*

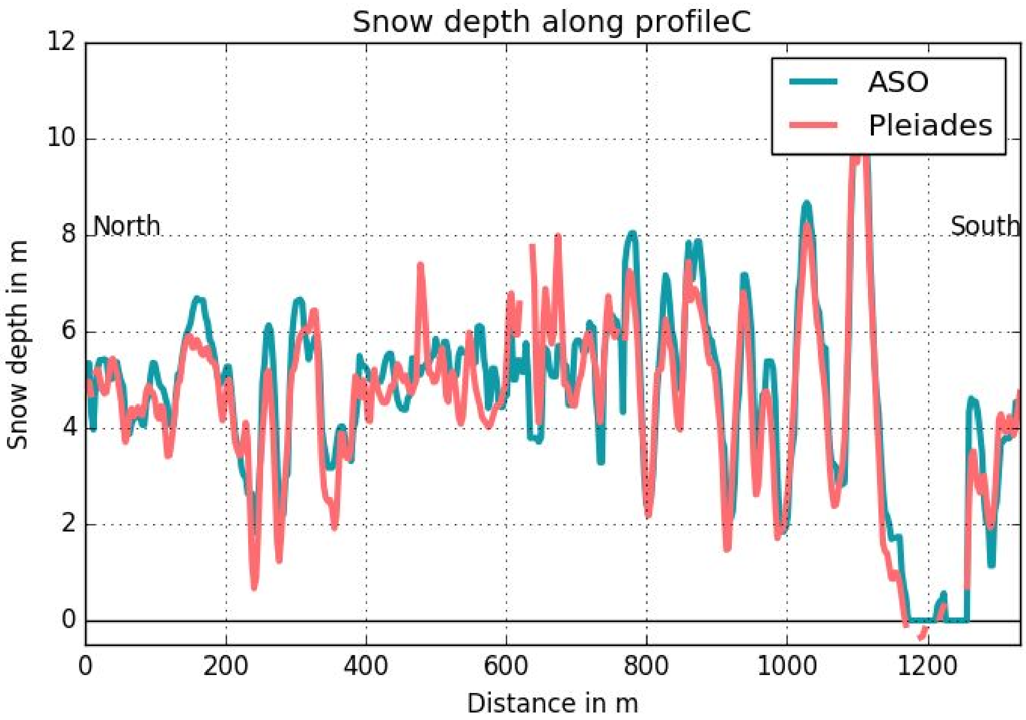


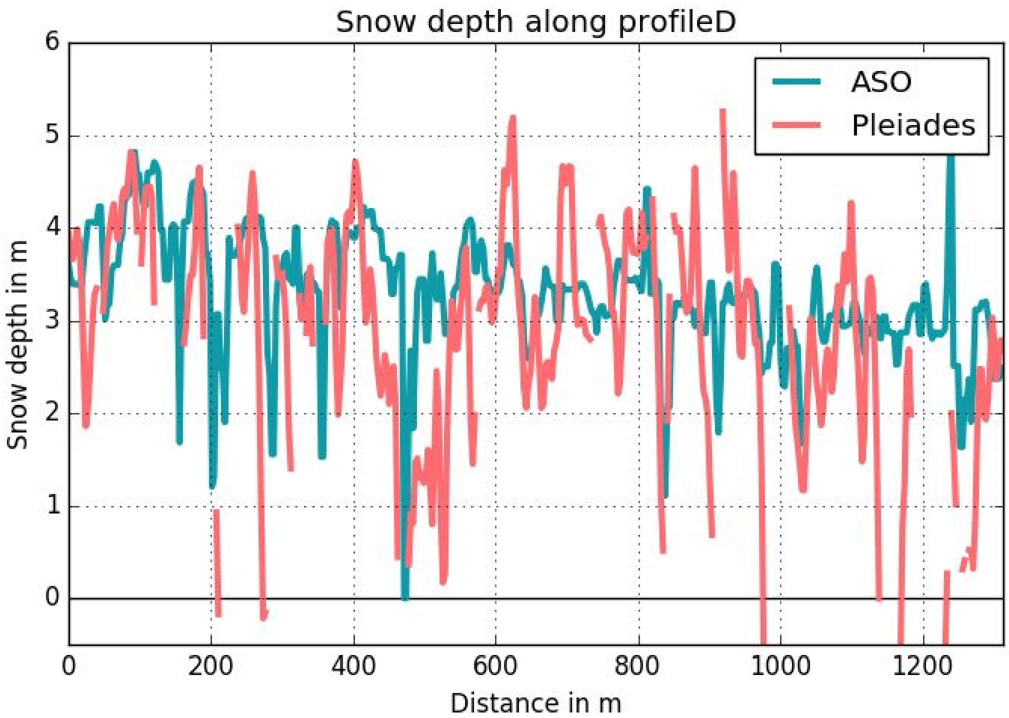


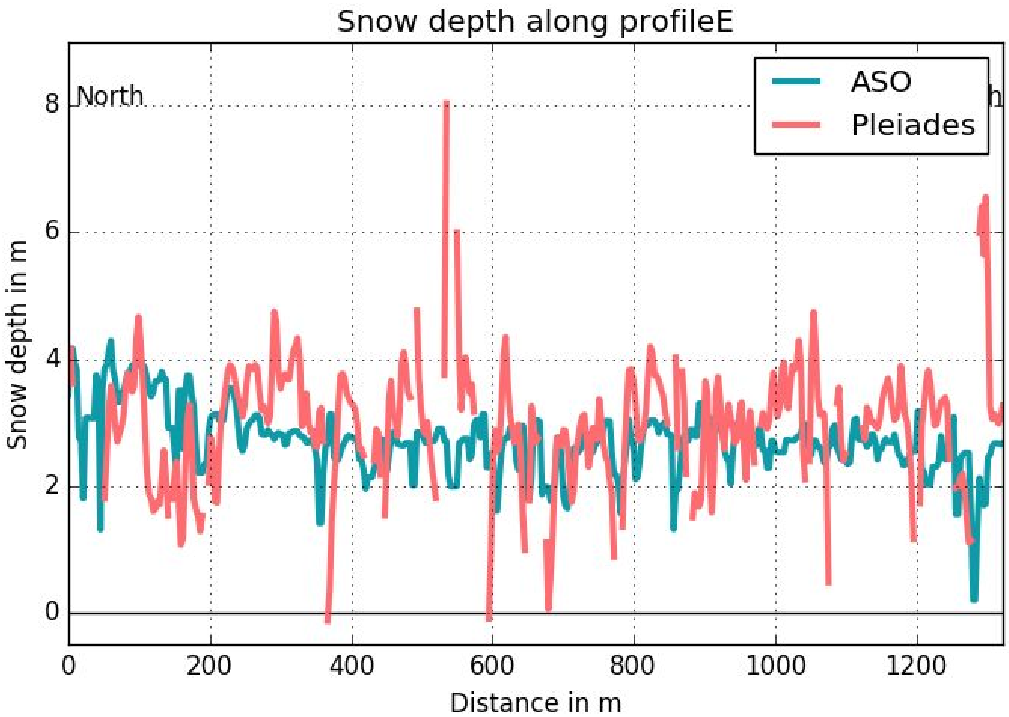
Snow depth along profileA
ASO
Pleiades
Snow depth in m
Distance in m
0
1
2
3
4
5
6
0
200
400
600
800
1000
1200



Snow depth along profileB
ASO
Pleiades
North
South
Snow depth in m
Distance in m
0
2
4
6
8
10
12
14
0
200
400
600
800
1000
1200
1400


Snow depth along profileC
ASO
Pleiades
North
South
Snow depth in m
Distance in m
12
10
8
6
4
2
0
0
200
400
600
800
1000
1200


Snow depth along profileD
ASO
Pleiades
Snow depth in m
Distance in m
0
1
2
3
4
5
6
0
200
400
600
800
1000
1200


Snow depth along profileE
North
ASO
Pleiades
Snow depth in m
Distance in m
0
2
4
6
8
0
200
400
600
800
1000
1200

P21Fig6: The ASO HS median is mostly lower than the Pléiade HS even though the ASO uses a DTM as snow-free surface (penetrating low vegetation). Therefore, we would expect the LiDAR HS to be slightly higher. Can you comment on that?

We agree with this suggestion. We did not take into account the fact that ASO snow-off elevation is a DTM while Pléiades snow-off elevation is a DSM although we do not know precisely which surface is represented by Pléiades elevation. Evaluation of the residual against vegetation height suggest that Pléiades do not represent the top of the vegetation at least for low vegetation. Further work should evaluate this point.

Following the suggestion of comparing individual DEMs (snow-on and snow-off) separately we conclude that Pléiades snow-on DEM bias of +0.21 m is partially compensated by the fact that we used a DSM for snow-off, resulting in a lower bias on HS +0.08 m. This is commented in answer to the point 1.

We also added this in the discussion L425:

"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 co-registration. This bias is low given the differences in the characteristics of the ASO and the Pléiades products. It 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 do 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."