# Answer to the review of the anonymous referee #1:

## General comments:

1. The current manuscript introduces the application of PRACTISE, but does not provide a detailed enough description to be able to fully understand how it works. Considering there are already two detailed papers on PRACTISE by the authors in GMD, I would hope citations to them could provide the reader with a satisfactory description. Then this manuscript could be refined to provide more focus on the results of the RCZ / VF comparison, and greater detail on the scaling question. Figures 2 through 5 could be removed to focus more on the results of this study (Fig 6 through 9) and expand analysis around figure 10 – which is of great interest, but under-analyzed in the current manuscript.

**Answer to 1.** We thank you for the comment and we will revise the paragraphs on the application of PRACTISE in this study to clarify the workflow used in this study by including more references to our previous papers on PRACTISE. However, we also think that the figures 2 to 5 are important here as they outline the workflow in PRACTISE graphically. This makes it simpler for non-experts to understand the processing steps taken in this specific study and it was a recommendation of the editor to our initial submission. Moreover, with these figures, readers are not obliged to read our previous papers where they probably face much more options and detail than they need. The second part of your comment is proposing to expand the analysis on the scaling effects between different NDSI thresholds. We agree with you on this point (see the answer to comment 2).

## Manuscript changes to 1.

p.6, l1 to p.7, l5: "In a first step information about the camera location and orientation was needed for georectification of the photography. This information was automatically optimized by using ground control points (GCPs, Fig. 4a; Sect. 3.3 in Härer et al., 2013). The calculated viewpoint and viewing direction were by default used to perform a viewshed analysis (Fig. 4b; Sect. 3.1 in Härer et al., 2013). The viewshed was needed for an identification of areas which were visible from the viewpoint and which were not obscured by topographical features or within a user-specified buffer area around the camera. The respective DEM pixels were then projected to the photo plane (Fig. 4c; Sect. 3.2 in Härer et al., 2013).

Now, the snow classification module was activated to distinguish between snow covered and snowfree DEM pixels (Fig. 4d). Two major procedures were available for classification: a statistical analysis which was using the blue RGB band (Salvatori et al. 2011; Sect. 3.4 in Härer et al., 2013) and a principal component analysis (PCA) based approach (Sect. 3.1 in Härer et al. 2016). The first was used for shadow-free scenes, the second for scenes with shaded areas. Section 3.4 in Härer et al. (2013) gives more insights into a third manual option if none of the two classification routines could be applied successfully. The photograph snow cover maps did have even in the case that an insufficient classification algorithm was used for a specific situation less than 5% misclassified pixels in the worst case region of the photograph in Chapt. 4 in Härer et al. (2013). It was also shown in an earlier publication that the classification of shadow-affected photographs are of the same quality as photographs without shadows (Chapt. 4 in Härer et al., 2016). As for this study, every classified image was visually inspected and no major snow classification errors comparable to our worst case example in the previous publication were found, we expect a relative misclassification error of 1%. For this example photograph, the snow classification algorithm utilizing a principal component analysis (PCA) was selected to account for the shadow-affected areas in the upper left part of the photograph (Fig. 4d, enlarged view in Fig. 4e).

After the photograph rectification and classification, the remote sensing routine of PRACTISE began with the identification of satellite pixels that spatially overlap with the photograph snow cover map (Sect. 3.2 in Härer et al., 2016). The used photograph and satellite image were thereby recorded within one (RCZ) to three (VF) hours. Moreover, a cloud- and shadow-free satellite image is generated by using fmask (Zhu et al., 2015). The needed NDSI map was calculated in accordance to Eq. (1) by PRACTISE (Fig. 5a).

If both, the NDSI satellite map and the corresponding high resolution photograph snow cover map were processed, an iterative calibration of the NDSI threshold value was started. The Landsat image was thereby resampled to the finer resolution of the photograph in the calibration to avoid losing any information by the aggregation of the photograph snow cover map. The best agreement between the local scale (photograph) and the large scale (Landsat) snow cover map was detected by maximizing the accuracy which is the ratio of identically classified pixels to the overall number of photograph-satellite image pixel pairs *n* (Aronica et al., 2002):

$$F = \frac{(a+d)}{n},\tag{2}$$

a represents the number of correctly identified snow pixels and d the same for no snow pixels. F is between 0 and 1 and becomes 1 for a perfect agreement between the two images.

The calibrated NDSI threshold was finally applied to the Landsat data with 30m pixel size to generate the final Landsat snow cover map. Figure 5b shows the resulting satellite snow cover map superimposed on the photograph snow cover map and a Landsat Look image. A cutout is shown for more detail in Fig. 5c. "

2. The comparison between RCZ / VF is robust and well quantified (Fig 7). The influence of rock reflectance is well described (Fig 8) and provides a valuable process basis to the quadratic relationships (Fig 9). However, the scaling question, while well illustrated by the example of 16 September (Fig 10a), suggests that all data have been used to create relationships presented in (Fig 10c). How was Fig 10c constructed and what does the 'cumulated probability' mean in this context? Can you show that the increase in the identical nature of snow cover maps is not simply a function of decreasing number of snow maps pairs (again this links to greater clarification as to what is meant by cumulated probability). If this relationship is statistically robust, could more be made of this message, as understanding the influence of measurement resolution by satellite imagery is important to understanding the fate of snow and ice in small glacierized basins.

**Answer to 2.** We really appreciate that you value our work presented using the figures 7 to 10. With respect to figure 10c, you are firstly right that all analysed data at Vernagtferner and Zugspitze is used to generate the graph. The term 'cumulated probability' in figure 10c might be unclear, we will change it to the 'number of identical snow cover maps' where the total number is 63. The figure hence answers

the question how many of the snow maps generated with the calibrated and the standard threshold are identical in the complete catchment areas for the different pixel sizes between 30 m and 990 m. For example, 58 of the 63 snow maps (over 90%) are completely identical at a pixel size of 510 m. Now, we come to the final part of your question where you asked if the identical nature of snow cover maps is not simply a function of decreasing number of snow map pairs and if this relationship is statistically robust. We can neither agree nor deny this statement from our data. We thought about this effect and our data shows often increase but also at other dates decrease in agreement for coarser resolutions. The simple reason is that one pixel being different has a stronger relative effect with coarser resolution as our catchments are small and the number of pixels becomes low. We therefore decided to not look at the relative increase in agreement of the calibrated and standard snow cover maps with larger pixel sizes but to focus on the pixel size at which the snow cover maps become completely identical as this is independent of relative changes with aggregation. Fig. 10 c outlines when total agreement of the snow cover maps for the different NDSI thresholds was found.

#### Manuscript changes to 2.

Figure 10



p.11, l1 to 16: "Our aggregation experiment of the Landsat snow cover maps for the different *NDSI* thresholds shows that the SCA deviation between standard and calibrated snow cover maps diminishes for coarser resolution data in most cases. Figure 10 a outlines this error reduction with spatial aggregation for a Landsat 7 scene of Vernagtferner catchment on 16 September 2011. Figure 10 b shows the simultaneously captured photograph used for calibration. We however cannot draw an absolute conclusion from fig. 10 a that the difference in snow cover maps between the different thresholds is always reduced with a coarser resolution. The simple reason is that with larger pixel sizes,

the number of pixels in the catchment becomes lower and the relative weight of a pixel being different for different thresholds has a larger relative weight. Therefore, we decided to investigate at which spatial resolution the standard and calibrated snow cover maps become identical for the 63 cases investigated in the two catchments. This variable is absolute and thus independent of relative weights and changes with spatial aggregation. The aggregation step to 510m is thereby of major importance as more than 90% (58 of 63) of SCA maps become identical at this pixel size. Thus, using the standard threshold of 0.4 instead of the higher *NDSI* thresholds at VF and the lower *NDSI* values at RCZ seems to be accurate in most cases with a pixel size of 500m. For applications at this scale, the positive effect of using camera calibrated data diminishes and might rarely justify the effort."

# Specific comments:

Note: We do not show each of the manuscript changes for the specific comments here as the changes are obvious by the answer. Nonetheless, the changes will be denoted in the new manuscript.

**Pg1, In19**: Quantify how different the statistically insignificant correlation was to the standard threshold.

**Answer:** We add the correlation coefficient here.

Pg1, In 20: what is the 'another literature value'? State it here.

<u>Answer:</u> The other literature value is the locally optimized 0.7 threshold value of Maher et al. (2012) which was also found for single events at Vernagtferner. We add this to the abstract.

**Pg1, In 21:** replace 'case' with 'cases where'. <u>Answer:</u> Thank you for the correction.

Pg2, In 5: 'precipitation water' – just say 'precipitation'?Answer:, We delete the word 'water'.

**Pg2, In 10-13:** avoid single sentence paragraphs. Change this throughout the manuscript.

**Answer:** You are right, we change the single sentence paragraphs in our manuscript.

Pg2, In28: 'In this context' is superfluous and could be removed.<u>Answer:</u> We remove it, thank you.

Pg3, Ln 23: 'built up' is poor terminology for geological composition<u>Answer:</u> We agree and change the terminology.

Pg3, In 25: 'pending' is strange terminology. Do you mean 'underlying rock' or 'substrate'<u>Answer:</u> Thank you, we use 'underlying rock' now.

Pg3, In 29: no need for 'for' in the statement 'guarantees for comparable'<u>Answer:</u> Thank you for the correction.

Pg4, In 1: do you mean 'dates' rather than 'cases'? Stick to constant terminology.<u>Answer:</u> We totally agree, we change it.

Pg4, In 4: is 'rectifaciton' as spelling error?Answer: Yes, you are right, it should be 'rectification'.

Pg5, In 1: no need for 'It has to be mentioned that'<u>Answer:</u> We agree and remove it accordingly.

Pg5, In 22: – remove 'an'<u>Answer:</u> Thank you for the correction.

Pg5, ln 30 & 32: remove 'used' <u>Answer:</u> We remove 'used' in both sentences.

Pg5, In 32: 'misclassificied' is misspelt<u>Answer:</u> Thank you for finding this spelling error.

**P6, In 4-6:** tenses are used interchangeably. Suggest sticking to past tense consistently throughout the methods section

**Answer:** You are right, we change the tense in the methods section to past tense where appropriate.

Pg6, In 9: remove 'thereby' Answer: We correct it.

P6, In 13: no need for 'It has to be mentioned that'<u>Answer:</u> You are right, we remove it.

**Pg7, In 10:** what does 'underline' mean in this context, I think the wrong word is being used here. <u>Answer:</u> This is true, we wanted to clarify that the minimal differences between the Otsu method and the standard threshold does not justify the additional effort needed for the Otsu method. We use 'justify' now."

Pg7, In 20: – if statistically significant, then present the stats here (r-value and p-value).
<u>Answer:</u> We change 'significantly weaker' to 'weaker' as the expression was not meant in a statistically quantitative way here. It is a qualitative statement.

**Pg8. Ln 17:** – remove 'in percents' **Answer:** Thank you for the correction.

Pg8, In24: – what chose a threshold of 0.7? Provide some justification.

<u>Answer:</u> We have chosen the 0.7 threshold as it is in the range of plausible NDSI threshold values (0.35 to 0.7) that might result from a single date calibration at VF. Moreover, Maher et al. stated this value in their study area when they also calibrated the NDSI threshold only for one date due to the lack of additional data. We simply want to show here that a NDSI threshold calibrated at a single date does not give too much insight how the NDSI threshold might look like at other dates but at the same time gives investigators a false sense of confidence. We clarify it.

### Pg10, In 11: - what is a 'date by date transfer'?

<u>Answer:</u> The term was misleading. We wanted to state that we tested if the NDSI threshold calibrated at one catchment can be used at another catchment in the same Landsat scene. We rephrase the sentence.

**Pg10, In 23:** – 'jeopardous' is probably not the correct term to use here. 'inappropriate' or something similar may be better.

**Answer:** Thank for the suggestion, we change it.

**Figures:** Use of titles within figures and sub-figures is unnecessary (e.g. above each sub-figure in Fig 7). Instead use the associated caption to clearly describe each figure. Often current captions are a mix of methods and results, rather than sticking to the bare minimum need to adequately describe what is presented. The main body of the text should instead be used to explain methodological procedures and results.

**<u>Answer</u>**: Thank you, we will change the figure captions and the respective text in the main body of the text.